

AUTOMATIC MUSIC EMOTION CLASSIFICATION USING ARTIFICIAL NEURAL NETWORK BASED ON VOCAL AND INSTRUMENTAL SOUND TIMBRES

Mudiana Binti Mokhsin, Nurlaila Binti Rosli, Suzana Zambri,
Nor Diana Ahmad and Saidatul Rahah Hamidi

Department of Information Technology,
Faculty of Computer and Mathematical Sciences, UiTM Shah Alam, Selangor, Malaysia

Received 2014-03-06; Revised 2014-05-21; Accepted 2014-09-19

ABSTRACT

Detecting emotion features in a song remains as a challenge in various area of research especially in Music Emotion Classification (MEC). In order to classify selected song with certain mood or emotion, the algorithms of the machine learning must be intelligent enough to learn the data features as to match the features accordingly to the accurate emotion. Until now, there were only few studies on MEC that exploit audio timbre features from vocal part of the song incorporated with the instrumental part of a song. Timbre features is the quality of a musical features or sound that distinguishes different types of sound production in human voices and musical instruments such as string instruments, wind instruments and percussion instruments. Most of existing works in MEC are done by looking at audio, lyrics, social tags or combination of two or more classes. The question is does exploitation of both timbre features from both vocal and instrumental sound features helped in producing positive result in MEC? Thus, this research present works on detecting emotion features in Malay popular music using artificial neural network by extracting audio timbre features from both vocal and instrumental sound clips. The findings of this research will collectively improve MEC based on the manipulation of vocal and instrumental sound timbre features, as well as contributing towards the literature of music information retrieval, affective computing and psychology.

Keywords: Music Emotion Classification, Artificial Neural Network, Audio Timbres Extraction, Vocal, Instrumental Sound

1. INTRODUCTION

The world is now facing the new era of massive data also known as big data. Lots of data from various types of sources are generated, created and manipulated every single day in support of miscellaneous range of applications. Musical data too are not excluded. The popularity of the internet and the use of the great quality of music data format such as MP3, has triggered remarkable phenomenon in digital music libraries (Soriano *et al.*, 2014). Single person might have thousands of thousands of

multimedia data collections such as images, audio and video data and the numbers might rise to some extent that is hard to organize. Thus, it is essential to conduct research as to analyze the similarities among music pieces based on which music can be organized in groups and recommended to user with suitable tastes (Tzanetakis, 2014).

1.1. Music

Music has been played in almost every single day of our daily routines. Music is played during housekeeping, reading, learning, cooking, bathing and

Corresponding Author: Mudiana Binti Mokhsin, Department of Information Technology,
Faculty of Computer and Mathematical Sciences, UiTM Shah Alam, Selangor, Malaysia

the list goes on and on. The world will be in a death silent without music. According to (Levitin, 2011), music can be referred to as a super-stimulus for the perception of musicality, where “musicality” is actually a perceived property of speech that determines how “good” the music is, how strong an emotional effect the music contain and how much one enjoy listening to the music. Basically, vocal or instrumental sounds (or both) combined in such a way as to produce beauty of form, harmony and expression of emotion defines the music.

1.2. Vocal

According to Song *et al.* (2013), voice or vocal is an essential aspect of human. Vocal music normally features harmonic words called lyrics, although there are prominent examples of vocal music that are performed by means of non-linguistic syllables, sounds, or noises, occasionally as musical onomatopoeia. A short piece of vocal music with lyrics is generally termed a song. It is most likely the oldest form of music, since it does not require any instrument besides the human voice. The quality of vocal sound is differing between one individual to another. The quality of any musical sound depends on the relative power of the fundamental tone and of the overtones that accompany it. The less the fundamental tone is disturbed by overtones, the clearer and better is the voice. This is due to the perfectness of the elasticity, the relation of thickness to length, surface smoothness and other physical conditions of the cords themselves and the exactitude with which the muscles can adapt the surfaces (Levitin, 2011).

1.3. Instrumental Sound

An instrumental is a musical composition or recording without lyrics, or singing, although it might include some non-articulate vocal input; the music is primarily or exclusively produced by musical instruments.

1.4. Music and Emotion

How singing voice and instrumental sound in music can evoke human emotion? Scherer and de Vries (2013), have stated in their work on emotional response to music, that there are numbers of elements such as rhythm, pitch, intensity and timbre features can be used to define emotion in a certain music.

According to Friberg *et al.* (2014), timbre which also known as tone colour or tone quality from psychoacoustics, is the quality of a musical features or

sound that distinguishes different types of sound production in human voices and musical instruments such as string instruments, wind instruments and percussion instruments. When piano and violin are played with the same loudness and pitch, they still sound different. That is because the timbre of the instruments is different. Timbre is determined by the physical characteristics of the sound, such as the spectral envelope, the rise, duration and decay time envelope, the prefix to the sound, micro-intonation and the range between tonal and noise like character.

Generally, music can recall an appearance of sadness by a gradual downward movement, or by utilizing underlying patterns of unresolved tension, or by employing dark timbres, or thick harmonic bass textures (Bhatara *et al.*, 2014). According to Lin *et al.* (2014), the timbre controls any emotion associated with the sound.

Until recently, most of existing work on MEC is done by looking at features such as audio, lyrics, social tags or combination of two or more features as stated above (Bhat *et al.*, 2014). There were very few study on MEC that exploit features from vocal part of the song (Yang and Chen, 2012). Technically, the use of vocal part features in classifying emotion in certain music might be something novel though the processes of classification are as same as audio based MEC. It has been proved that, the timbral of the singing voice, such as aggressive, breathy, gravelly, high-pitched, or rapping is often directly related to human emotion perception and important for valence perception (Yang and Chen, 2012). Thus it is suggested that vocal timbre should be incorporated to MEC.

1.5. Related Works

Generally, MEC is part of music data mining and Artificial Intelligence (AI) area of science. Music conveys and evokes feeling. Therefore, many studies that correlate music with emotion have been done. According to Oxford dictionaries website (Music, 2013), music can be defined as vocal or instrumental sound and its common elements are pitch, rhythm, dynamics and timbre. Emotion whereas, is refer to as a strong feeling deriving from one’s circumstances, mood or relationship with others. There are 6 primary emotion suggested in work done by Ghazi *et al.* (2014), namely happiness, sadness, anger, surprise, disgust and fear.

Emotion in certain music can be classified by employing two main processes namely, signal

modelling and pattern matching. Based on work done in Gilkes *et al.*, (2012), signal modelling is referred to method of translating music audio signal into a set of musical features parameters. While, pattern matching is the process of parameter sets discovery from memory which strongly matches the parameter set obtained from the input music audio signal.

All of this process automatically carried out using AI machine classifier such as Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree and recently Baume *et al.* (2014) have exploits 47 different types of audio features and were evaluated using a five-dimensional support vector regressor, trained and tested on production music, in order to find the combination which produces the best performance for MEC. Numbers of multimedia systems have applied MEC techniques which includes subjective test, musical features extraction, machine learning algorithm. Normally, a subjective test is conducted to collect the ground truth needed for training the computational model of emotion prediction. Subjective test can be done by numbers of annotation process, where selected annotators, manually listen to certain song and classify it based on group or classes.

In most MEC, features of music such as timbre, rhythm and harmony are extracted to signify the audio parameters of a music clips. Several machine learning algorithms also applied to learn the relationship between music features and emotion labels. The most used machine learning algorithm in MEC are as follows; Artificial Neural Network (ANN), Support Vector Machines (SVM), k-Nearest Neighbour (k-NN), Bayesian.

Audio based classification for instances have been done by manipulating features from instrumental part of song and just like other MEC process, extracted features will be trained and tested as to produced classification result in determining whether the song with affective parameters are suitable to be categorized as sad, happy, angry or calm (Egger *et al.*, 2014).

The main reason for exploiting ANNs algorithm in those areas is because ANNs are very compatible as it can cater problem in various field. ANNs learn by examples from the supervised and unsupervised training just same as how human brain operates. This is because ANNs is one of the information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information.

(Anderson, 2014). Learning in biological systems involves adjustments to the synaptic connections that exist between the neurones and this is true of ANNs as well. During learning, the network is trained to correlate outputs with input patterns. When the network is used, it identifies the input pattern and tries to output the associated output pattern.

In this study, the ANN detection algorithms are implied to develop emotion classification system for Malay popular music from the year of 2000-present. The final system should be able to use musical timbre features extracted from vocal part and instrumental part of a song as well as able to classify the type of emotion in music.

2. MATERIALS AND METHODS

Generally, for every automatic MEC system, there are number of important factor that need to be considered. For example the chosen of machine learning as a classifier, audio features that need to extracted, emotions that need to be classify throughout the classification process and so on.

For a better and accurate mining result, the music data first must go through few pre-processing phase. That is to standardize audio data format into Wav format and to split the original audio data into 30 sec long. In work done by Su *et al.* (2014) it is stated that the reasons of using 30 seconds long audio data is to avoid wrong features extraction which may lead to inaccuracy of training and testing result.

Music data is collected by completing two important processes namely, subjective test and isolating process. Audio data are processed and standardized using audacity software. Matlab, NN Toolbox and MIR Toolbox have been used throughout this project to extract timbre features and to create ANN for music data training and testing. **Figure 1** below illustrates the overall MEC process for this study.

2.1. Timbres Extraction

Music features extraction is the most crucial part in this study. It involves audio features extraction which has taking place as to determine the accuracy of data generation in the database. Generally, this project only focuses on timbre features which comprises of Spectral Rolloff, Zero-Cross and Spectral Centroid. Matlab programming is used to extract all of those selected features from every part of audio data (vocal part and instrumental part).

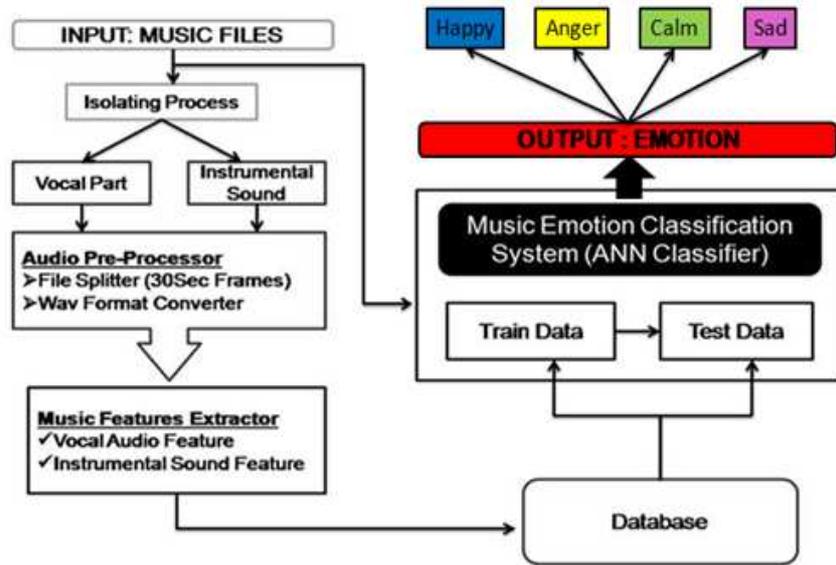


Fig. 1. Overall MEC process

2.2. Spectral Rolloff

Spectral Rolloff point is described as the Nth percentile of the power spectral distribution, where N is usually 85% or 95%. The Rolloff point is the frequency below which the N% of the magnitude distribution is determined. This evaluation is useful in distinguishing voiced speech from unvoiced. This is because the unvoiced speech has a high proportion of energy contained in the high-frequency range of the spectrum, where most of the energy for voiced speech and music is contained in lower bands. The Spectral Rolloff is a representation of the spectral shape of a sound and they are strongly correlated (Osmalsky *et al.*, 2014). From Equation 1, it's defined as the frequency where 85% of the energy in the spectrum is below that frequency. If K is the bin that fulfils, then the Spectral Rolloff frequency is $f(K)$, where $x(n)$ represents the magnitude of bin number n and $f(n)$ represents the center frequency of that bin:

$$\sum_{n=0}^k x(n) = 0.85 \sum_{n=0}^{N-1} x(n) \tag{1}$$

2.3. Spectral Centroid

Spectral centroid is a measure used in digital signal processing to exemplify a spectrum. It indicates where the “center of mass” of the spectrum is. The spectral centroid is commonly associated with the measure of the brightness of a sound (Osmalsky *et al.*, 2014).

This measure is obtained by evaluating the “center of gravity” using the Fourier transform’s frequency and magnitude information. The individual centroid of a spectral frame is described as the average frequency weighted by amplitudes, divided by the sum of the amplitudes. In practice, centroid finds this frequency for a given frame and then finds the nearest spectral bin for that frequency. The centroid is frequently a lot higher than one might instinctively anticipate, because there is so much more energy above (than below) the fundamental which contributes to the standard.

It is calculated as the weighted mean of the frequencies present in the signal, determined using a Fourier transform, with their magnitudes as the weights. Equation 2 is a formula to find the amount of spectral centroid in certain song:

$$\text{Spectral centroid} = \frac{\sum_{k=1}^N kF[k]}{\sum_{k=1}^N F[k]} \tag{2}$$

2.4. Zero-Cross

Zero Crossing rates have been proved to be useful in exemplifying divergent audio signals and have been universally used in speech and music classification problems. Zero-Crossing is a point where the sign of a function changes (e.g., from positive to negative),

represented by a Crossing of the axis (zero value) in the graph of the function. The zero-Crossing rate is the rate of sign-changes along a signal, (i.e., the rate at which the signal changes from positive to negative or back). (Ramalingam and Dhanalakshmi, 2013).

Zero-Cross is the number of times a sound signal crosses the x-axis, this accounts for noisiness in a signal and is calculated using the following Equation 3, where sign is 1 for positive arguments and 0 for negative arguments. $X[n]$ is the time domain signal for frame t :

$$Z_t = \frac{1}{2} \sum_{n=1}^N |sign(x[n]) - sign(s[n-1])| \quad (3)$$

2.5. ANN Model Created

Throughout this research, the network used is Feedback network. Feedback networks can have signals travelling in both directions by introducing loops in the network. Feedback networks are dynamic as their 'state' is changing continuously until they reach an equilibrium point. They remain at the equilibrium point until the input changes and a new equilibrium needs to be found (Rybarsch and Bornholdt, 2014)

In terms of the network layer, for audio timbre features that consist of six timbres parameters, where three of the timbre features were taken from vocal part and another three from instrumental part of song, the network therefore consist of six input nodes, one output nodes and four hidden nodes. The input nodes will be decrease into three nodes if only three audio timbre features from vocal part are used. Same goes when only three audio timbre features from instrumental part are used.

During training session, the system will calculate the error, which is defined as the square of the difference between the actual and the desired activities. The weight of each connection is changed so as to reduce the error. This training process is repeated by using different audio data of each different timbre features until the network classifies every audio data correctly. To implement this procedure, the error derivative for the weight (EW) are then calculated in order to change the weight by an amount that is proportional to the rate at which the error changes as the weight is changed.

The network is created to classify emotion in selected music by comparing happy, anger, calm and sad audio data features that have been successfully

extracted and grouped based on type of emotion. The network was built, trained and tested using the MATLAB programming language.

Figure 2 below shows the ANN architecture that represents six input nodes that consist of three timbre features from vocal part and three timbre features from instrumental part of song. Neural network toolbox in MATLAB was utilized to train the neural network. It includes several variations of the standard back propagation.

Back propagation is the algorithm that compute the error derivative for the weight (EW) by computing the rate at which the error changes as the activity level of a unit is changed or (EA). For output units, the EA is simply the difference between the actual and the desired output. To compute the EA for a hidden unit in the layer just before the output layer, all the weights between that hidden unit and the output units to which it is connected must be identified. Those weights then multiply by the EAs of those output units and add the products. This sum equals the EA for the chosen hidden unit. After calculating all the EAs in the hidden layer just before the output layer, repeat computation to the EAs for other layers, moving from layer to layer in a direction opposite to the way activities propagate through the network. This is what gives back propagation its name. Once the EA has been computed for a unit, it is straight forward to compute the EW for each incoming connection of the unit. The EW is the product of the EA and the activity through the incoming connection (Du and Swamy, 2014).

A variable learning rate that is a combination of adaptive learning rate and momentum training is used to train music clips data. 800 vocal audio data (comprises with 200 happy+200 anger+200 calm+200 sad songs) and another 800 instrumental audio data (also comprises with 200 sad+200 happy+200 calm+200 excited songs) data were used to train the ANN classifier. All training data were in the standardized audio format. Training data was obtained from various sources in the internet and Malaysia's radio station.

2.6. Neural Network Testing

Testing process in MEC take place after database comprises with musical features are generated. Music data says for example "*Muara Hati*" one of the Malay popular songs is entered to the system. Automatically system will extract musical features from that particular song before ANN classifier can classify category of emotion contained in the song.

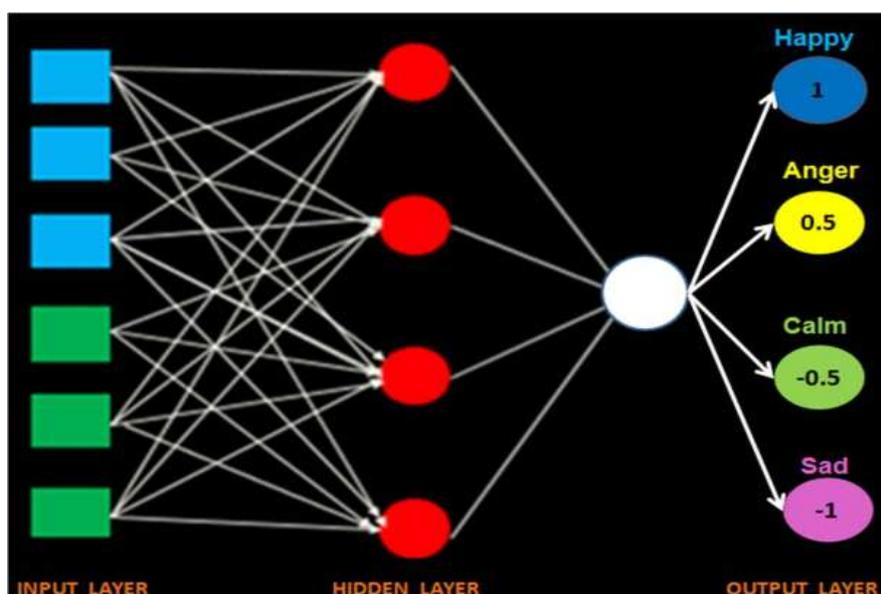


Fig. 2. ANN architecture

During the classification process, ANN classifier will get the information from the database or (memorized value of musical features) from previous training process. ANN classifier then can classify emotion from the song by scheming the music features vector as to produced result that close to 0.5 (anger), close to 1 (happy), close to -0.5 (calm) or close to -1 (sad).

Songs with output ranging from $0.5 < x \leq 1$ were considered as happy songs, songs with an output ranging from $0 < x \leq 0.5$ were considered as anger songs, songs with output ranging from $0 > x \geq -0.5$ were considered as calm songs and songs with output ranging from $-0.5 > x \geq -1$ were considered as sad songs. These tests were further verified using neural networks.

2.7. Testing Using Different Data

The testing process can be done using three different algorithms. This is to see the differences in classification rate when using different data. The details of algorithm used are as follow:

- *Neural Network+Only Vocal Timbres Data*
Audio timbre features data extracted from vocal part of the song are trained and tested using the same neural network architecture to classify emotion in selected malay pulular music.
- *Neural Network+Only Instrumental Timbres Data*
Audio timbre features data extracted from instrumental part of the song are trained and tested

using the same neural network architecture to classify emotion in selected malay pulular music.

- *Neural Network+Vocal+Instrumental Sound Timbres*
Audio timbre features data extracted from both vocal and instrumental part of the song are trained and tested using the same neural network architecture to classify emotion in selected malay pulular music.

3. RESULTS AND DISCUSSION

3.1. Timbre Extraction Results

About 800 vocal data and 800 instrumental sound data comprises of four classes of music emotion namely happy, anger, calm and sad are extracted. From the results, it is found that the average values of zero-cross features extracted from all classes of emotion are differed from each other. Same goes with spectral centroid and spectral roll off. **Table 1** below, illustrated the average value of timbre features extraction from both happy and sad music data.

The average extraction values are obtained by summing up extraction value of selected timbre and divided it to the total number of selected timbre features. Timbre features such as spectral centroid, spectral roll off and zero cross in audio clips contains repeated patterns and these repeated patterns are correlated with emotion parameters. All of the extracted timbre features are stored in the database for training and emotion detection in the ANN machine classifier.

Table 1. Average timbre extraction value

Timbre features	Average extraction value for instrumental sound data				Average extraction value for vocal data			
	Happy	Anger	Calm	Sad	Happy	Anger	Calm	Sad
Spectral centroid	3908.7	4898.2	766.5	1679.8	3288.7	3822.1	932.6	1534.6
Spectral rolloff	8916.9	9129.1	2108.0	4293.9	8966.6	9332.8	2900.8	4113.9
Zero-cross	1628.9	1776.6	432.1	555.9	1439.1	1698.1	305.3	498.4

Table 2. Test results

Music description	No. of data	%
Happy	30	100
Classified as happy	25	83
Anger	30	100
Classified as anger	24	80
Calm	30	100
Classified as calm	26	85
Sad	30	100
Classified as sad song	23	76

Table 3. Classification rate using different training data

Algorithm	% Accur.
NN + Only vocal features	71
NN+ Only instrumental sound features	78
NN +Vocal+ instrumental sound features	85

3.2. Results Classification

The accuracy of the classification result can be measured by dividing number of correctly classified songs with the total number of songs. This performance measurement is based on the evaluation taken from Music Information Retrieval Evaluation Exchange (MIREX), as done in work by Beveridge and Knox (2012). 30 songs that were categorized as happy, anger, calm and sad, were used to test the algorithm. Summary of the results is shown in **Table 2 and 3**.

Based on the results, the average accuracy of the algorithm is approximately 78%. A comparison of the accuracy of using only vocal features and the combination of vocal and instrumental sound features is shown in **Table 3**. The tests were administered using the same set of test music. Results show that the ANN detection algorithm that uses audio timbres data from both vocal and instrumental part of song are more competitive than ANN algorithm that uses only vocal or only instrumental audio timbre as training data. This finding will collectively shed light on the fundamental question of how vocal and instrumental sounds features interact with one another with regard to music emotion. Besides, this research will contribute to the literature of music information

retrieval, audio analysis, affective computing as well as music psychology.

4. CONCLUSION

Generally, there are three timbre features, which are to be exact spectral centroid, spectral rolloff and zero-cross are extracted based on its attribute in identifying Happy, Anger, Calmness and Sad (HACS) audio features. The final system is able to detect the presence of the HACS audio timbres in selected Malay popular music by evaluating the approximation of the audio parameters between the music samples and all of the timbre features extracted from vocal part and instrumental part of a song in the database.

Thus, the music classification algorithm developed is proven to be up to 75% accurate. The use of vocal and instrumental features and ANN can provide successful music emotion classification. Data from timbre extraction for both vocal and instrumental sound is used as training data to the neural network. Vocal and instrumental sound features were combined to improve testing and classification accuracy.

ANN learns to recognize emotion in music based on timbre musical texture as exist in the database. This

system is developed through learning by training all the audio data before it is tested to finally classified emotion in selected malay popular music rather than just depending on the system programming. However, ANN is still unpredictable. It may take some time to learn a sudden drastic change.

4.1. Future Works

This project has been manipulating four basic emotions as to categorize emotion in selected music which is to be precise, happy anger, calm and sad. Besides, this study only focus on extracting timbre vectors in the music data, in which previous studies have recommend that timbre can be used to strongly determined the emotion or behaviour in both vocal and instrumental sound data. For future study, it is suggested that another types of music excerpt such as pitch, energy and harmony can be used to improved musical features database for training and testing process.

As for machine classifier, this project has been using ANN detection algorithm. Though, ANN model training and testing in this study have been proved to be able to generate positive result, it is extremely recommended that if other types of machine learning techniques or approaches can be implied as part of the classification accuracy and system performance comparison in hope to improve automatic music emotion classification in the future.

5. ACKNOWLEDGEMENT

This study is supported by a grant from the RMI, Universiti Teknologi MARA, Malaysia.

5.1. Author's Contributions

All authors equally contributed in this work.

5.2. Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

6. REFERENCES

- Anderson, J.R., 2014. Rules of the Mind. 1st Edn., Psychology Press, ISBN-10: 1317782038, pp: 336.
- Baume, C., G. Fazekas, M. Barthelet, D. Marston and M. Sandler, 2014. Selection of audio features for music emotion recognition using production music. Proceedings of the 53rd International Conference: Semantic Audio, (SA' 14), London, pp: 1-3.
- Beveridge, S. and D. Knox, 2012. A feature survey for emotion classification of western popular music. Proceedings of The 9th International Symposium on Computer Music Modeling and Retrieval, (CMMR): Music and Emotions, Jun. 19-22, Queen Mary University of London.
- Bhat, A.S., V.S. Amith, N.S. Prasad and D.M. Mohan, 2014. An efficient classification algorithm for music mood detection in western and hindi music using audio feature extraction. Proceedings of the 5th International Conference on Signal and Image Processing, Jan. 8-10, IEEE Xplore Press, Jeju Island, pp: 359-364. DOI: 10.1109/ICSIP.2014.63
- Bhatara, A., P. Laukka and D.J. Levitin, 2014. Expression of emotion in music and vocal communication: Introduction to the research topic. *Front. Psychol.*, 5: 399-399.
- Du, K.L. and M.N.S. Swamy, 2014. Multilayer Perceptrons: Architecture and Error Backpropagation. In: *Neural Networks and Statistical Learning*, Du, K.L. and M.N.S. Swamy (Eds.), Springer, London, ISBN-10: 144715570X, pp: 83-126.
- Egger, S., P. Reichl and K. Schoenenberg, 2014. Quality of Experience and Interactivity. In: *Quality of Experience*, Möller, S. and A. Raake (Eds.), Springer, pp: 149-161.
- Friberg, A., E. Schoonderwaldt, A. Hedblad, M. Fabiani and A. Elowsson, 2014. Using perceptually defined music features in music information retrieval. KTH Royal Institute of Technology.
- Ghazi, D., D. Inkpen and S. Szpakowicz, 2014. Prior and contextual emotion of words in sentential context. *Comput. Speech Lang.*, 28: 76-92. DOI: 10.1016/j.csl.2013.04.009
- Gilkes, M., P. Kachare, R. Kothalikar, V. Pius and R.M. Pednekar, 2012. MFCC-based vocal emotion recognition using ANN. Proceedings of the International Conference on Electronics Engineering and Informatics, (EEI' 12), IACSIT Press, Singapore, pp: 150-154.
- Levitin, D.J., 2011. *This is Your Brain on Music: Understanding a Human Obsession*. 1st Edn., Atlantic Books Ltd., New York, ISBN-10: 0857895141, pp: 300.
- Lin, Y.P., Y.H. Yang and T.P. Jung, 2014. Fusion of electroencephalographic dynamics and musical contents for estimating emotional responses in music listening. *Front. Neuroprosthet.*, 8: 94-94. PMID: 24822035
- Music, 2013. In Oxford dictionary online.

- Osmalsky, J., M. Van Droogenbroeck and J.J. Embrechts, 2014. Performances of low-level audio classifiers for large-scale music similarity. Proceedings of the International Conference on Systems, Signals and Image Processing, May 12-15, IEEE Xplore Press, Dubrovnik, pp: 91-94.
- Ramalingam, T. and P. Dhanalakshmi, 2013. Speech/music classification using wavelet based feature extraction techniques. J. Comput. Sci., 10: 34-44. DOI: 10.3844/jcssp.2014.34.44
- Rybarsch, M. and S. Bornholdt, 2014. Avalanches in self-organized critical neural networks: A Minimal model for the neural SOC universality class. PloS One, 9: e93090-e93090. DOI: 10.1371/journal.pone.0093090
- Scherer, L.D. and M. de Vries 2013. Faith in deliberation: Examining the consequences of deliberative versus intuitive decision strategies in patient decision making. J. Behav. Decision Mak.
- Song, L., M. Li and Y. Yan, 2013. Automatic vocal segments detection in popular music. Proceedings of the 9th International Conference on Computational Intelligence and Security, Dec. 14-15, IEEE Xplore Press, Leshan, pp: 349-352. DOI: 10.1109/CIS.2013.80
- Soriano, A., F. Paulovich, L.G. Nonato and M.C.F. Oliveira, 2014. Visualization of music collections based on structural similarity. University of Sao Paulo.
- Su, L., C. Yeh, J. Liu, J. Wang and Y. Yang, 2014. A systematic evaluation of the bag-of-frames representation for music information retrieval. IEEE Trans. Multimedia, 16: 1188-1200. DOI: 10.1109/TMM.2014.2311016
- Tzanetakis, G., 2014. Music information retrieval.
- Yang, Y.H. and H.H. Chen, 2012. Machine recognition of music emotion: A review. ACM Trans. Intell. Syst. Technol., 3: 40-40. DOI: 10.1145/2168752.2168754