International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 9, No.3 (May-2020)

http://dx.doi.org/10.12785/ijcds/090308

A Survey on Autonomous Techniques for Music Classification based on Human Emotions Recognition

Deepti Chaudhary[1], Niraj Pratap Singh[1] and Sachin Singh[2]

[1]*Department of Electronics Engineering, National Institute of Technology Kurukshetra, Haryana 136119, India*
[2]*Department of Electrical and Electronics Engineering, National Institute of Technology, Delhi 110040, India*

Received 3 Jun. 2019, Revised 28 Mar. 2020, Accepted 13 Apr. 2020, Published 1 May 2020

Abstract: Music is one of the finest element to trigger emotions in human beings. Each and every human being feels the music and emotions are automatically provoked by listening music. Music is considered as strong stress reliever. With the increase in size of music dataset available online and advancement of automation technologies the emotions from the music are to be recognized automatically so that the online database of music can be organized and browsed in an efficient manner. Automation of music emotion classification (MEC) helps the people to listen the music of their interest without wasting time on surfing the internet. It helps the psychologists in treatment process of patients. It also helps the musicians and artists to work on specific type of music and to classify them. This paper aims to provide the overview and survey related to autonomous technique for music classification (ATMC). In this article, the basic steps such as database collection, preprocessing, database analysis, feature extraction, classification and evaluation parameters involved in ATMC are explained and comprehensive review related to the basic steps is summarized. Research issues and solutions related to ATMC along with future scope are also discussed in this article.

Keywords: Music emotion classification (MEC), Feature extraction, Classification techniques, Evaluation Parameters

1. Introduction

With the vast increase in the digital music data available online and offline the demand for efficient techniques for tag based content search, proper organization of online metadata and classification has also been increased. Music can be organized and described on the basis of various parameters such as genre, lyrics, artist, mood and emotion. With the increase in electronic media and interactive access the automatic task for classification of largely available music data is required.

Everyone feels the emotion in music that is induced by musician, singer or artist. Even a kid starts responding in various ways to music of different genres. Every work becomes interesting by playing music in background. It is considered as strong stress reliever. According to C.C. Pratt, music is considered as mode of expression for emotions[1]. It cannot be composed, realized or entertained without affection involvement. Emotion is the energy that brings a person in motion and music is the energy to induce emotions in humans. Thus music is the energy to control the human motion or in other words music acts as the driving force for human beings depending on the emotion induced by the songs. The medium of music has evolved specifically for the expression of emotions, and it is natural process for humans to organize music in terms of its emotional expressions but quantifying it empirically proves to be a very difficult task. Various types of emotions like peace, relax, angry, surprise, affection and lonely etc. can be sensed from music. Millions of songs are freely available online for immediate download. The humans naturally judge the music on the basis of emotions induced by them. Music and emotion are highly correlated to each other. The need of music management and recognition systems arises with the increase in wireless network bandwidth, widespread use of the internet, increase in number of the mobile users, online and offline availability of music in various stores, handy devices to play and record music, musical games, music therapies used by doctors etc. The music can be managed by considering many factors such as language, title, artist name, album name, genres, mood, emotion etc. The music information retrieval evaluation exchange (MIREX) is the evaluation campaign for music information techniques and systems coordinated by International Symposium on music information retrieval (ISMIR) annually. The aim of this symposium is to provide the exploration to various techniques and

E-mail:deeptinitk@gmail.com, nirajatnitkkr@gmail.com, sachinsingh.iitr@gmail.com

algorithms for research in MIR. MIREX introduces the automatic music classification (AMC) task in 2007 [2]. ATMC is an interdisciplinary research area that includes the detailed study of information retrieval mechanism from music clips, study and implementation of classification techniques and processes that are associated for classifying music on the basis of emotion. ATMC systems are designed for handling the huge amount of digital database of music that can be accessed for entertainment, research and other issues. The ATMC system based on human computer interaction is used to automatically detect the emotion of musical clips[3] and this automatic system is designed by conducting five basic steps shown in figure 1 are described below.

1) Dataset collection- A large database that covers all types of music belonging to different genres is considered for research by keeping in view that the database should not be affected by album effect or artist effect [4].The various types of standard datasets such as MediaEval 2017[5], DEAP[6], CAL 500 [7] are also available online for research.
2) Preprocessing- The music database is reformed to a precise standard format such as sampling frequency (44100 Hz), precision (16 bits) for fair evaluation. The emotion perceived from a song is not stable for its entire duration. The complete music can contain sections of different emotions. So, most representative 20-30 second segment of a song is considered and is used for emotion detection [2].
3) Database analysis- The category of emotion belonging to songs is commented by a group of people called subjects. The average opinion of subjects is considered as the category of emotion for a particular song. The categories used by subjects are defined by categorical and dimensional approaches. These approaches are used to divide the dataset in different classes on the basis of emotion.
4) Feature extraction- Features such as energy, pitch, timbre, tonality and rhythm that contain the information of music clips are elicited to represent the perceptual dimensions of songs. Various tools such as Psysound [8], Marsyas [9] and MIR toolbox [10] are used to extract the features.
5) Classification- The database belonging to various categories with different features is used for classifier's training to determine the relationship between emotion and music. The various type of classifiers are support vector machines (SVM), Gaussian mixture model (GMM), K nearest neighbor (KNN) and combination of Artificial Neural Network (ANN), Back Propagation (BP) Neural Network, Convolution Neural Network(CNN) etc. Finally performance of the system is determined in terms of evaluation parameters such as accuracy, precision, recall, f-score, g-mean etc.

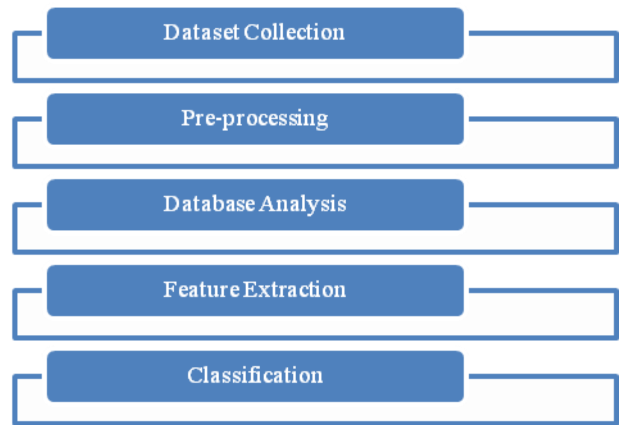

Figure 1. Block diagram of ATMC

The research related to ATMC is interdisciplinary in nature. Database preparation and music collection is related to musicians and artists, Database analysis and categorization is related to psychologists and sociology research field. Classification algorithms are developed by the researchers working in the field of machine learning, audio signal processing and affective computing. Researchers working in the field of musicology, psychology, sociology and affective computing can share their knowledge and combine it to design efficient ATMC.

By noticing the multidisciplinary nature, the aim of this chapter is mentioned with the help of following points.

1) To describe the relationship between music and emotion.

2) To explore the methods used to describe and analyze the music emotions

3) To provide the detailed study and comprehensive review of ATMC. This article will be helpful to the researchers working in different areas involved in ATMC.

In this paper survey related to music database collection is given in section 2. Section 3 describes the pre-processing technique used in ATMC. Database analysis described in section 4. Section 5 is dedicated to feature extraction, classification techniques are described in section 6 and evaluation parameters are reviewed in section 7.

2. Music Database

A large database consisting of all the genres related to different languages is used for ATMC. The database should not be from the same album and of same singer and artist. The database should be collected widely from various albums and websites for research. The database available with MIREX can be used by the researchers by signing the agreement for not sharing the database commercially. The datasets such as Remote collaborative and affective interaction RECOLA [11] , Magna Tag A tune[12], Million Song Dataset (MSD) [13], AMG 1608[14], MER 60[15], DEAP, MediaEval [5], GTZAN [16], CAL500 are freely available online by various institutes or research centers to enhance the research on ATMC. Some authors prefer to collect their own dataset to apply MEC on different languages. The review for the available dataset is represented in table 1.

TABLE 1. DATASET REVIEW

DATASET	RELATED WORK
Self collected	[17], [18], [19], [20], [21], [22]
RECOLA	[23]
CAL500	[24], [25], [26]
Magna Tag A Tune	[27], [28]
MSD	[27], [29]
AMG 1608	[30], [31], [32]
MER60	[30], [33]
DEAP	[34], [35], [36]
Mediaeval	[37], [38], [39]
GTZAN	[16]
Marsyas	[40]

3. PREPROCESSING

As the emotion perceived from the song is not constant throughout the entire song and it varies across with the segments of songs, a short time segment of the song is considered for the research. The song length considered by researchers to avoid emotion variation is 25 to 45 seconds. If the length of the clip is less than this range then for such short duration clips the emotion cannot be judged correctly and for longer clips the emotion of within the song is not stable. A variety of methods are adopted by different researchers to attain the short time segment of the song. The segmentation process can be carried out by various methods such as:

a) The short segment of 25 to 45 second duration of the song is considered by neglecting the first 30 seconds of the song.

b) The short segment of 25 to 45 seconds that represents the most influencing emotion of the song is considered.

c) The 25 to 45 seconds chorus part of the song is considered.

d) The mid section of duration 25 to 45 seconds is considered.

e) The last 25 to 45 seconds of the song is considered.

The short segment song requires considerable less time resulting in increased constancy in user's ratings. It has been noted from the review that a 30 segment clip is common choice[41].

The music clips are also not available in standard format. It is the prime requirement to convert the music database in standard format for their comparative analysis. The standard format normally considered by researchers is 22050Hz maximum frequency and 44100 Hz sampling frequency, keeping in view the frequency range of audio signals i.e. 20Hz to 20 KHz and 16 bits precision and mono-channel. Music clips also undergo the normalization process. In this process the windowing and framing techniques are used. Windowing is directly in co-operated with the Fourier transform function. The sound signals are non-stationary, thus the analysis of sound signals is carried out by considering short time signals. The process of transforming the sound signal in short time signals is framing. Authors make use of hamming [7], [42]–[44] and hanning windows [45], [46] for preprocessing the signals. Gabor function can also be used for preprocessing the signals [47]. An automated tool named Cool Edit Pro is also used to preprocess the music signals [48], [49].

4. DATABASE ANALYSIS

In this section music database is analyzed in order to categorize it into different emotion classes. Emotion related to music varies from person to person, thus it is considered as a subjective concept. The database is collected from various sources and annotated by a group of subjects. The subjective analysis of the music clips can be carried out either by a group of experts or untrained group of people [41]. The expert group consists of less number of people generally less than five who have in-depth knowledge of music and are employed for the task of database analysis [50]. In untrained group the database analysis task is given to more than ten people and each song is annotated by the whole group. The average opinion of the subjects is considered as final category of the music clip. The database analysis process can be carried out by considering either categorical approach or dimensional approaches of emotion classification as described in sections 4.1 and 4.2.

Huron described the four parameters style, genre, emotion and similarity on the basis of which classification of music can be carried out [50]. The study related to emotion labeling has been reviewed in this section. Emotions can be classified as expression, perceiving and felting emotion. Expression emotion is the emotion induced by the performer for effective communication. Perceiving and felting emotions are the emotional responses of listeners. Both of the emotions are dependent on interplay among musical, personal and situational factors. Perceiving emotion is intrinsically subjective and can be perceived differently. Felting emotion refers to an emotion that is actually experienced by the listeners. It is similar to perceiving emotion. In the research field of music the keywords used for emotions are well defined by psychologists and they use the words that are used by human beings to express their emotion. From the literature study two main types of approaches: categorical and dimensional are identified to define emotion models. Categorical approach is defined discretely and makes use of clusters using adjective terms to define the emotion and dimensional approach is defined dimensionally and represents the emotions on the basis of their positions on the emotion planes.

A. Categorical Approach

The relationship between emotion and music is explored by Hevner in 1936 [51]. In this model, author makes use of the described discrete cluster of emotions. The adjectives related to emotions are used in eight different categorical clusters as shown in figure 2.

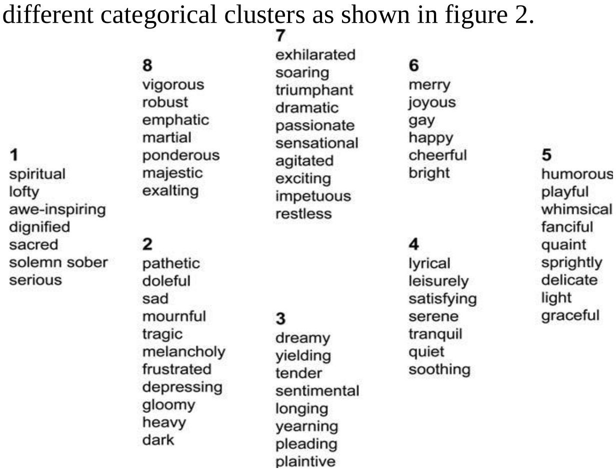

Figure 2 Hevner's model of emotion [51]

The emotional clusters formed by Hevner were reexplored by Farnsworth [52] by using ten groups of emotional terms and in nine groups by Schubert in 2003 [53] as represented in table 2.

TABLE 2 SCHUBERT'S EMOTION MODEL [53]

Cluster	Emotions in Each Cluster
1	Bright, cheerful, happy, joyous
2	Humorous, light, lyrical, merry, playful
3	Calm, delicate, graceful, quiet, relaxed, serene, soothing, tender, tranquil
4	Dreamy, sentimental
5	Dark, depressing, gloomy, melancholy, mournful, sad, solemn
6	Heavy, majestic, sacred, serious, spiritual, vigorous
7	Tragic, yearning
8	Agitated, angry, restless, tense
9	Dramatic, exciting, exhilarated, passionate, sensational, soaring,triumphant

MIREX makes use of categorical approach and makes use of five emotion clusters to define emotions as represented in table 3 [2].

TABLE 3. MIREX EMOTION CLUSTERS

Clusters	Emotional terms
Cluster_1	Passionate, rousing, 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

B. Dimensional Approach

Dimensional approach used for emotion categorization is based on the positions of emotions on dimensional plane. The dimensions on the plane are given by considering the relationship between basic factors that are used to differentiate the emotions. The placement of the emotion on dimensional graph depends on the correlation between the axes scales and the large number of terms is used to describe the varying emotions on the bases of their variability on axes of emotion plane. In 1980 Robert Plutchik proposed first 2-dimensional wheel model85 and 3-dimensional model in cone-shape to represent relationship between different types of emotions [54]. Authors considered eight basic types of emotions: anger, disgust, fear, joy, sadness, surprise, anticipation and trust and arranged them circularly as shown in figure 3. The emotions are represented by different colors in this model. As shown in figure similar colors are used to represent the similar type of emotions with variable strengths (e.g. ecstasy-joy-serenity) and terms representing opposite emotions are placed against each other (e.g. joy-sad). The emotions shown in table in different colors can be mixed up to obtain different emotions (e.g. combination of serenity and acceptance create love emotion). By using this basic differentiation of emotion along the axes various dimensional models are proposed by authors and these models represent the emotions in continuous plane by considering two or three dimensions. These dimensions are related with valence, arousal and dominance. Valence term deals with the positive and negative types of emotional terms, arousal term deals with the energy or stimulation level of song and dominance deals with the level of measuring strength of influencing power.

The three dimensions pleasure, arousal and dominance related to emotion were described by A. Mehrabian and J.A. Russell in 1974. Another two dimensional circumplex model of emotion had been proposed by Russel in 1980 [55]. In this model valence and arousal are considered as major dimensions. The horizontal dimension of the model is related with positive and negative emotions whereas vertical axis of the model is related with positive arousal and negative arousal as shown in Figure 4.

The same types of emotions are placed in the same quadrant and opposite emotions are placed in the opposite quadrant. For example the first quadrant of the model deals with positive arousal – positive valence emotions covering the emotions such as happy, glad, delighted excited etc., second quadrant deals with positive arousal- negative valence types of emotions covering the emotions such as angry, tense, frustrated etc., third quadrant is related with negative arousal- negative valence type emotions such as sad, bored, tired etc. and fourth quadrant consists of negative arousal and positive valence type of emotions such as calm, relax, satisfied etc.

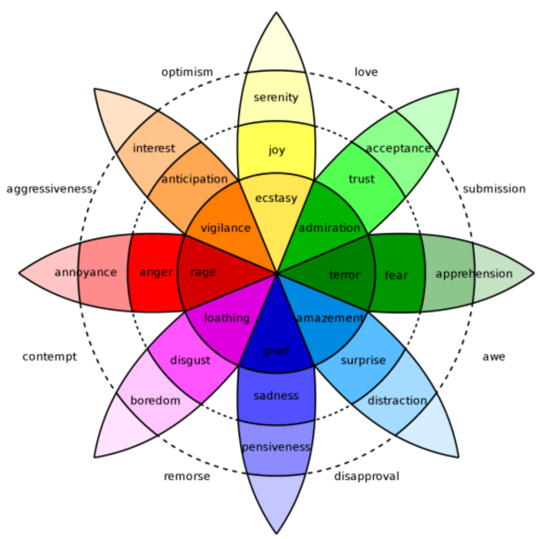

Figure 3. Plutchik’s model of emotion [54]

Authors make use of 28 adjective terms related to emotion in four different ways by making use of Ross technique [56] to obtain the model. This technique is used for ordering the variables in circular pattern, implementation of a multidimensional scaling technique on similar emotional terms and one-dimensional scaling on presumed degree of valence and arousal dimensions.

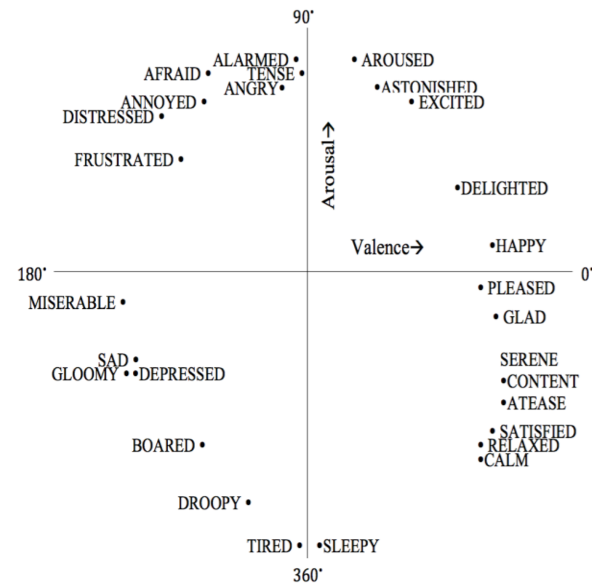

Figure 4. Russel’s model of emotion [55]

Another two dimensional emotion model is proposed by Thayer in 1989 [57]. In this model the relationship between tension and arousal is described in two dimensions. First quadrant includes the emotional terms related to positive energy and positive tension. The emotions such as happy and exciting exist in this quadrant. Second quadrant includes the emotional terms related to positive energy and negative tension. The emotions such as anxious and angry belong to second quadrant. Third quadrant includes the emotional terms related to negative energy and negative tension. The emotions such as sad and depressed exist in this quadrant. Fourth quadrant includes the emotional terms related to negative energy and positive tension. The emotions such as relaxed and calm are considered in this quadrant.

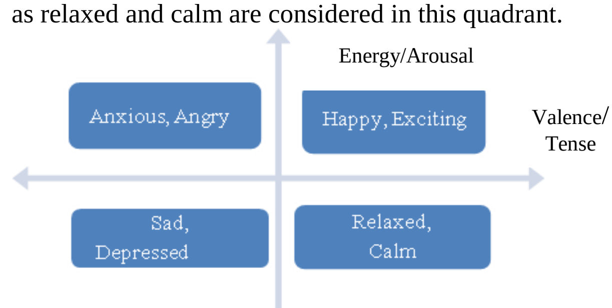

Figure 5. Thayer’s model of emotion[57]

E. Bigand et al. [58] proposed a 3-dimensional space to represent emotions, by considering arousal, valence and dominance as primary factors. J. Fontaine et al.[59] proposed four dimensions to represent the emotions. The emotional dimensions proposed by author are evaluation evaluation-pleasantness, potency-control, activation-arousal, and unpredictability.

Geneva Emotional Music Scales (GEMS) is the instrumental device designed for measurement of emotions that are perceived by listening music [60]. This model consists of 45 labels to describe the emotional terms related to music and these emotional terms can be categorized in 9 different groups.

Navrasa is a set of the nine emotional terms that are used to describe the Indian classical music [19]. The model is represented in figure 6

Sringar (Love/Beauty)	Hasya (Laughter/Happy)	Karuna (Sorrow/Sad)
Raudra (Anger)	Veera (Heroism/Courage)	Bhayanak (Terror/Fear)
Vibhadsa (Disgust)	Adbhut (Surprise/Wonder)	Shantha (Peace/Tranquility)

Figure 6. Navrasa emotional model [19]

Researchers make use of the different emotion models to work in the field of emotion recognition. The work related to ATMC is represented in table 4. As database analysis process is subjective, it is time consuming and costly as one have to search the group of experts in music and psychology to determine the correct class of the dataset. Authors consider various methods of database analysis to overcome the gaps[49]. Some reduce the length of music piece to reduce the time taken for database analysis process.

TABLE 4. RELATED WORK OF DATABASE ANALYSIS PROCESS

Database analysis model	**Related work**
Thayers	[17],[61],[62],[48],[63], [19],[64],[65]
Categorical	[66]
Russel's	[18], [40], [67]
2-d	[34]
GEMS	[68]
Indian classical model	[19]
Hevner	[64]

Authors may also provide the list of adjective terms and their synonyms to categorize the songs to reduce the time consumption. Some example songs with defined categories may also be provided to the group for better judgment of the class belonging to particular song. A user friendly interface may also be provided to the group for database analysis process. Some training lectures may also help the group for database analysis process and enhance its quality. A clear set of instructions may also be provided to the group of annotators to understand the purpose and method of database analysis. It may save the time and enhance the quality of database analysis process. Web based games are also available for database analysis process. In such games multiple users are allowed to play the game simultaneous and given the task of database analysis. D. Turnbull's listen designed a game in which a list of related words is displayed and the player is asked to choose the best and worst words to represent the emotion of the songs [69]. The score of the players is given on the basis of choices of other player's playing simultaneously. Aljanki also designed such game named emotify for database analysis[44]. It a game on emotions provoked by listening the songs. The choice of the player related to emotion of a particular song is used for research on music emotion by Utrecht University.

5. FEATURE EXTRACTION

Feature extraction is the process of determining the attributes related to the input data to perform the desired task. Music study is multidimensional and various parameters such as genre, emotion and mood can be perceived from music by considering various features related to them. Thus feature extraction is considered as one of the important step for ATMC. The emotional dimensions of the music are broadly represented by five features i.e. energy, rhythm, temporal, spectral and harmony[49]. These features are further divided in subcategories as shown in table 5.

Deepti et al. compared all the features and concluded that spectral features provide better results than other features[70]. Fu et. al. categorized the features for audio signals in three levels i.e. low level features, mid level features and high level features as shown in figure7.

Low-level features consist of timbre and tonality. Timbre is related with the sound quality of the music clips. Temporal features deals with the variation of timbral features with respect to time [71]. Timbre includes various low level features such as zero crossing rate, spectral centroid, spectral roll off, spectral flux, spectral crest factor etc. Temporal features include various subfeatures such as statistical moments, amplitude modulation and autoregressive modeling. Low level features are widely used in ATMC due to its better performance.

TABLE 5. BASIC TYPES OF FEATURES [49]

Type	Subfeatures
Energy	Dynamic loudness
Rhythm	Beat histogram, Rhythm pattern, rhythm histogram, tempo, Rhythm strength, rhythm regularity, rhythm clarity, average onset frequency, and average tempo
Temporal	Zero-crossings, temporal centroid, and log attack time
Spectral	Spectral centroid, spectral rolloff, spectral flux, spectral flatness measures, spectral crest factors, mel-frequency cepstral coefficients, spectral contrast, Daubechies wavelets coefficient histogram, tristimulus, even-harmonics, odd-harmonics, roughness, irregularity, and inharmonicity
Harmony	Salient pitch, chromagram centroid, key clarity, musical mode, harmonic change, pitch histogram, sawtooth waveform inspired pitch estimate

Mid-level features consist of rhythm, pitch and harmony. Rhythm represents the pulses of varying strength. It includes tempo and meter. Pitch is related with the perceived fundamental frequency of the sound [49]. Harmony is related with the analysis of superposition of sound and deals with simultaneous occurring frequencies. Top level features described the general method of categorizing the song by listeners i.e. genre, mood, style, artist, instrument etc.

Different types of toolboxes such as MIRtoolbox [10], Marsyas [9], jAudio[72], Psysound [8], openSmile[73] etc. are available for feature extraction of music signals. MIRtoolbox is open source toolbox based on MATLAB programming and it provides the MATLAB functions for different features such as dynamics, tonality, rhythm, spectrum etc. The researchers can directly use these functions for extracting features of music signals. Statistical analysis of features is also provided by this toolbox. Signal processing toolbox is required for MIRtoolbox.

Music Analysis, Retrieval and Synthesis for Audio Signals (MARSYAS) generally provide the efficient structure for audio signal processing and emphasized on music information retrieval. It provides large number of modules for audio signal processing. These modules consists of command line programs to extract audio features, C++ library consists of basic units for audio processing, a programming language names MARSYAS script that makes the processing of audio signals easier and a program named MARSYAS run for execution of MARSYAS scripts.

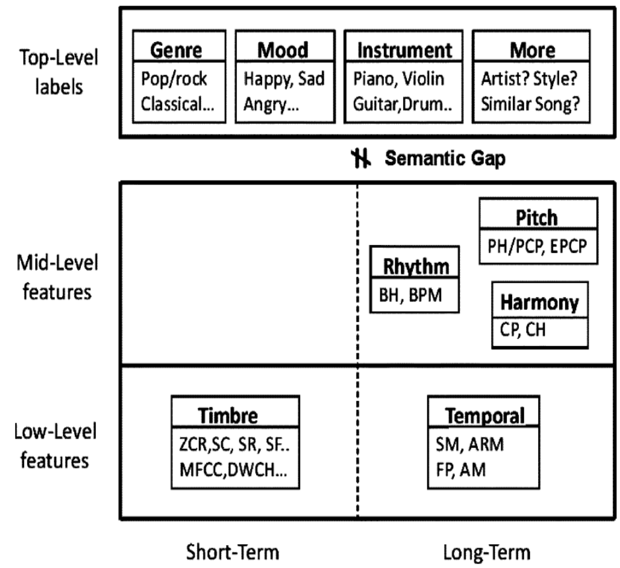

Figure 7. Basic categorization of features [71]

Essentia is a open source tool for the analysis of music signals and retrieval of features related to music [74]. This tool can be used by using C++ library with python bindings. The library consists of collection of reusable algorithms for analysis of audio signals. jAudio is user friendly application program for feature extraction of audio signals. jAudio can be used to extract primary features that are defined in the application and these primary features can be used to derive other features named derived features. These features can further be used by machine learning tools such as WEKA tool to extract the unknown class belonging to the song. A GUI or command-line interface with embedding support can be used to run this application.

Yet Another Audio Feature Extractor (YAAFE) is another toolbox for audio analysis. This toolbox is user friendly and can be used to extract large number of music features[75]. It supports WAV and MP3 music files and can be implemented by using C++, Python or MATLAB applications.

PsySound3 is a tool used for analyzing audio signals. It simplifies the process of feature extraction of audio signals. Features such as roughness, loudness, tempo and articulation etc can be extracted by using this toolbox. This program can be installed in MATLAB environment.

Open smile toolbox is used for feature extraction and pattern recognition tool. SMILE is an abbreviation used for speech and music interpretation by large-space extraction. This tool is used to extract real time large audio and music features. This tool can be implemented by using C++.

The Tempogram Toolbox is also available to extract tempo and pulse related features of audio signals. The Chroma Toolbox is also MATLAB based toolbox used to extract pitch and chroma based features of audio and sound signals. The Sound Description Toolbox (SDT) is used to extract features such as energy, harmonic, perceptual, spectral etc from WAV audio and sound files [76]. Rhythm Pattern Extractor is used to extract rhythm features of music if tempo feature is known. Table 6 represents the software used for implementation of toolboxes used by various authors.

TABLE 6. STATE-OF-ART FEATURE EXTRACTION TOOLBOXES

Toolbox	Software used	Related work
MIRToolbox	MATLAB	[15], [17], [20], [30], [31], [33], [34], [37], [38], [48], [62], [77], [78]
MARSYAS	C++ library, MARSYAS script	[3], [15], [28] [40], [67], [79], [80]
Essentia	C++ library with python bindings	[40]
jAudio	GUI, Command line interface	[19], [22]
YAAFE	C++, Python or MATLAB	[31]
PsySound	MATLAB	[3], [30], [48]
OpenSmile	C++	[22],[44]
Tempogram	MATLAB	[30]
Chroma toolbox	MATLAB	[30]
SDT	MATLAB	[15], [37], [38], [48]
Rhythm pattern extractor	MATLAB	[15]

The feature normalization techniques are used after feature extraction for fair comparison of value of each feature. Feature normalization can be carried out by two methods.

1) Linear normalization – In this method, the range of each feature is set between zero and one [0, 1].
2) Z-score normalization- In this method, each feature is normalized to zero mean value and the standard deviation is set to unity.

Before classification process the feature selection techniques are applied to minimize the number of random variables by selecting the principal variables from the features that are extracted. The various techniques such as Principal component analysis [81], ReliefF [82], Sequential forward floating selection (SFFS), Genetic search [83], Sequential backward search can be used to select the appropriate features from all the extracted features.

Principal component analysis (PCA) is a statistical procedure to convert a group of correlated features by using orthogonal transformation and creates a new set of uncorrelated features [81].

ReliefF depends on k- nearest neighbors, where k parameter represent the nearest hits and nearest miss for each feature of the samples[82]. The importance of features is estimated depending on the performance of algorithm to distinguish between the variables on the basis of feature variability.

Sequential feature selection methods finds a reduced set of features by selecting m-dimensions from feature space consisting of n dimensions where m<n [84]. The Sequential forward floating sequence (SFFS) and sequential backward floating sequence (SBFS) finds the new redundant set of features by adding or deleting the features from the subset of features. Genetic algorithm is a stochastic method for function optimization based on the mechanics of natural genetics and biological evolution[83]. Table7 represent the work related to feature selection algorithm.

TABLE 7. STATE-OF-ART FEATURE SELECTION ALGORITHM

Feature selection algorithm	Related Work
RRelief	[3], [18], [35], [62]
Sequential forward floating selection(SFFS)	[48], [61]
PCA	[38], [85]
Genetic search	[85]
Ranker	[85]

6. CLASSIFICATION

This step consists of training and testing process. After feature extraction the training process of classifiers is carried out by using data belonging to different classes. The different types of classification process such as regression models, k-NN, SVM, GMM, ANN, deep learning networks and Naive Bayes etc are described in this section.

i) Regression models are used as classifiers to determine relationship between dependent and independent variables. The performance parameter for regression models is R^2 statistics that is used to fit the data to the regression line [48], [64], [86].
ii) KNN classifiers are used to store the data for all the categories and new classes are classified on the basis of distance functions. The test data is classified based on the majority vote of its neighbors [87], [88].
iii) SVM classifiers are based on supervised learning techniques and algorithms. The dataset is divided into training and testing part [89]–[93]. The training data is used to train the SVM by marking the particular category. Further the test data is analyzed to check their category by determining

the hyper plane that maximizes the distance between the classes.

iv) Gaussian Mixture models (GMM) are basically used to detect the likeliness for ordinarily scattered data within overall dataset. The presumption of scattered data belongingness to particular class is not required in this case resulting in unsupervised learning [26], [33], [94].

v) Random Forest classifier creates a group of decision trees by using a random subset of training dataset [11], [19], [95]. The decision about the class of test data is based on the aggregate of the result of different decision trees whereas a decision tree is a flowchart based structure in which experiments are represented by internal node and the results of the experiments are represented by branch and the class labels are represented by leaf nodes of the tree.

vi) BP neural network is multi- layer feed forward network whose training process is based on error back propagation algorithm [16], [96]. These networks can be used to store the mapping relations of input-output models, and prior knowledge of these relations is not required in training process.

vii) NB technique is used as classification algorithm. The class labels are assigned to problem instances and described by using feature values in vector form [78]. Class labels can be chosen by any one of the method described above.

viii) Deep learning is a branch of neural network that makes use of multiple hidden layers for feature extraction and transformation. The output of a layer becomes the input of next layer. The deep learning techniques can be used for supervised and unsupervised tasks. Deep learning requires large amount of labeled data and the features are not extracted separately. The deep learning architectures learn the features directly from the dataset. CNN and recurrent neural networks are commonly used deep learning networks [44], [97], [98]. A CNN network is formed by combination of input layers, hidden layers such as convolution layers, RELU layer or pooling layer and an output layer. The audio signals go through the pre processing steps before applying to the CNN layers. Spectrograms of the audio signals are generated in pre processing. In RNN the output of previous step becomes the input of first step and consists of hidden state with memory to remember the information about a sequence. The classifiers described in this section are used by various authors across the world to detect the emotion automatically. The state-of-art for the classifiers is summarized table 8.

TABLE 8. STATE-OF-ART CLASSIFIERS

Classifiers	Related Work
Regression	[15], [30], [40], [38], [99]
Convolution LSTM	[34],[100]
Naïve Bayes	[18], [37]
SVM	[3], [18], [25], [37], [38], [48], [61], [62], [66], [67], [68], [79], [101]
CNN	[27], [36], [100]
Random Forest Classifier	[19], [38]
GMM	[26], [32], [33], [79], [102]
BP neural network	[103]
KNN	[78]

7. EVALUATION PARAMETERS

The performance of the ATMC systems can be measured on the basis of various evaluation parameters. In this section the evaluation parameters used by various authors are reviewed. The basic evaluation parameters are True Negatives (TN), True Positives (TP), False Negatives (FN) and False Positives (FP) from which other parameters can be derived. In above terms positive terms are related to the presence of particular class and negative term deals with the absence of particular class. TP is the outcome of the classifier when it predicts the class when that particular class is actually present. TN is result of the classifier when is does not detect the absent class. A FP is an outcome of classifier when it detects the absent class and FP is an outcome where the classifier incorrectly predicts the present class.

A. *Accuracy*

Accuracy is described as correctly projected outcomes out of total test class [104]. Accuracy is an important factor for any work. The greater accuracy percentage is important for any implementation. The accuracy is computed using the formulas given below

$$\text{Accuracy}=\frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

B. *Specificity*

Specificity is related to true negative rate. It measures the number of correctly projected data samples not belonging to particular category[104]. The specificity is computed by the formula as shown.

$$\text{Specificity}=\frac{TN}{TN+FP} \quad (2)$$

C. Precision

Precision is the fraction of relevant projected data out of projected data by the classifier [104].

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

D. Recall

Recall is the fraction of relevant projected data out of total relevant samples belonging to particular category present in the database [104].

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

E. F-measure

It is the harmonic mean of precision and recall [105].

$$F - \text{measure} = \frac{2 * (\text{Precision} * \text{Recall})}{\text{Precision} + \text{Recall}} \quad (5)$$

F. RMSE

A small deviation in the observation is identified using the RMSE [106]. The RMSE is computed using the formula

$$\text{RMSE} = \sqrt{\sum_{i=1}^{k} \frac{(\text{obtained result} - \text{original result})^2}{N}} \quad (6)$$

G. Pearson Correlation Coefficient(PCC)

It measures the degree of linear relationship between variables represented by a line [105]. It depends on the covariance and standard deviation of two variable X and Y.

$$\text{PCC} = \frac{\text{cov}(X,Y)}{SD(X) * SD(Y)} \quad (7)$$

Where cov is the covariance operator, and SD represents the standard deviation of X and Y. The value of correlation lies between [-1, 1].

H. R^2 statistics

R2 is a statistical measure of the closeness of data to the fitted regression line [105]. It is denoted by the formula given below

$$R^2\left(Y, r(X)\right) = \frac{(\text{cov}(Y, r(X)))^2}{\text{var}(Y) * \text{var}(r(x))} \quad (8)$$

Where Y is the true value, $r(X)$ stands for regression prediction model, cov is the covariance operator, var is the variance operator, R2 represents the proportion of underlying data variation in fitted regression model. The value of R2 lies in the range [$-\infty$, 1], 1 means the model perfectly fits the data, while a negative R2 means the model is even worse than simply taking the sample mean.

I. Area under ROC curve (AuC)

It is the area under the ROC curve, where ROC curve is the curve between True positives rate and false positives rate. AuC ranges between [0,1] [105]. Zero value of AuC means the predictions of the classifier are 100% false and 1 represents 100% correct predictions.

J. Equal error rate

It provides the threshold value for which the false acceptance rate and false rejection rate [77].

K. Low mean squared error

It is the average squared difference between the predicted values and what is actual values of the sample [66]. The evaluation parameters used by different authors are summarized in table 9.

TABLE 9. EVALUATION PARAMETERS REVIEW

Evaluation Parameters	**Related work**
Low mean squared error	[66]
F-measure	[18], [24], [25], [78], [107]
Accuracy	[19], [37], [38], [39], [48], [61], [67], [78], [99] , [101], [108]
Specificity	[61]
AuC	[27]
Pearson Correlation Coefficient	[68],[109]
R^2 statistics	[3], [15], [38], [40], [65], [79],
Precision	[24], [25], [26] , [78], [107]
Recall	[24], [26], [78], [107]
RMSE	[39]
Equal error rate	[77]

8. RESEARCH ISSUES

ATMC is the multidiscipline research field and it includes various steps as described in above mentioned sections. ATMC is still not much developed field of research and many research issues are anticipated from this review in this article. First research issue is related to the database. The standard database available is limited due to which the researcher's have to create their own dataset. To deal with such issues the researcher can collect the large database from freely available online websites depending on the interest of researcher. The collection of dataset should not face artist, language or genre effect. Second research issues are related to database analysis process. As the analysis process is subjective this process is expensive and time consuming. The major problems in analysis faced by researchers are to distinguish between induced and perceived emotion. The perceived emotion varies from person to person and depends on situational factors. Database analysis process also undergo the problem of granularity as the number of emotion classes is small as compared to the emotion classes that can be perceived from the song and it also faces ambiguity issue as the same emotion term can be defined by a number of adjective terms. To deal with analysis issue one should appoint the expert annotators to annotate the songs or one should try to make this process algorithmic. It is also seen in the review that most researchers are trying to increase the number of features related to music for better accuracy, but increasing features will improve the system to an extent. Further improvement in the ATMC system can be done by selecting the best features among all with the help of appropriate feature redundant techniques. Inspite of so many classification algorithms the ATMC process still possess limitations mentioned below.

1) Millions of songs are available online, but the MEC is limited to thousands of songs.

2) Techniques used to classify the songs are also limited to few types and languages of songs.

9. CONCLUSION & FUTURE SCOPE

In this article an extensive review of the ATMC has been presented. The detailed discussion of datasets used for ATMC, database analysis methods, pre-processing, audio features, classification techniques and evaluation parameters is provided. As it has already been discussed that emotion is considered as parameter for music classification by MIREX in 2007, still there are many open issues that are to be considered as discussed in previous section. The issues regarding collection of large music dataset and their proper database analysis is still unsatisfactory and needs lot of attention so that the songs of all the genres and languages can be considered by researchers working in this field. The classifiers used in literature are based on subjective database analysis; efforts can be made by researchers to design the classifiers that are not based on subjective database analysis. As the computation time is large for machine learning algorithms, new classification algorithms can also be designed. This article will help to gain inspiring knowledge to multidisciplinary researchers about ATMC. From the above discussion it can be noticed that there is much more space for the researchers of various fields and design improved ATMC.

REFERENCES

[1] C.C. Prat, Music as the language of emotion. The Library of Congress, 1950.

[2] X. Hu, J. S. Downie, C. Laurier, M. Bay, and A. F. Ehmann, "The 2007 MIREX Audio Mood Classification Task : Lessons Learned," in Proceedings of 9th International Conference on Music Information Retrieval, September 14 - 18, 2008, pp. 462–467, Philadelphia, PA, United States

[3] Y. H. Yang, Y.C. Lin, Y.F. Su, and H. H. Chen, "A Regression Approach to Music Emotion Recognition," in IEEE Transactions of Audio, Speech and Language Processesing, vol. 16, no. 2, pp. 448–457, 2008.

[4] Y. E. Kim, D. S. Williamson, and S. Pilli, "Towards Quantifying the 'Album Effect' in Artist Identification," in Proceedings of 7th International Conference on Music Information Retrieval, October 8-12, 2006 pp. 393–394, Victoria, Canada.

[5] B. Bischke, P. Helber, C. Schulze, V. Srinivasan, A. Dengel, and D. Borth, "The multimedia satellite task at mediaeval 2017: Emergency response for flooding events," in CEUR Workshop Proceedings, 2017, 13–15,September, Ireland, Dublin.

[6] S. Koelstra et al., "DEAP: A database for emotion analysis; Using physiological signals," in IEEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18–31, 2012.

[7] S.-Y. W. J.-C. W. Y.-H. Y. and H.-M. Wang, "Towards Time - Varying Music Auto-Tagging Based on CAL500 Expansion," in International Conference on Multimedia and Expo., 14-18 July 2014, Chengdu, China

[8] D. Cabrera, S. Ferguson, and E. Schubert, "Psysound3 : Software for acoustical and psychoacoustical analysis of sound recordings' Proceedings of the 13th International Conference on Auditory Display, June 26-29, 2007, Montréal, Canada.

[9] G. Tzanetakis and P. Cook, "MARSYAS: A framework for audio analysis," Organised Sound, vol. 4, no. 3, pp. 169–175, 2000.

[10] P. T. Olivier Lartillot, "A Matlab Toolbox for Musical Feature Extraction from Audio," in Proceedings of International Conference on Digital Audio Effects, Bordeaux, 2007.

[11] M. Valstar et al., "AVEC 2016 - Depression, Mood, and Emotion Recognition Workshop and Challenge," in Proceedings of The Audio-Visual Emotion Challenge and Workshop, 16 October 2016, Amsterdam, NL.

[12] E. Law, K. West, M. Mandel, and M. Bay, "Evaluation of algorithms using games: The case of music tagging," in Proceedings of 10th International Conference on Music Information Retrieval, October 26-30, 2009, Kobe, Japan.

[13] T. Bertin-Mahieux, D. P. W. Ellis, B. Whitman, and P. Lamere, "The Million Song Dataset,"in Proceedings of Proceedings 12th Internation Conference on Music Information Retrieval , August 9th to 13th, 2011, pp. 591–596, Utrecht, Netherlands.

[14] S. Shandilya and P. Rao, "Detection of the Singing Voice in Musical Audio," in Proceedings of 114th Convention of Audio Engineering Society, March, 2003, Amsterdam.

[15] Y. H. Yang and H. H. Chen, "Prediction of the Distribution of Perceived Music Emotions Using Discrete Samples," IEEE Transactions on Audio, Speech and Language Processing, vol. 19, no. 7, pp. 2184–2196, 2011.

[16] X. Yang, Y. Dong, and J. Li, "Review of data features-based music emotion recognition methods," in Multimed. Syst., pp. 1–25, 2017.

[17] I. Binanto, "A method of mood classification on keroncong music," IEEE Symp. Comput. Appl. Ind. Electron., pp. 19–24, April 28-29, Penang, Malaysia.

[18] R. Malheiro, R. Panda, P. Gomes, and R. P. Paiva, "Emotionally - Relevant Features for Classification and Regression of Music Lyrics," IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 45–55, 2016.

[19] A. M. Ujlambkar, "Mood classification of Indian Popular Music," in Sixth Asia Modelling Symposium, pp. 278–283, September 3–5, 2012, Pune, Maharashtra, India.

[20] Y. H. Yang and H. H. Chen, "Ranking-Based Emotion Recognition for Music Organization and Retrieval," in IEEE Transactions Audio, Speech Lang. Process., vol. 19, no. 4, pp. 762–774, 2011.

[21] T. Greer, K. Singla, B. Ma, and S. Narayanan, "Learning Shared Vector Representations of Lyrics and Chords in Music," vol. 0, no. 1, pp. 3951–3955, 12-17 May 2019, Brighton, United Kingdom.

[22] B. G. Patra, D. Das, and S. Bandyopadhyay, "Multimodal Mood Classification of Hindi and Western Songs," in Journal of Intelligent Information Systems, pp. 1–18, 2018.

[23] Z. Zhang, J. Han, J. Deng, X. Xu, F. Ringeval, and B. Schuller, "Leveraging Unlabeled Data for Emotion Recognition with Enhanced Collaborative Semi-Supervised Learning," in IEEE Access, vol. 6, pp. 22196–22209, 2018.

[24] J.-C. Wang, Y.-S. Lee, Y.-H. Chin, Y.-R. Chen, and W.-C. Hsieh, "Hierarchical Dirichlet Process Mixture Model for Music Emotion Recognition," in IEEE Transactions on Affective Computing, vol. 6, no. 3, pp. 261–271, 2015.

[25] B. L. Sturm, "Evaluating music emotion recognition: Lessons from music genre recognition," IEEE International Conference on Multimedia and Expo Workshops, July 8-12, 2013 , San Jose, California, USA.

[26] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet, "Semantic Annotation and Retrieval of Music and Sound Effects," IEEE Transactions on Audio, Speech and Language Processing, vol. 16, no. 2, pp. 467–476, 2008.

[27] J. Lee and J. Nam, "Multi-Level and Multi-Scale Feature Aggregation Using Sample-level Deep Convolutional Neural Networks for Music Classification," Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, pp. 3–5, August 6-11, 2017.

[28] X. Hu, J. S. Downie, and A. F. Ehmann, "Lyric text mining in music mood classification," in 10th International Society for Music Information Retrieval Conference, pp. 411–416, Kobe, Japan. October 26-30, 2009.

[29] G. Verma, E. G. Dhekane, and T. Guha, "Learning Affective Correspondence between Music and Image," IEEE International Conference on Acoustics, Speech and Signal Processing pp. 3975–3979, May 12-17 2019, Brighton, United Kingdom.

[30] X. Hu and Y. H. Yang, "Cross-Dataset and Cross-Cultural Music Mood Prediction: A Case on Western and Chinese Pop Songs," IEEE Transactions on Affective Computing., vol. 8, no. 2, pp. 228–240, 2017.

[31] Y. Chen, Y. Yang, J. Wang, and H. Chen, "THE AMG1608 Dataset for Music Emotion Recognition, " IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 693–697, April 19-24, 2015, Brisbane, QLD, Australia

[32] H.-M. W. and G. L. Ju- Chiang Wang, "A histogram density modeling approach to music emotion recognition," in IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 698–702, 19-24 April 2015, Brisbane, QLD, Australia.

[33] J. C. Wang, Y. H. Yang, H. M. Wang, and S. K. Jeng, "Modeling the affective content of music with a Gaussian mixture model," IEEE Transactions on Affective Computing., vol. 6, no. 1, pp. 56–68, 2015.

[34] B. H. Kim and S. Jo, "Deep Physiological Affect Network for the Recognition of Human Emotions," IEEE Transactions on Affective Computing., vol. 14, no. 8, 2018.

[35] S. W. Byun, S. P. Lee, and H. S. Han, "Feature selection and comparison for the emotion recognition according to music listening," Int. Conf. Robot. Autom. Sci., pp. 172–176, August 26-29, 2017, Hong Kong, China.

[36] P. Keelawat, "Subject-Independent Emotion Recognition During Music Listening Based on EEG Using Deep Convolutional Neural Networks," in Proceedings of IEEE 15th International Colloquium on Signal Processing & Its Applications, March 8-9, 2019, pp. 21–26, Penang, Malaysia.

[37] J. B. et Al., "Music emotion recognition by cognitive classification methodologies," in 16th International Conference on Cognitive Informatics and Cognitive Computing , pp. 121–129, July 26-28,2017, Oxford, UK.

[38] D. L. Hgrqj et al., "Dimensional music emotion recognition by valence-arousal regression," in Cognitive Informatics & Cognitive Computing, 2016, pp. 42–49.

[39] S. Mo and J. Niu, "A Novel Method Based on OMPGW Method for Feature Extraction in Automatic Music Mood Classification," IEEE Transactions on Affective Computing., vol. 3045, no. c, 2017.

[40] J. Grekow, "Audio features dedicated to the detection of arousal and valence in music recordings," in IEEE International Conference on Innovations in Intelligent SysTems and Applications, pp. 40–44, July 3-5, 2017 Gdynia, Poland

[41] K. F. MacDorman, S. Ough, and C. C. Ho, "Automatic emotion prediction of song excerpts: Index construction, algorithm design, and empirical comparison," J. New Music Res., vol. 36, no. 4, pp. 281–299, 2007.

[42] V. Rao and P. Rao, "Vocal melody detection in the presence of pitched accompaniment using harmonic matching methods," in Proceedings of 11th International Conference on Digital Audio Effects, September 1-4, 2008, pp. 1–8, Espoo, Finland.

[43] C. Y. Chang, C. K. Wu, C. Y. Lo, C. J. Wang, and P. C. Chung, “Music Emotion Recognition with Consideration of Personal Preference,” in The 2011 International Workshop on Multidimensional (nD) Systems, 5-7 Sept. 2011, pp. 1–4, Poitiers, France.

[44] A. Aljanaki, “Emotion in Music: representation and computational modeling,” September, 2016.

[45] M. Kos, Z. Kacic, and D. Vlaj, “Acoustic classification and segmentation using modified spectral roll-off and variance-based features,” Digit. Signal Processing, A Review Journal, vol. 23, no. 2, pp. 659–674, 2013.

[46] E. M. Schmidt, D. Turnbull, and Y. E. Kim, “Feature selection for content based, time varying musical emotion regression,” in Proceedings of the international conference on Multimedia information retrieval, p.p. 267- 274, March 29–31, 2010 Philadelphia, Pennsylvania, USA.

[47] S. Mo and J. Niu, “A Novel Method based on OMPGW Method for Feature Extraction in Automatic Music Mood Classification,” IEEE Transactions on Affective Computing, vol. 3045, 2017.

[48] S. Pouyanfar and H. Sameti, “Music emotion recognition using two level classification,” in Proceedings of Iranina Conference on Intelligent Systems, pp. 1–6, 2014.

[49] Y.-H. Yang, Y.-F. Su, Y.-C. Lin, and H. H. Chen, Music Emotion Recognition. Boca Raton, CRC Press, 2011.

[50] T. Li and M. Ogihara, “Detecting emotion in music,”in Proceedings of 4th Intenational Conference on Music Information Retrieval, pp. 239–240, October 26-30 2003, Baltimore, Maryland (USA)

[51] D. Huron, “Perceptual and cognitive applications in music information retrieval,” in Proceedings of 1st International Symposium on Music Information Retrieval, October 23-25, 2000, Plymouth, Massachusetts, USA.

[52] P. R. Farnsworth, “A Study of the Hevner Adjective List,” in *Journal of Aesthetics & Art Criticism*, vol. 13, no. 1, pp. 97–103, 1954.

[53] E. Schubert, “Modeling perceived emotion with continuous musical features,” in Music Perception: *An Interdisciplinary Journal*, vol. 21, no. 4, pp. 561–585, 1390.

[54] T. A. Burns, “The Nature of Emotions,” in International Journal of Philosophical Studies, vol. 27, no. 1, pp. 103–106, 2019.

[55] J. A. Russell, “A circumplex model of affect,” in Journal of Personality and Social Psychology, vol. 39, no. 6, pp. 1161–1178, 1980.

[56] R. T. Ross, “A statistic for circular series.,” J. Educ. Psychol., vol. 29, no. 5, pp. 384–389, 1938.

[57] R. E. Thayer, The biopsychology of mood and arousal. New York, NY, US: Oxford University Press, 1989.

[58] E. Bigand, S. Vieillard, F. Madurell, J. Marozeau, and A. Dacquet, “Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts,” in Cognition and Emotion, vol. 19, no. 8, pp. 1113–1139, 2005.

[59] J. R. J. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, “The world of emotions is not two-dimensional,” in Psychological Science, vol. 18, no. 12, pp. 1050–7, 2007.

[60] M. Zentner, D. Grandjean, and K. R. Scherer, “Emotions Evoked by the Sound of Music: Characterization, Classification, and Measurement,” in Emotion, vol. 8, no. 4, pp. 494–521, 2008.

[61] Y. L. Hsu, J. S. Wang, W. C. Chiang, and C. H. Hung, “Automatic ECG-Based Emotion Recognition in Music Listening,” IEEE Transactions on Affective Computing, pp. 1–16, 2017.

[62] C. Lin, M. Liu, W. Hsiung, and J. Jhang, “Music Emotion Recognition Based on Two-Level Support Vector Classification,” Proceedings of the 2016 International Conference on Machine Learning and Cybernetics, July 10-13, 2016, Jeju, South Korea.

[63] A. S. Bhat, V. S. Amith, N. S. Prasad, and D. M. Mohan, “An efficient classification algorithm for music mood detection in western and Hindi music using audio feature extraction,” in Proceedings of 5th International Conference on Signal and Image Processing, January 8-10, 2014, pp. 359–364, Bangalore, India.

[64] Y. Panagakis and C. Kotropoulos, “Automatic music mood classification via low-rank representation,” in Proceedings of 19th European Signal Processing Conference, pp. 689–693, August 29 - September 2, 2011, Barcelona, Spain.

[65] M. J. Yoo and I. K. Lee, “Affecticon: Emotion-based icons for music retrieval,” IEEE Computer Graphics and Applications, vol. 31, no. 3, pp. 89–95, 2011.

[66] B. Zhang and J. Lin, “An efficient content based music retrieval algorithm,” in Proceedings of International Conference on Intelligent Transportation, Big Data & Smart City, January 25-26,2018, pp. 617–620, Xiamen, China.

[67] X.Hu, “A Framework for Evaluating Multimodal Music Mood Classification,” in Journal of the Association for Information Science and Technology, vol. 68, no. 2, pp. 273–285, 2017.

[68] J. Jakubik and H. Kwasnicka, “Music emotion analysis using semantic embedding recurrent neural networks,” in Proceedings of IEEE International Conference on INnovations in Intelligent Systems and Applications, July 3-5, 2017, pp. 271–276, Gdynia, Poland.

[69] D. Turnbull, R. Liu, L. Barrington, and G. Lanckriet, “A Game-Based Approach for Collecting Semantic Annotations of Music,” in Proceedings of International Conference on Music Information Retrieval, September 23-30, 2007, pp. 535–538,Vienna, Austria.

[70] D.Chaudhary, N.P. Singh and S.Singh, “Classification of Music Signals Based on Emotion,” in Next Generation Computing Technologies"(NGCT-2018) on the theme “Computational Intelligence,” November 21-22,2018.

[71] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, “A Survey of Audio-Based Music Classification and Annotation,” IEEE Transactions on Multimedia, vol. 13, no. 2, pp. 303–319, 2011.

[72] D. McEnnis, C. McKay, I. Fujinaga, and P. Depalle, “jAudio: A feature extraction library,” Proceedings of 6th International Conference on Music Information Retrieval, London, UK, September 11-15 2005, pp. 600–603, London, UK.

[73] F. Eyben and B. Schuller, “Opensmile: the munich versatile and fast open-source audio feature extractor,”in Proceedings of the 18th ACM international conference on Multimedia, October 25–29, 2010, pp. 1459-1462, Firenze, Italy.

[74] D. Bogdanov et al., “ESSENTIA: An audio analysis library for music information retrieval,” in Proceedings of International Society for Music Information Retrieval Conference, pp. 493–498, November 4-8 2013, Curitiba, Brazil.

[75] B. Mathieu, S. Essid, T. Fillon, J. Prado, and G. Richard, “YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software,” in Proceedings of the 11th International Society for Music Information Retrieval Conference, August 9-13, 2010, pp. 441-446, Utrecht, Netherlands.

[76] E. Benetos, M. Kotti, and C. Kotropoulos, “Large scale musical instrument identification,” in Proceedings of Sound and Music Computing Conference, July 11 - 13 2007, Lefkada.

[77] Y. H. Chin, C. H. Lin, E. Siahaan, and J. C. Wang, “Happiness detection in music using hierarchical SVMs with dual types of kernels,” in Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference , October 29- November 1, 2013, Kaohsiung, Taiwan.

[78] P. Saari, T. Eerola, and O. Lartillot, “Generalizability and Simplicity as Criteria in Feature Selection: Application to Mood Classification in Music,” in IEEE Transactions on Audio, Speech and Language Processesing, vol. 19, no. 6, pp. 1802–1812, 2011.

[79] K. Markov and T. Matsui, “Music Genre and Emotion Recognition Using Gaussian Processes,” in IEEE access, vol. 2, pp. 2–3, 2013.

[80] X. Sun and Y. Tang, “Automatic Music Emotion Classification Using a New Classification Algorithm,” in Proceedings of Second International Symposium on Computational Intelligence and Design, pp. 540–542, December 12-14, 2009, Changsha, China.

[81] F. Song, Z. Guo, and D. Mei, “Feature Selection Using Principal Component Analysis,” in International Conference on System Science, Engineering Design and Manufacturing Informatization, November 12-14, 2010, pp. 27–30 Yichang, China

[82] F. Falceto and K. Gawedzki, “Chern-Simons states at genus one,” in Communications in Mathematical Physics, vol. 159, no. 3, pp. 549–579, 1994.

[83] N. Chaikla and Y. Qi, “Genetic algorithms in feature selection,” in Proceedings of International Conference on Systems, Man, and Cybernetics, October 12-15, 1999, pp. 538–540, Tokyo, Japan.

[84] T. Rückstieß, C. Osendorfer, and P. van der Smagt, “Sequential Feature Selection for Classification BT," in Proceedings of Australasian Conference on Advances in Artificial Intelligence, 5-8 December , pp. 132–141, Perth, WA, Australia

[85] M. M. Ruxanda, B. Y. Chua, A. Nanopoulos, and C. S. Jensen, “Emotion Based Music Retrieval on a Well Reduced Audio Feature Space," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, April 19-24, 2009, pp. 181–184, Taipei, Taiwan.

[86] V. Adolphs, R. & Janowski, “Emotion recognition,” in Social Neuroscience, vol. 16, no. September, pp. 252–264, 2011.

[87] H. Ahsan, V. Kumar, and C. V. Jawahar, “Multi-label annotation of music,” in Proceedings of 8th International Conference on Advances in Pattern Recognition, January 4-7, 2015, pp. 1-5, Kolkata, India.

[88] A. Sen, D. Popat, H. Shah, P. Kuwor, and E. Johri, “Music Playlist Generation using Facial Expression Analysis and Task Extraction,” in Intelligent Communication and Computational Technologies, Springer, , pp. 129–139, 2018, Singapore.

[89] J. Wang, S. Xin, and J. Li, “Emotional classification based on the tempo and mutation degrees,” in Proceedings of 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation December 23-24, 2013, pp. 444–446,Toronto, Canada.

[90] R. Sawata, T. Ogawa, and M. Haseyama, “Novel Audio Feature Projection Using KDLPCCA-Based Correlation with EEG Features for Favorite Music Classification,” IEEE Transactions on Affective Computing, vol. 3045, no. JUNE 2016, pp. 1–14, 2017.

[91] K. C. Tseng, B. S. Lin, C. M. Han, and P. S. Wang, “Emotion recognition of EEG underlying favourite music by support vector machine,” in Proceedings of 1st International Conference on Orange Technologies, March 12-16, 2013, pp. 155–158, Tainan, Taiwan.

[92] and C.-J. L. Chih-Wei Hsu, Chih-Chung Chang, “A Practical Guide to Support Vector Classification,” BJU International., vol. 101, no. 1, pp. 1396–400, 2008.

[93] F. Abdat, C. Maaoui, and A. Pruski, “Human-Computer Interaction Using Emotion Recognition from Facial Expression,” in Proceedings of 5th European Symposium on Computer Modeling and Simulation, November 16-18, 2011, pp. 196–201, Madrid, Spain.

[94] V. Rao, S. Ramakrishnan, and P. Rao, “Singing Voice Detection in North Indian Classical Music,” National Conference on Communications, **01-03 February 2008,** Indian Institute of Technology, Bombay.

[95] L. C. Wai and G. Lu, “Music emotion annotation by machine learning,” in Proceedings of IEEE 10th on Multimedia Signal Processing, pp. 580–585, October 8-10, 2008, Cairns, Qld, Australia.

[96] Y. Feng, Y. Zhuang, and Y. Pan, “Popular Music Retrieval by Detecting Mood,” in Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, July 28 - August 01, 2003, pp. 375–376, Toronto, Canada.

[97] X. Liu, Q. Chen, X. Wu, Y. Liu, and Y. Liu, “CNN Based Music Emotion Classification,” 2017.

[98] J. Lee and J. Nam, “Multi-Level and Multi-Scale Feature Aggregation Using Pretrained Convolutional Neural Networks for Music Auto-Tagging,” in IEEE Signal Processing Letters, vol. 24, no. 8, pp. 1208–1212, 2017.

[99] X. Xu, J. Deng, E. Coutinho, C. Wu, L. Zhao, and B. W. Schuller, “Connecting subspace learning and extreme learning machine in speech emotion recognition,”in IEEE Transactions on Multimedia, vol. 21, no. 3, pp. 795–808, 2019.

[100] Z. Zhang, J. Han, E. Coutinho, and B. W. Schuller, “Dynamic Difficulty Awareness Training for Continuous Emotion Prediction,” in IEEE Transactions on Multimedia, vol. 21, no. 5, pp. 1289–1301, 2018.

[101] Y. P. Lin et al., “EEG-based emotion recognition in music listening,” in IEEE Transactions on Biomedical Engineering, vol. 57, no. 7, pp. 1798–1806, 2010.

[102] Y. A. Chen, J. C. Wang, Y. H. Yang, and H. Chen, "Linear Regression-Based Adaptation of Music Emotion Recognition Models for Personalization," in IEEE International Conference on Acoustics, Speech and Signal Processing, May 4-9, 2014, pp. 2149–2153, Florence, Italy.

[103] Z. Wei, X. Li, and L. Yang, "Extraction and evaluation model for the basic characteristics of MIDI file music," in Proceedings of 26th Chinese Control and Decision Conference, pp. 2083–2087, May 31 - June 2, 2014, Changsha, China.

[104] T. Fawcett, "An Introduction to ROC Analysis," Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, 2006.

[105] D. Powers, "Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation," in Journal of Machine Learning and Technologies, vol. 2, no. 1, pp. 37–63, 2011.

[106] C. J. Willmott et al., "Statistics for the Evaluation and Comparison of Models," in Journal of Geophysical Research: Oceans, vol. 90, no. C5, pp. 8995–9005, 1985.

[107] J. J. Valero-Mas and J. M. Iñesta, "Interactive user correction of automatically detected onsets: approach and evaluation,"in Eurasip Journal of Audio, Speech, Music Processing, vol. 2017, no. 1, 2017.

[108] K. Mannepalli, P. N. Sastry, and M. Suman, "Analysis of Emotion Recognition System for Telugu Using Prosodic and Formant Features," in Speech Langage and Processing for Human-Machine Communication, Advances in Intelligent Systems and Computing, vol. 664, pp. 137–144, 2018.

[109] A. Rodà, S. Canazza, and G. De Poli, "Clustering affective qualities of classical music: Beyond the valence-arousal plane," IEEE Transactions on Affective Computing, vol. 5, no. 4, pp. 364–376, 2014.

Deepti Chaudhary is an assistant professor in the Department of Electronics and Communication Engineering at University Institute of Engineering and Technology, Kurukshetra University Kurushetra, Haryana. Presently she is pursuing Ph.D from the Department of Electronics and Communication Engineering at National Institute of Technology, Kurukshetra, Haryana, India. She received her B. Tech. degree in Electronics and Communication Engineering from Kurukshetra University, Kurukshetra in 2006, Haryana, India.
She received her M.Tech from Department of Electronics and Communication Engineering at National Institute of Technology, Kurukshetra, Haryana, in 2009. Her research interest is signal processing and audio signal processing.

Niraj Pratap Singh is an associate professor in the Department of Electronics and Communication Engineering at National Institute of Technology, Kurukshetra. He received his B.E. and M.E. degrees in Electronics and Communication Engineering from Birla Institute of Technology, Mesra, Ranchi, India in 1991 and 1994, respectively. He received Ph.D. degree in Electronics and Communication Engineering from National Institute of Technology, Kurukshetra, India. His research interests are radio resource management,
interworking architectures design; and mobility management of next generation wireless networks.

Sachin Singh is assistant professor in department of Electrical & Electronics Engg. at National Institute of Technology Delhi. He received his MTech in solid state electronics and materials from Indian Institute of Technology, Roorkee, Uttarakhand, India, in 2010. He did his PhD from the Department of Electrical Engineering, Indian Institute of Technology, Roorkee, and Uttarakhand, India in 2015. His area of interests are speech enhancement, speech recognition, digital speech processing and
biomedical instrumentation.