Dimensional Music Emotion Recognition: Combining Standard and Melodic Audio Features

R. Panda[1], B. Rocha[1] and R. P. Paiva[1],

[1] CISUC – Centre for Informatics and Systems of the University of Coimbra, Portugal
{panda, bmrocha, ruipedro}@dei.uc.pt

Abstract. We propose an approach to the dimensional music emotion recognition (MER) problem, combining both standard and melodic audio features. The dataset proposed by Yang is used, which consists of 189 audio clips. From the audio data, 458 standard features and 98 melodic features were extracted. We experimented with several supervised learning and feature selection strategies to evaluate the proposed approach. Employing only standard audio features, the best attained performance was 63.2% and 35.2% for arousal and valence prediction, respectively (R^2 statistics). Combining standard audio with melodic features, results improved to 67.4 and 40.6%, for arousal and valence, respectively. To the best of our knowledge, these are the best results attained so far with this dataset.

Keywords: music emotion recognition, machine learning, regression, standard audio features, melodic features.

1 Introduction

Current music repositories lack advanced and flexible search mechanisms, personalized to the needs of individual users. Previous research confirms the fact that "music's preeminent functions are social and psychological", and so "the most useful retrieval indexes are those that facilitate searching in conformity with such social and psychological functions. Typically, such indexes will focus on stylistic, mood, and similarity information" [1]. This is supported by studies on music information behavior that have identified emotions as an important criterion for music retrieval and organization [2].

Music Emotion Recognition (MER) research has received increased attention in recent years. Nevertheless, the field still faces many limitations and open problems, particularly on emotion detection in audio music signals. In fact, the present accuracy of current audio MER systems shows there is plenty of room for improvement. For example, in the Music Information Retrieval (MIR) Evaluation eXchange (MIREX), the highest attained classification accuracy in the Mood Classification Task was 67.8%.

Several aspects make music MER a challenging subject. First, perception of emotions evoked by a song is inherently subjective: different people often perceive different, sometimes opposite, emotions. Besides, even when listeners agree in the kind of emotion, there's still ambiguity regarding its description (e.g., the employed

terms). Additionally, it is not yet well-understood how and why music elements create specific emotional responses in listeners [3]. Another issue is the lack of standard, good quality audio emotion datasets available to compare research results. A few initiatives were created to mitigate this problem, namely MIREX annual comparisons. Still, these datasets are private, exclusively used in the contest evaluations and most studies use distinct datasets created by each author.

Our main objective in this work is to study the importance of different types of audio features in dimensional MER, namely standard audio (SA) and melodic audio (MA) features. Most previous works on MER are devoted to categorical classification, employing adjectives to represent emotions, which creates some ambiguity. From the ones devoted to continuous classification, most seem to use only standard audio features (e.g., [3], [4]). However, other audio features, such as melodic characteristics directly extracted from the audio signal have already been used successfully in other tasks such as genre identification [5].

In this work, we combine both types of audio features (standard and melodic) with machine learning techniques to classify music emotion in the dimensional plane. This strategy is motivated by recent overviews (e.g., [2], [6]) where several emotionally-relevant features are described, namely, dynamics, articulation, pitch, melody, harmony or musical form. This kind of information is often difficult to extract accurately from audio signals. Nevertheless, our working hypothesis is that melodic audio features offer an important contribution towards the extraction of emotionally-relevant features directly from audio. .

This strategy was evaluated with several machine learning techniques and the dataset of 189 audio clips created by Yang et al. [3]. The best attained results in terms of the R^2 metric were 67.4% for arousal and 40.6% for valence, using a combination of SA and MA features. These results are a clear improvement when compared to previous studies that used SA features alone [3], [4]. This shows that MA features offer a significant contribution to emotion detection.

To the best of our knowledge, this paper offers the following original contributions, which we believe are relevant to the MIR/MER community:

- the first study combining standard and melodic audio features in dimensional MER problems;
- the best results attained so far with the employed dataset.

This paper is organized as follows. In section 2, related work is described. Section 3 introduces the used dataset and the followed methodology for feature extraction and emotion classification. Next, experimental results are presented and discussed in section 4. Finally, conclusions from this study as well as future work are drawn in section 5.

2 Related Work

For long, emotions have been a major subject of study in psychology, with researchers aiming to create the best model to represent them. However, given the complexity of such task and the subjectivity inherent to emotion analysis, several

proposals have come up over the years. Different people have different perceptions of the same stimulus and often use different words to describe similar experiences.

The existing theoretical models can be divided into two different approaches: categorical and dimensional models. In categorical models, emotions are organized in different categories such as anger, fear, happiness or joy. As a result, there is no distinction between songs grouped in the same category, even if there are obvious differences in terms of how strong the evoked emotions are. On the other side, dimensional models map emotions to a plane, using several axes, with the most common approach being a two dimensional model using arousal and valence values. While the ambiguity of such models is reduced, it is still present, since for each quadrant there are several emotions. As an example, emotions such as happiness and excitation are both represented by high arousal and positive valence. To solve for this, dimensional models have been further divided into discrete – described above, and continuous. Continuous models eliminate the existing ambiguity since each point on the emotion plan denotes a different emotional state [3].

One of the most known dimensional models was proposed by Russell in 1980 [7]. It consists in a two dimensional model based on arousal and valence, splitting the plane into four distinct quadrants: Contentment, representing calm and happy music; Depression, referring to calm and anxious music; Exuberance, referring to happy and energetic; and Anxiety, representing frantic and energetic music (Figure 1). In this model, emotions are placed far from the origin, since it is where arousal and valence values are higher and therefore emotions are clearer. This model can be considered discrete, with the four quadrants used as classes, or continuous, as used in our work.

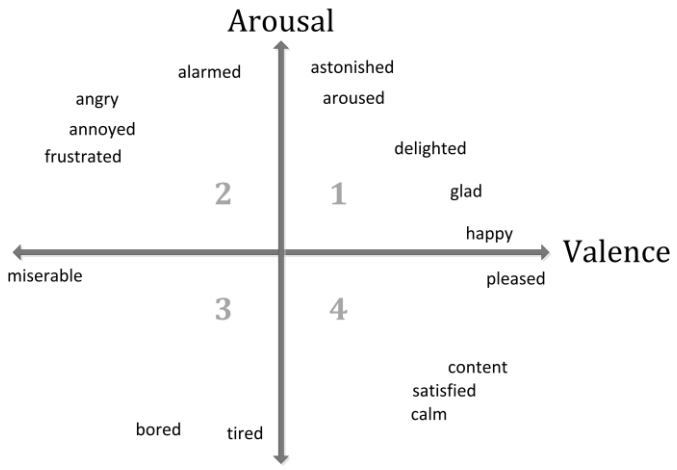

Fig. 1. Russell's model of emotion (picture adapted from [9]).

Another commonly used, two dimensional, model of emotion is Thayer's model [8]. In contrast to Russell, Thayer's theory suggests that "emotions are represented by components of two biological arousal systems, one which people find energizing, and the other which people describe as producing tension" (energetic arousal versus tense arousal).

Research on the relations between music and emotion has a long history, with initial empirical studies starting in the 19th century [10]. This problem was studied more actively in the 20th century, when several researchers investigated the relationship between emotions and particular musical attributes such as mode, harmony, tempo, rhythm and dynamics [2].

One of the first works approaching MER in audio signals was carried out by Feng et al. [11] in 2003. Two musical attributes – tempo and articulation – were extracted from 200 audio clips and used to classify music in 4 categories (happiness, sadness, anger and fear) recurring to neural networks. Feng et al. attained a precision and recall of 67% and 66% respectively, but some limitations exist in this first work. Namely, only 23 pieces were used during the test phase, as well as the low number of features and categories making it hard to provide evidence of generality. Most of the described limitations were still present in following research works (e.g., [12], [13], [14]).

Contrasting to most approaches based on categorical models, Yang et al. [3] proposed one of the first works using a continuous model. In his work, each music clip is mapped to a point in the Russell's arousal-valence (AV) plane. Several machine learning and feature selection techniques were then employed. The authors evaluated their system with recourse to R^2 statistics, having achieved 58.3% for arousal and 28.1% for valence.

Another interesting study tackling MER as a continuous problem was proposed by Korhonen et al. [15]. Employing the Russell's AV plane, the authors propose a methodology to model the emotional content of music as a function of time and musical features. To this end, system-identification techniques are used to create the models and predict AV values. Although the average R^2 is 21.9% for valence and 78.4% for arousal, it is important to note that only 6 pieces of classical music were used.

Finally, in a past work by our team [4], we used Yang's dataset and extracted features from the MIR toolbox, Marsyas and PsySound frameworks. We achieved 63% and 35.6% arousal and valence prediction accuracy, respectively. These were the best results attained so far in Yang's dataset. As will be seen, in the present study we achieved a significant improvement by employing melodic audio features.

3 Methodology

3.1 Yang Dataset

In our work we employ the dataset and AV annotations provided by Yang et al. in his work [3]. Originally the dataset used by Yang et al. contained 194 excerpts. However, five of the clips and AV annotations provided to us did not match the values available at the author's site[1] and were ignored. Thus, only 189 clips were used in our study. Each clip consists in 25 seconds of audio that better represent the emotion of the original song. These clips were selected by experts and belong to various genres,

[1] http://mpac.ee.ntu.edu.tw/~yihsuan/MER/taslp08/

mainly Pop/Rock from both western and eastern artists. The clips were selected by specialists, representing the 25 seconds that best represented the emotion content of each song, besides containing one single emotion. Each clip was later annotated with arousal and valence values ranging between -1.0 and 1.0, by at least 10 volunteers each. All clips were converted to WAV PCM format, with 22050 Hz sampling rate, 16 bits quantization and mono.

In a previous study, we have already identified some issues in this dataset [4]. Namely, the number of songs between the four quadrants of the model is not balanced, with a clear deficit in quadrant two. In addition, many clips are placed near the origin of Russell's plane. This could have been caused by a significant difference in annotations for the same songs, which could be a consequence of the high subjectivity in the emotions conveyed by those songs. According to [3], the standard deviation of the annotations was calculated to evaluate the consistency of the dataset. Almost all music samples had a standard deviation between 0.3 and 0.4 for arousal and valence, which in a scale of [-1, 1] reflects the subjectivity problems mentioned before. Although these values are not very high, they may explain the positioning of music samples in the origin, since samples with symmetric annotations (e.g., positioned in clusters 1 and 3) will result in an average AV close to zero.

3.2 Audio Feature Extraction

Several researchers have studied the hidden relations between musical attributes and emotions over the years. In a recent overview, Friberg [2] lists the following features as relevant for music and emotion: timing, dynamics, articulation, timbre, pitch, interval, melody, harmony, tonality and rhythm. Other musical characteristics commonly associated with emotion not included in that list are, for example, mode, loudness or musical form [6]. In the same study, it was found that major modes are frequently related to emotional states such as happiness or solemnity, whereas minor modes are associated with sadness or anger. In addition, simple, consonant, harmonies are usually happy, pleasant or relaxed. On the contrary, complex, dissonant, harmonies relate to emotions such as excitement, tension or sadness, as they create instability in a musical piece.

However, many of these musical attributes are usually hard to extract from audio signals or still require further study from a psychological perspective. As a result, many of the features normally used for MER were originally developed or applied in other contexts such as speech recognition and genre classification. These features usually describe audio attributes such as pitch, harmony, loudness and tempo, mostly calculated recurring to the short time spectra of the audio waveform.

Standard Audio Features.

Due to the complexity to extract meaningful musical attributes, it is common to extract standard features available in common audio frameworks. Some of those features, the so called low level features descriptors (LLD), are generally computed from the short-time spectra of the audio waveform, e.g., spectral shape features such as centroid, spread, skewness, kurtosis, slope, decrease, rolloff, flux, contrast or MFCCs. Other higher-level attributes such as tempo, tonality or key are also extracted.

In this work, three audio frameworks were used to extract features from the audio clips – PsySound, MIR Toolbox and Marsyas.

PsySound3 is a MATLAB toolbox for the analysis of sound recordings using physical and psychoacoustical algorithms. It does precise analysis using standard acoustical measurements, as well as implementations of psychoacoustical and musical models such as loudness, sharpness, roughness, fluctuation strength, pitch, rhythm and running interaural cross correlation coefficient (IACC). Since PsySound2, the framework was rewritten in a different language and the current version is unstable and lacks some important features. Due to this and since the original study by Yang used PsySound2 [3], we decided to use the same feature set containing 44 features. A set of 15 of these features are said to be particularly relevant to emotion analysis [3].

The MIR Toolbox framework is an integrated set of functions written in MATLAB, that are specific to the extraction and retrieval of musical information such as pitch, timbre, tonality and others [16]. This framework is widely used and well documented, providing extractors for a high number of both low and high-level audio features.

Marsyas (Music Analysis, Retrieval and Synthesis for Audio Signals) is a framework developed for audio processing with specific emphasis on MIR applications. Written in highly optimized C++ code, it stands out from the others due to its performance, one of the main reasons for its adoption in a variety of projects in both academia and industry. Some of its pitfalls are the complexity and the lack of some features considered relevant to MER.

A total of 458 standard audio features were extracted, 44 using PsySound, 177 with MIR Toolbox and 237 using Marsyas. Regarding the analysis window size used for frame-level features and hop size, all default options were used (512 samples for Marsyas and 0.05 seconds for MIR Toolbox). These features are then transformed in song-level features by calculating mean, variance, kurtosis and skewness. This model implicitly assumes that consecutive samples of short-time features are independent and Gaussian distributed and, furthermore, that each feature dimension is independent [17]. However it is well known, that the assumption that each feature is independent is not correct. Nevertheless, this is a commonly used feature integration method that has the advantage of compactness, a key issue to deal with the curse of dimensionality [17].

A small summary of the extracted features and their respective framework is given in Table 1.

Table 1. List of audio frameworks used for feature extraction and respective features.

Framework	Feature
Marsyas (237)	Spectral centroid, rolloff, flux, zero cross rate, linear spectral pair, linear prediction cepstral coefficients (LPCCs), spectral flatness measure (SFM), spectral crest factor (SCF), stereo panning spectrum features, MFCCs, chroma, beat histograms and tempo.
MIR Toolbox (177)	Among others: root mean square (RMS) energy, rhythmic fluctuation, tempo, attack time and slope, zero crossing rate, rolloff, flux, high frequency energy, Mel frequency cepstral coefficients (MFCCs), roughness, spectral peaks variability (irregularity), inharmonicity, pitch, mode, harmonic change and key.
PsySound2 (44)	Loudness, sharpness, volume, spectral centroid, timbral width, pitch multiplicity, dissonance, tonality and chord, based on psycho acoustic models.

Melodic Audio Features. The extraction of melodic features from audio resorts to a previous melody transcription step. To obtain a representation of the melody from polyphonic music excerpts, we employ the automatic melody extraction system proposed by Salamon et al. [18]. Figure 2 shows a visual representation of the contours output by the system for one excerpt.

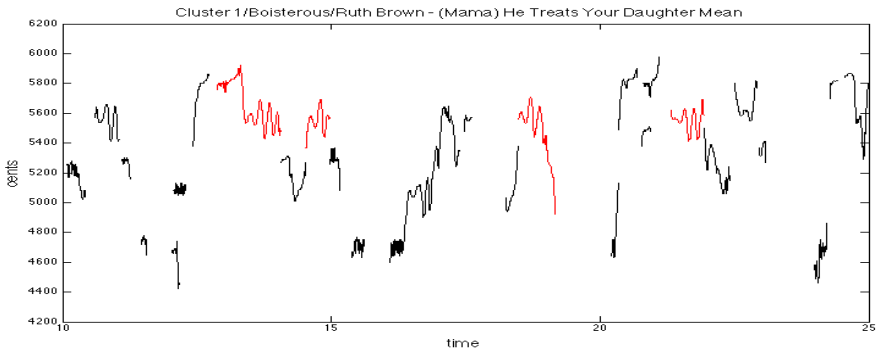

Fig. 2. Melody contours extracted from an excerpt. Red indicates the presence of vibrato.

Then, for each estimated predominant melodic pitch contour, a set of melodic features is computed. These features, explained in [19] and [5], can be divided into three categories. Then in *Global features* we show how the contour features are used to compute global per-excerpt features for use in the mood estimation.

Pitch and duration features

Three pitch features are computed: mean pitch height, pitch deviation, pitch range, and interval (the absolute difference in cents between the mean pitch height of one contour and the previous one). The duration (in seconds) is also calculated.

Vibrato features

Vibrato is a voice source characteristic of the trained singing voice. It corresponds to an almost sinusoidal modulation of the fundamental frequency [20]. When vibrato is detected in a contour, three features are extracted: vibrato rate (frequency of the variation, typical values 5-8 Hz); vibrato extent (depth of the variation, typical values 10-300 cents [21]; vibrato coverage (ratio of samples with vibrato to total number of samples in the contour).

Contour typology

Adams [22] proposed a new approach to study melodic contours based on "the product of distinctive relationships among their minimal boundaries". By categorizing the possible relationship between a segment's initial (I), final (F), highest (H) and lowest (L) pitch, 15 "contour types" are defined. We adopt Adams' melodic contour typology and compute the type of each contour.

Global features

The contour features are used to compute global excerpt features, which are used for the classification. For the pitch, duration and vibrato features we compute the mean, standard deviation, skewness and kurtosis of each feature over all contours. The contour typology is used to compute a type distribution describing the proportion of each contour type out of all the pitch contours forming the melody. In addition to these features, we also compute: the melody's highest and lowest pitches; the range between them; the ratio of contours with vibrato to all contours in the melody.

This gives us a total of 51 features. Initial experiments revealed that some features resulted in better classification if they were computed using only the longer contours in the melody. For this reason we computed for each feature (except for the interval features) a second value computed using only the top third of the melody contours when ordered by duration. This gives us a total of 98 features.

Applying these features to emotion recognition presents a few challenges. First, melody extraction is not perfect, especially when not all songs have clear melody, as is the case of this dataset. Second, these features were designed with a very different purpose in mind: to classify genre. As mentioned, emotion is highly subjective. Still, we believe melodic characteristics may give an important contribute to music emotion recognition.

3.3 Emotion Regression and Feature Selection

A wide range of supervised learning methods are available and have been used in regression problems before. The idea behind regression is to predict a real value, based on a previous set of training examples. Since the Russell's model is a continuous representation of emotion, a regression algorithm is used to train two distinct models – one for arousal and another for valence. Three different supervised machine techniques were tested: Simple Linear Regression (SLR), K-Nearest Neighbours (KNN), and Support Vector Regression (SVR). These algorithms were run using both Weka and the libSVM library using MATLAB.

In order to assess each feature's importance and improve results, while reducing the feature set size at the same time, feature selection and ranking was also performed. To this end, the RReliefF algorithm [23] and Forward Feature Selection (FFS) [24] were used. In RReliefF, the resulting feature ranking was then tested to determine the number of features providing the best results. This was done by adding one feature at a time to the set and evaluating the corresponding results. The best top-ranked features were then selected.

All experiments were validated using 10-fold cross validation with 20 repetitions, reporting the average obtained results. Moreover parameter optimization was performed, e.g., grid parameter search in the case of SVR.

In order to measure performance of the regression models, R^2 statistics were used. This metric represents the coefficient of determination, "which is the standard way for measuring the goodness of fit for regression models" [3].

4 Experimental Results

We conducted various experiments to evaluate the importance of standard audio and melodic audio features, as well as their combination in dimensional MER.

A summary of the results is presented in Table 2. The experiments were run first for SA and MA features separately, and later with the combination of both feature groups. For each column, two numbers are displayed, referring to arousal and valence prediction in terms of R^2. In addition to the results obtained with all features, results from feature selection are also presented (marked with *).

Table 2. Regression results for standard and melodic features (arousal / valence).

Classifier	SA	MA	SA+MA
SLR	42.17 / -1.30	37.45 / -7.84	42.17 / -1.29
SLR*	54.62 / 3.31	37.45 / -4.12	54.62 / 3.31
KNN	59.23 / 1.00	33.79 / -6.79	56.81 / 1.50
KNN*	62.03 / 11.97	49.99 / 0.01	61.07/ 11.97
SVR	58.19 / 14.79	45.20 / 2.72	58.03 / 16.27
SVR*	63.17 / 35.84	49.96 / 2.61	**67.39 / 40.56**

As expected, the best results were obtained with Support Vector Regression and a subset of features from both groups. These results, 67.4% for arousal and 40.6% valence are a clear improvement over the previous results obtained with SA features only: 58.3/28.1 % in [3] and 63/35.6% in a previous study by our team [4]. The standard audio features achieve better results than the melodic ones when isolated, especially for valence. Here, melodic features alone show poor performance. In fact, these features rely on melody extraction, which is not perfect, especially when not all songs have clear melody, as is the case of this dataset. However, the combination of SA and MA features improves results by around 5%. The best results were obtained with 67 SA features and 5 MA features for arousal, and 90 SA features plus 12 MA features for valence.

These results support our idea that combining both standard and melodic audio features is important for MER.

Table 3. List of the top 5 features for each feature set (rank obtained with ReliefF). Avg, std, skw and kurt stand for average, standard deviation, skewness and kurtosis, respectively.

Feature Set	Feature Name
SA (arousal)	1) Linear Spectral Pair 7 (std), 2) MFCCs 2 (kurt), 3) Key, 4) Loudness A-weighted (min), 5) Key Minor Strength (max)
SA (valence)	1) Tonality, 2) Spectral Dissonance, 3) Key Major Strength (max), 4) MFCCs 6 (skw), 5) Chord
MA (arousal)	1) Pitch Range (std), 2) Vibrato Rate (std), 3) Pitch Standard Deviation (std), 4) Higher Pitch, 5) Vibrato Rate (kurt)[2]
MA (valence)	1) Vibrato Extent (std)[1], 2) Shape Class 6[1], 3) Vibrato Extent (avg)[1], 4) Lower Pitch, 5) Lower Pitch[1]

[2] computed using only the top third lengthier contours

A list of the 5 most relevant features for each feature set is presented in Table 3. As for standard audio features, key, mode (major/minor), tonality and dissonance seem to be important. While in melodic features, some of the most relevant are related with pitch and vibrato, similarly to the results obtained in a previous study related to genre prediction [5].

5 Conclusions and Future Work

We studied the combination of standard and melodic audio features in dimensional MER. The influence of each feature to the problem was also assessed.

Regarding AV accuracy, we were able to outperform the results previously obtained by both Yang and us using only standard audio features. Additionally, we were also able to improve results by combining both sets, resulting in a new maximum of 67.4% for arousal and 40.6% for valence. Although MA features perform considerable worse than SA features, especially for valence, they were found to be relevant when working in combination.

Despite the observed improvements in results, there is still much room for improvement, especially regarding valence. To this end, we will continue researching novel audio features that best capture valence.

Acknowledgements

This work was supported by the MOODetector project (PTDC/EIA-EIA/102185/2008), financed by the Fundação para Ciência e a Tecnologia (FCT) and Programa Operacional Temático Factores de Competitividade (COMPETE) - Portugal.

References

1. Huron, D.: Perceptual and Cognitive Applications in Music Information Retrieval, International Symposium on Music Information Retrieval (2000)
2. Friberg, A.: Digital Audio Emotions - An Overview of Computer Analysis and Synthesis of Emotional Expression in Music. Proc. 11th Int. Conf. on Digital Audio Effects. pp. 1–6. , Espoo, Finland (2008)
3. Yang, Y.-H., Lin, Y.-C., Su, Y.-F., Chen, H.H.: A Regression Approach to Music Emotion Recognition. IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, pp. 448–457 (2008)
4. Panda, R., Paiva, R.P.: Automatic Creation of Mood Playlists in the Thayer Plane: A Methodology and a Comparative Study. 8th Sound and Music Computing Conference, Padova, Italy (2011)
5. Salamon, J., Rocha, B., and Gómez, E. Musical Genre Classification using Melody Features Extracted from Polyphonic Music Signals. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, (2012)

6. Meyers, O.C.: A mood-based music classification and exploration system, MSc thesis, Massachusetts Institute of Technology (2007)
7. Russell, J.A.: A circumplex model of affect. Journal of Personality and Social Psychology. 39, 1161–1178 (1980).
8. Thayer, R.E.: The Biopsychology of Mood and Arousal. Oxford University Press, USA (1989)
9. Calder, a J., Lawrence, a D., Young, a W.: Neuropsychology of fear and loathing. Nature reviews. Neuroscience. 2, 352–63 (2001)
10. Gabrielsson, A., Lindström, E.: The Influence of Musical Structure on Emotional Expression. Music and Emotion: Theory and Research. pp. 223–248. Oxford University Press (2001)
11. Feng, Y., Zhuang, Y., Pan, Y.: Popular Music Retrieval by Detecting Mood, Proc. 26th Annu. Int. ACM SIGIR Conf. on Research and Development in Information Retrieval, vol. 2, no. 2, pp. 375–376 (2003)
12. Lu, L., Liu, D., Zhang, H.-J.: Automatic Mood Detection and Tracking of Music Audio Signals, IEEE Trans. on Audio, Speech and Language Processing, vol. 14, no. 1, pp. 5-18 (2006)
13. Yang, D., Lee, W.: Disambiguating Music Emotion Using Software Agents. Proc. 5th Int. Conf. on Music Information Retrieval. pp. 52–58. , Barcelona, Spain (2004)
14. Liu, D., Lu, L.: Automatic Mood Detection from Acoustic Music Data, Int. J. on the Biology of Stress, vol. 8, no. 6, pp. 359-377 (2003)
15. Korhonen, M. D., Clausi, D. a, Jernigan, M. E.: "Modeling Emotional Content of Music Using System Identification". IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 36(3), 588–599, (2006).
16. Lartillot, O., Toiviainen, P.: A Matlab Toolbox for Musical Feature Extraction from Audio. Proc. 10th Int. Conf. on Digital Audio Effects. pp. 237–244. , Bordeaux, France (2007).
17. Meng, A., Ahrendt, P., Larsen, J., Hansen, L. K.: "Temporal Feature Integration for Music Genre Classification". IEEE Trans. on Audio, Speech and Language Processing, 15(5), pp. 275–9, (2007).
18. Salamon, J. and Gómez, E.: Melody Extraction from Polyphonic Music Signals using Pitch Contour Characteristics. IEEE Transactions on Audio, Speech and Language Processing, 20(6), pp. 1759-1770, (2012)
19. Rocha, B.: Genre Classification based on Predominant Melodic Pitch Contours. MSc thesis, Universitat Pompeu Fabra, Barcelona, Spain, (2011)
20. Sundberg, J.: The science of the singing voice. Northern Illinois University Press, Dekalb, (1987)
21. Seashore, C.: Psychology of music. Dover, New York, (1967)
22. Adams, C.: Melodic contour typology. Ethnomusicology, 20, pp. 179-215, (1976)
23. Robnik-Šikonja, M., Kononenko, I.: Theoretical and Empirical Analysis of ReliefF and RReliefF. Machine Learning, vol. 53, no 1–2, pp. 23–69 (2003)
24. Chiu, S.L.: Selecting input variables for fuzzy models. Journal of Intelligent and Fuzzy Systems, vol. 4, pp. 243–256 (1996)