Multi-Modal Music Emotion Recognition: A New Dataset, Methodology and Comparative Analysis

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

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

Abstract. We propose a multi-modal approach to the music emotion recognition (MER) problem, combining information from distinct sources, namely audio, MIDI and lyrics. We introduce a methodology for the automatic creation of a multi-modal music emotion dataset resorting to the AllMusic database, based on the emotion tags used in the MIREX Mood Classification Task. Then, MIDI files and lyrics corresponding to a sub-set of the obtained audio samples were gathered. The dataset was organized into the same 5 emotion clusters defined in MIREX. From the audio data, 177 standard features and 98 melodic features were extracted. As for MIDI, 320 features were collected. Finally, 26 lyrical features were extracted. We experimented with several supervised learning and feature selection strategies to evaluate the proposed multi-modal approach. Employing only standard audio features, the best attained performance was 44.3% (F-measure). With the multi-modal approach, results improved to 61.1%, using only 19 multi-modal features. Melodic audio features were particularly important to this improvement.

Keywords: music emotion recognition, machine learning, multi-modal analysis.

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" [8]. This is supported by studies on music information behavior that have identified emotions as an important criterion for music retrieval and organization [4].

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%.

Some of the major difficulties in MER are related to the fact that the perception of emotions evoked by a song is inherently subjective: different people often perceive different emotions when listening to the same song. Besides, even when listeners agree in the perceived emotion, there is still much ambiguity regarding its description (e.g., the adjectives employed). Additionally, it is not yet well-understood how and why music elements create specific emotional responses in listeners [30].

Another issue is the lack of standard, good quality audio emotion datasets. For this reason, most studies use distinct datasets created by each author, making it impossible to compare results. Some efforts have been developed to address this problem, namely the MIREX mood classification dataset. However, this dataset is not publicly available and exclusively used in the MIREX evaluations.

Our main goal in this work is to evaluate to what extent a multi-modal approach to MER could be effective to help break the so-called glass ceiling effect. In fact, most current approaches, based solely on standard audio features (as the one followed in the past by our team [22]), seem to have attained a glass-ceiling, which also happened in genre classification. Our working hypothesis is that employing features from different sources, namely MIDI and lyrics, as well as melodic features directly extracted from audio, might help improve current results.

Our hypothesis is motivated by recent overviews (e.g., [4], [19]) where several emotionally-relevant features are described, namely, timing, dynamics, articulation, timbre, pitch, interval, melody, harmony, tonality, rhythm, mode or musical form. Many of these features are score-oriented in nature and have been studied in the MIDI domain. However, it is often difficult to extract them accurately from audio signals, although this is the subject of active research (e.g., pitch detection [25]). Hence, we believe that combining the generally employed standard audio features with melodic audio and MIDI features can help us break the glass ceiling. Moreover, song lyrics bear significant emotional information as well [29] and, therefore, are also exploited.

To this end, we propose a new multi-modal dataset, supporting audio signals, MIDI and lyrical information for the same musical pieces, and study the importance of each in MER, as well as their combined effect. The created dataset follows the same organization as the one used in the MIREX mood classification task, i.e., 5 emotion clusters.

We evaluate our approach with several supervised learning and feature selection strategies. Among these, best results were attained with an SVM classifier: 64% F-measure in the set of 903 audio clips (using only standard and melodic audio features) and 61.1% in the multi-modal subset (193 audio clips, lyrics and midi files).

We believe this paper offers a number of relevant original contributions to the MIR/MER research community:

- a MIREX-like audio dataset (903 samples)
- a new multi-modal dataset for MER (193 audio, lyrics and midi samples);
- a methodology for automatic emotion data acquisition, resorting to the AllMusic platform;
- a multi-modal methodology for MER, combining audio, MIDI and lyrics, capable of significantly improving the results attained with standard audio features only;
- the first work employing melodic audio features in categorical MER problems.

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

2 Related Work

Emotions have long been a major subject of study in psychology, with several theoretical models proposed over the years. Such models are usually divided into two major groups: categorical and dimensional models. Categorical models consist of several categories or states of emotion, such as anger, fear, happiness or joy.

An example of the categorical paradigm is the emotion model that can be derived from the four basic emotions - anger, fear, happiness and sadness - identified by Ekman [1]. These four emotions are considered the basis from which all the other emotions are built on. From a biological perspective, this idea is manifested in the belief that there might be neurophysiological and anatomical substrates corresponding to the basic emotions. From a psychological perspective, basic emotions are often held to be the primitive building blocks of other, non-basic emotions.

Another widely known categorical model is Hevner's adjective circle [6]. Kate Hevner, best known for her research in music psychology, concluded that music and emotions are intimately connected, with music always carrying emotional meaning in it. As a result, the author proposed a grouped list of adjectives (emotions), instead of using single words. Hevner's list is composed by 67 different adjectives, organized in eight different groups in a circular way. These groups, or clusters, contain adjectives with similar meaning, used to describe the same emotional state.

In addition to these, the categorical paradigm is also employed in the MIREX Mood Classification Task, an annual comparison of state of the art MER approaches held in conjunction with the ISMIR conference. This model classifies emotions into five distinct groups or clusters, each comprising five to seven related emotions (adjectives). However, as will be discussed, the MIREX taxonomy, is not supported by psychological models.

Dimensional models, on the other hand, use several axes to map emotions into a plan. The most frequent approach uses two axes (e.g., arousal-valence (AV) or energy-stress), with some cases of a third dimension (dominance) [30]. In this paper, we follow the categorical paradigm, according to the five emotion clusters defined in MIREX.

Researchers have been studying the relations between music and emotions since at least the 19th century [5]. The problem was more actively addressed in the 20th century, when several researchers investigated the relationship between emotions and particular musical attributes such as mode, harmony, tempo, rhythm and dynamics [4].

To the best of our knowledge, the first MER paper was published in 1988 by Katayose et al. [9] There, a system for sentiment analysis based on audio features from polyphonic recordings of piano music was proposed. Music primitives such as melody, chords, key, rhythm features were used to estimate the emotion with heuristic rules.

One of the first works on MER using audio signals was conducted by Feng in 2003 [3]. Using 4 categories of emotion and only two musical attributes: tempo and articulation, Feng achieved an average precision of 67%. Some of the major limitations of this work were the very small test corpus with only 23 songs, the limited number of audio features (2) and categories (4).

From the various research works addressing emotion recognition in audio music (e.g. [12], [14], [27] and [28]), one of the first and most comprehensive using a categorical view of emotion was proposed by Lu et al. [14]. The study used the four quadrants of the Thayer's model to represent categorical emotions and intensity, timbre and rhythm features were extracted. Emotion was then detected with Gaussian Mixture Models and feature de-correlation via the Karhunen-Loeve Transform, testing hierarchical and non-hierarchical solutions. Although the algorithm reached 86.3% average precision, this value should be regarded with caution, since the system was only evaluated on a corpus of classical music.

More recently, Wang et al. [27] proposed an audio classification system using a semantic transformation of the feature vectors based on music tags and a classifier ensemble, obtaining interesting results in the MIREX 2010 mood classification task.

Some recent studies have also proposed multi-model approaches, combining different strategies for emotion detection. McVicar et al [17] proposed a bi-modal approach, combining the study of the audio and the lyrics of songs to identify common characteristics between them. This strategy is founded on the authors' assumption that "the intended mood of a song will inspire the songwriter to use certain timbres, harmony, and rhythmic features, in turn affecting the choice of lyrics as well". Using this method, the Pearson's correlation coefficient between each of the audio features and lyrics AV values were computed, finding many of the correlations to be extremely statistically significant, but below 0.2 in absolute value.

Other bi-modal work also using both audio and lyrics was presented Yang et al [29]. The authors explore the usage of lyrics, rich in semantic information, to overcome a possible emotion classification limit from using audio features only. This limit is attributed to the "semantic gap between the object feature level and human cognitive level of emotion perception" [29]. Using only four classes, the accuracy of the system went from 46.6% to 57.1%. The authors also highlight the importance of lyrics to enhance of classification accuracy of valence.

An additional study by Hu et al [7] demonstrated that, for some emotion categories, lyrics outperform audio features. In these cases, a strong and obvious semantic association between lyrical terms and categories was found.

Although few multi-model strategies have been proposed, none of the approaches we are aware of employ MIDI as well.

3 Methods

3.1 Dataset Acquisition

To create our multi-modal dataset we built on the AllMusic knowledge base, organizing it in a similar way to the MIREX Mood Classification task test bed. It contains five clusters with several emotional categories each: 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.

The MIREX taxonomy, although not supported by psychological models, is employed since this is the only base of comparison generally accepted by the music emotion recognition community. Moreover, we chose the AllMusic database because, unlike other popular databases like Last.FM, annotations are performed by professionals instead of a large community of music listeners (as happens in Last.FM). Therefore, those annotations are likely more reliable. However, the annotation process is not made public and, hence, we cannot critically analyze it.

The first step consisted in accessing automatically the AllMusic API to obtain a list of songs with the MIREX mood tags and other meta-information, such as song identifier, artists and title. To this end, a script was created to fetch existing audio samples from the same site, mostly being 30-second mp3 files.

The next step was to create the emotion annotations. To do so, the songs containing the same mood tags present in the MIREX clusters were selected. Since each song may have more than one tag, the tags of each song were grouped by cluster and the resulting song annotation was based in the most significant cluster, i.e., the one with more tags (for instance, a song with one tag from cluster 1 and three tags from cluster 5 is marked as cluster 5). A total of 903 MIREX-like audio clips, nearly balanced across clusters, were acquired: 18.8% cluster 1, 18.2% cluster 2, 23.8% cluster 3, 21.2% cluster 4 and 18.1% cluster 5.

Next, we developed tools to automatically search for lyrics and MIDI files of the same songs using the Google API. In this process, three sites were used for lyrical information (lyrics.com, ChartLyrics and MaxiLyrics), while MIDI versions were obtained from four different sites (freemidi.org, free-midi.org, midiworld.com and cool-midi.com). After removal of some deficient files, the interception of the 903 original audio clips with the lyrics and MIDIs resulted in a total of 764 lyrics and 193 MIDIs. In fact, MIDI files proved harder to acquire automatically.

As a result, we formed 3 datasets: an audio-only (AO) dataset with 903 clips, an audio-lyrics (AL) dataset with 764 audio clips and lyrics (not evaluated here) and a combined multi-modal (MM) dataset with 193 audio clips and their corresponding

lyrics and MIDIs. All datasets were nearly balanced across clusters (maximum and minimum representativity of 25 and 13%, respectively).

Even though the final MM dataset is smaller than we intended it to be (an issue we will address in the future), this approach has the benefit of exploiting the specialized human labor of the AllMusic annotations to automatically acquire a music emotion dataset. Moreover, the proposed method is sufficiently generic to be employed in the creation of different emotion datasets, with different emotion adjectives than the ones used in this article.

The created dataset can be downloaded from http://mir.dei.uc.pt/resources/MIREX-like_mood.zip.

3.2 Feature Extraction

Several authors have studied the most relevant musical attributes for emotion analysis. Namely, 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 [19]. 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 [19]. In a recent overview, Friberg [4] describes the following features: timing, dynamics, articulation, timbre, pitch, interval, melody, harmony, tonality and rhythm. Other common features not included in that list are, for example, mode, loudness or musical form [19].

As mentioned previously, many of these features have been developed in the MIDI domain and it is often difficult to extract them accurately from audio signals. Thus, we propose the combination of standard audio features with melodic audio and MIDI features, as this has the potential to improve the results. Moreover, song lyrics carry important emotional information as well and are exploited.

Standard Audio (SA) 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.

Several audio frameworks can be used to extract such audio features. In this work, audio features from Marsyas, MIR Toolbox and PsySound were used.

PsySound 3 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 IACC.

The MIR toolbox is an integrated set of functions written in MATLAB, that are specific to the extraction of musical features such as pitch, timbre, tonality and others [11]. A high number of both low and high-level audio features are available.

Marsyas (Music Analysis, Retrieval and Synthesis for Audio Signals) is a software framework developed for audio processing with specific emphasis on MIR applications. It permits the extraction of features such as tempo, MFCCs and spectral features. It is written in highly optimized C++ code, but, on the less bright side, it lacks some features considered relevant to MER.

In Marsyas, the analysis window for frame-level features was set to 512 samples. MIR toolbox was used with the default window size of 0.05 seconds. These frame-level features are integrated to song-level features by calculating their mean and 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 [18]. 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 [18].

In total, 253 features were extracted using the three frameworks.

Melodic Audio (MA) 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. [25]. Then, for each estimated predominant melodic pitch contour, a set of 98 features is computed as in [25]. These features represent melodic characteristics such as pitch range and height, vibrato rate and extent, or melodic contour shape.

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. Emotion is highly subjective and it is susceptible to variations within a song. Still, we believe melodic characteristics may influence the way we perceive emotion. In any case, melodic features extracted from the corresponding MIDI files were extracted as described below.

MIDI Features. We used toolboxes that obtain features known to be relevant according to empirical results obtained both from literature ([5] and [13]) and from our experiments [21]. We focused only on global features (local features were not considered).

Three frameworks were employed to extract MIDI features: jSymbolic [16], MIDI Toolbox [2] and jMusic [26]. The jSymbolic framework extracts 278 features (e.g., average note duration and note density), the MIDI Toolbox extracts 26 features (e.g.,

melodic complexity and key mode) and jMusic extracts 16 features (e.g., climax position and climax strength).

In total, 320 MIDI features, belonging to six musical categories matching Friberg's list (instrumentation, dynamics, rhythm, melody, texture and harmony) were extracted.

Lyrical Features. Lyrical features were extracted resorting to common lyric analysis frameworks. One of the employed frameworks, JLyrics, is implemented in Java and belongs to an open-source project from the jMIR suite [15]. This framework extracts 19 features, predominantly structural (e.g., Number of Words, Lines per Segment Average), but including also a few more semantic features (e.g., Word Profile Match Modern Blues, Word Profile Match Rap).

We also used the Synesketch framework [10], a Java API for textual emotion recognition. It uses natural language processing techniques based on WordNet [20] to extract emotions according to Paul Ekman's model [1]. The extracted features are happiness, sadness, anger, fear, disgust and surprise weight.

A total of 27 lyrical features were extracted using the two frameworks.

3.3. Classification and Feature Selection

Various tests were run in our study with the following supervised learning algorithms: Support Vector Machines (SVM), K-Nearest Neighbors, C4.5 and Naïve Bayes. To this end, both Weka (a data mining and machine learning platform) and Matlab with libSVM were used.

In addition to classification, feature selection and ranking were also performed in order to reduce the number of features and improve the results. The Relief algorithm [24] was employed to this end, resorting to the Weka workbench. The algorithm outputs a weight for each feature, based on which the ranking is determined. After feature ranking, the optimal number of features was determined experimentally by evaluating results after adding one feature at a time, according to the obtained ranking.

For both feature selection and classification, results were validated with repeated stratified 10-fold cross validation (20 repetitions), reporting the average obtained accuracy. Moreover parameter optimization was performed, e.g., grid parameter search in the case of SVM.

4 Experimental Results

Several experiments were executed to assess the importance of the various features' sources and the effect of their combination in emotion classification.

We start with experiments using standard audio (SA) features and melodic audio (MA) features, in the audio-only (AO) dataset (see Table 1). In the last column, results (F-measure) obtained from their combination are shown. The F-measure attained with the set of all features as well as after feature selection (*) is presented (see Table 4 for details of the best features used).

Table 1. Results for standard and melodic audio features (F-measure) in the audio-only (AO) dataset.

Classifier	SA	MA	SA+MA
NaïveBayes	37.0%	31.4%	38.3%
NaïveBayes*	38.0%	34.4%	44.8%
C4.5	30.1%	53.5%	55.9%
C4.5*	30.0%	56.1%	57.3%
KNN	38.9%	38.6%	41.5%
KNN*	40.8%	54.6%	46.7%
SVM	44.9%	52.3%	52.8%
SVM*	46.3%	59.1%	64.0%

As can be seen, best results were achieved with SVM classifiers and feature selection. The commonly used standard audio features lag clearly behind the melodic features (46.3% against 59.1% F-measure). However, melodic features alone are not enough. In fact, combining SA and MA features, results improve even more to 64%. Also important is that this performance was attained resorting to only 11 features (9 MA and 2 SA) from the original set of 351 SA + MA features. These results strongly support our initial hypotheses that the combination of both standard and melodic audio features is crucial in music emotion recognition problems.

Table 2. Results for separate and combined multi-modal feature sets (F-measure).

Classifier	SA	MA	MIDI	Lyrics	SA+MA	Combined
SVM	35.6%	35.0%	34.3%	30.3%	39.1	40.2%
SVM*	44.3%	55.0%	42.3%	33.7%	58.3	61.2%

Table 2 summarizes the results for the MM dataset using each feature set separately (SA, MA, MIDI and lyrics), using SVM only. Again, in the last column, results obtained from their combination are shown.

In the MM dataset, the combination of SA and MA features clearly improved the results, as before (from 44.3% using only SA to 58.3%). Again, melodic features are greatly responsible for the obtained improvement.

Comparing SA and MIDI features, we observe that their performance was similar (44.2 against 42.7%). In fact, based on a previous study by our team following the dimensional emotion paradigm [23], SA features are best for arousal prediction but lack valence estimation capability. On the other hand, MIDI features seem to improve

valence prediction but are not as good as SA for arousal estimation. Therefore, a compensation effect exploited by their combination seems to occur.

As for the combined multi-modal feature set, results improved, as we have initially hypothesized: from 58.3% using only SA and MA to 61.2%. This was attained with only 19 multi-modal features out of the 698 extracted.

In Table 3, we present the confusion matrix for the best attained results in the MM dataset. There, cluster 4 had a performance significantly under average (51.5%), with all the others attaining similar performance. This suggests cluster 4 may be more ambiguous in our dataset.

Table 3. Confusion matrix for multi-modal datasets.

	C1	C2	C3	C4	C5
C1	63.6%	15.9%	4.5%	4.5%	11.4%
C2	20.9%	60.5%	11.6%	7.0%	0.0%
C3	4.2%	18.8%	64.6%	8.3%	4.2%
C4	12.1%	18.2%	9.1%	51.5%	9.1%
C5	12.0%	3.0%	12.0%	8.0%	64.0%

As mentioned before, best results were obtained with 19 features (5 SA, 10 MA, 4 MIDI and no lyrical features – see Table 4). The observed diversity in the selected features suggests that the proposed multi-modal approach benefits music emotion classification, with particular relevance to melodic audio features, as we have hypothesized. The only exception is that no lyrical features were selected. This is confirmed by the low performance attained with the employed features (33.7%), and is certainly explained by the lack of relevant semantic features in the used lyrical frameworks. This will be addressed in the future.

Table 4 lists the 5 most important features for each source (10 for MA). As for SA, the selected features mostly pertain to harmony and tonality. Only one spectral feature was selected. Regarding MA, the 10 top features were all computed using only the top third lengthier contours. Most are related with vibrato and are similar to the ones considered important to predict genre in a previous study [25]. As for MIDI, the features on the importance of middle and bass registers were most relevant, after which came the presence of electric instruments, particularly guitar. Unlike we initially expected, articulation features (such as staccato incidence) were not selected. The reason for this is that in our dataset, these performing styles had low presence. Finally, the two most important lyrical features pertain to fear and anger weight, extracted from Synesketch, but none of them was selected.

Finally, in the MIREX 2012 Mood Classification Task we achieved 67.8% (top result) with a similar classification approach, but resorting only to standard audio features. The difference between the results attained with the MIREX dataset and the dataset proposed in this article using only SAF features (46.3%) suggests our dataset might be more challenging, although it is hard to directly compare them.

Table 4. Top 5-10 features from each feature set. Avg, std, skw and kurt stand for average, standard deviation, skewness and kurtosis, respectively.

Feature Set	Feature Name
SA	1) Harmonic Change Detection Function (avg), 2) Tonal Centroid 4 (std), 3) Key, 4) Spectral Entropy (avg), 5) Tonal Centroid 3 (std)
MA	1) Vibrato coverage (VC) (skw), 2) VC (kurt), 3) VC (avg), 4) Vibrato Extent (VE) (avg),5) VE (kurt), 6) VC (kurt), 7) Vibrato Rate (VR) (std), 8) VE (std), 9) VR (avg), 10) VE (skw)
MIDI	1) Importance of Middle Register, 2) Importance of Bass Register, 3) Electric Instrument Fraction, 4) Electric Guitar Fraction, 5) Note Prevalence of Pitched Instruments
Lyrics	1) Fear weight, 2) Anger weight, 3) Word Profile Match Modern Blues, 4) Valence, 5) Word Profile Match Rap

5 Conclusions and Future Work

We proposed a multi-modal approach to MER, based on standard audio, melodic audio, MIDI and lyrical features.

A new dataset (composed of 3 sub-sets: audio-only, audio and lyrics and audio, midi and lyrics) and an automatic acquisition strategy resorting to the AllMusic framework are proposed.

The results obtained so far suggest that the proposed multi-modal approach helps surpassing the current glass ceiling in emotion classification when only standard audio features are used.

Comparing to models created from standard audio, melodic and midi features, the performance attained by the employed lyrical features is significantly worse. This is probably a consequence of using features predominantly structural, which do not accurately capture the emotions present in song lyrics. In the future, we plan to use semantic features with stronger emotional correlation.

Finally, we plan to increase the size of our multi-modal dataset in the near future. As mentioned, MIDI files are harder to acquire automatically. Therefore, we will acquire a larger audio set from AllMusic, from which we hope to obtain a higher number of corresponding MIDI archives.

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. Ekman, P.: Emotion in the Human Face, Cambridge University Press (1982).
2. Eerola, T., Toiviainen P.: "MIR in Matlab: The Midi Toolbox," ISMIR (2004).
3. 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).
4. Friberg, A.: "Digital Audio Emotions – An Overview of Computer Analysis and Synthesis of Emotional Expression in Music," DAFx, pp.1-6 (2008).
5. Gabrielsson, A., Lindström, E.: "The influence of musical structure on emotional expression", Music and Emotion: Theory and Research, pp.223–248 (2001).
6. Hevner, K.: "Experimental Studies of the Elements of Expression in Music". American Journal of Psychology, 48(2), pp. 246–268 (1936).
7. Hu, X., Downie, J.: "When lyrics outperform audio for music mood classification: a feature analysis," ISMIR, pp. 619-624 (2010).
8. Huron, D.: "Perceptual and Cognitive Applications in Music Information Retrieval," International Symposium on Music Information Retrieval (2000).
9. Katayose, H., Imai, M., Inokuchi, S.: "Sentiment extraction in music", Proceedings 9th International Conference on Pattern Recognition pp. 1083–1087 (1988).
10. Krcadinac, U.: Textual emotion recognition and creative visualization, Graduation Thesis, University of Belgrade (2008).
11. Lartillot O., Toiviainen, P.: "A Matlab Toolbox for Musical Feature Extraction from Audio," DAFx-07, p. 237–244 (2007).
12. 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).
13. Livingstone, S., Muhlberger, R., Brown, A., Loch, A.: "Controlling musical emotionality: an affective computational architecture for influencing musical emotion," Digital Creativity 18 (2007).
14. 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).
15. McKay, C.: Automatic music classification with jMIR, Ph.D. Thesis, McGill University, Canada (2010).
16. McKay, C., Fujinaga, I.: "jSymbolic: a feature extractor for Midi files," International Computer Music Conference (2006).
17. McVicar, M., Freeman, T.: "Mining the Correlation between Lyrical and Audio Features and the Emergence of Mood," ISMIR, pp.783-788 (2011).
18. 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).
19. Meyers, O.C.: A mood-based music classification and exploration system, MSc thesis, Massachusetts Institute of Technology (2007).
20. Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.: "WordNet: An online lexical database," Int. J Lexicograph, pp. 235-244 (1990).
21. Oliveira A., Cardoso, A.: "A musical system for emotional expression", Knowledge-Based Systems 23, 901–913 (2010).
22. Panda, R., Paiva, R.P.: "Music Emotion Classification: Dataset Acquisition and Comparative Analysis," DAFx-12 (2012).
23. Panda, R., Paiva, R.P.: "Automatic Creation of Mood Playlists in the Thayer Plane: A Methodology and a Comparative Study," in 8th Sound and Music Computing Conference, (2011).

24. Robnik-Šikonja, M., Kononenko, I.: "Theoretical and Empirical Analysis of ReliefF and RReliefF," Machine Learning, vol. 53, no. 1–2, pp. 23–69 (2003).
25. Salamon, J., Rocha, B., Gómez, E.: "Musical Genre Classification Using Melody Features Extracted from Polyphonic Music Signals," ICASSP (2012).
26. Sorensen A., Brown, A.: "Introducing JMusic," Australasian Computer Music Conference, pp. 68–76 (2000).
27. Wang, J., Lo, H., Jeng, S.: "Mirex 2010: Audio Classification Using Semantic Transformation and Classifier Ensemble," WOCMAT, pp.2-5 (2010).
28. Yang, D., Lee, W.: "Disambiguating Music Emotion Using Software Agents", ISMIR, pp. 52-58 (2004).
29. Yang, Y., Lin, Y., Cheng, H., Liao, I., Ho, Y., Chen, H.: "Toward multi-modal music emotion classification," PCM08, pp. 70-79 (2008).
30. Yang, Y., Lin, Y., Su, Y., Chen, H.: "A Regression Approach to Music Emotion Recognition," IEEE Trans. on Audio, Speech, and Language Processing, vol. 16, No. 2, pp. 448-457 (2008).