

Creating Reliable Database for Experiments on Extracting Emotions from Music

Alicja Wieczorkowska¹, Piotr Synak¹, Rory Lewis², and Zbigniew Ras²

¹ Polish-Japanese Institute of Information Technology,
Koszykowa 86, 02-008 Warsaw, Poland

² University of North Carolina, Charlotte, Computer Science Dept.,
9201 University City Blvd., Charlotte, NC 28223, USA

Abstract. Emotions can be expressed in various ways. Music is one of possible media to express emotions. However, perception of music depends on many aspects and is very subjective. This paper focuses on collecting and labelling data for further experiments on discovering emotions in music audio files. The database of more than 300 songs was created and the data were labelled with adjectives. The whole collection represents 13 more detailed or 6 more general classes, covering diverse moods, feelings and emotions expressed in the gathered music pieces.

1 Introduction

It is known that listeners respond emotionally to music [12], and that music may intensify and change emotional states [9]. One can discuss if feelings experienced in relation to music are actual emotional states, since in general psychology, emotions are currently described as specific process-oriented response behaviours, i.e. directed at something (circumstance, person, etc.). Thus, musical emotions are difficult to define, and the term "emotion" in the context of music listening is actually still undefined. Moreover, the intensity of such emotion is difficult to evaluate, especially that musical context frequently misses influence of real life, inducing emotions. However, music can be experienced as frightening or threatening, even if the user has control over it and can, for instance, turn the music off.

Emotions can be characterized in appraisal and arousal components, as shown in Figure 1 [11]. Intense emotions are accompanied by increased levels of physiological arousal [8]. Music induced emotions are sometimes described as mood states, or feelings. Some elements of music, such as change of melodic line or rhythm, create tensions to a certain climax, and expectations about the future development of the music. Interruptions of expectations induce arousal. If the expectations are fulfilled, then the emotional release and relaxation upon resolution is proportional to the build-up of suspense of tension, especially for non-musician listener. Trained listener usually prefer more complex music.

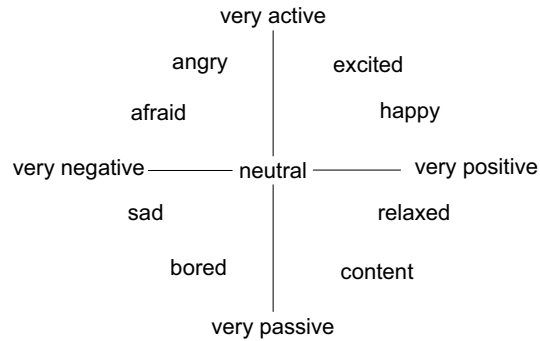


Fig. 1. Examples of emotions in arousal vs. appraisal plane. Arousal values range from very passive to very active, appraisal values range from very negative to very positive

2 Data Labelling

Data labelling with information on emotional contents of music files can be performed in various ways. One of the possibilities is to use adjectives, and if the data are grouped into classes, a set of adjectives can be used to label a single class. For instance, Hevner in [2] proposed a circle of adjective, containing 8 groups of adjectives. Her proposition was later redefined by various researchers, see for instance [5]. 8 categories of emotions may describe: fear, anger, joy, sadness, surprise, acceptance, disgust, and expectancy.

Other way of labelling data with emotions is represent emotions in 2 or 3-dimensional space. 2-dimensional space may describe amount of activation and quality, or arousal and valence (pleasure) [6],[11], as mentioned in Section 1. 3-dimensional space considers 3 categories, for instance: pleasure (evaluation), arousal, and domination (power). Arousal describes the intensity of emotion, ranging from passive to active. Pleasure describes how pleasant is the perceived feeling and it ranges from negative to positive values. Power relates to the sense of control over the emotion. Examples of emotions in 3-dimensional space can be observed in Figure 2.

In our research, we decided to use adjective-based labelling. The following basic labelling was chosen, yielding 13 classes, after Li and Ogihara [5]:

- cheerful, gay, happy,
- fanciful, light,
- delicate, graceful,
- dreamy, leisurely,
- longing, pathetic,
- dark, depressing,
- sacred, spiritual,
- dramatic, emphatic,

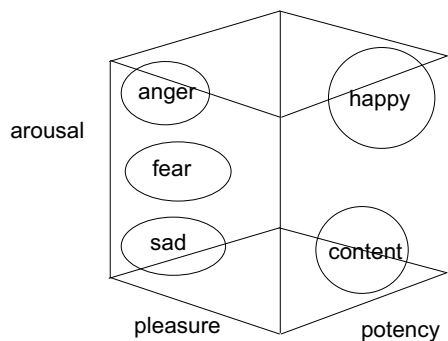


Fig. 2. Emotions represented in 3-dimensional space

- agitated, exciting,
- frustrated,
- mysterious, spooky,
- passionate,
- bluesy.

Since some of these emotions are very close, we also used more general labelling, which gathers emotions into 6 classes [5], as presented in Table 1.

Table 1. Emotions gathered into 6 classes

<i>Class Number</i>	<i>Class Name</i>	<i>Number of Objects</i>
1	<i>happy, fanciful</i>	57
2	<i>graceful, dreamy</i>	34
3	<i>pathetic, passionate</i>	49
4	<i>dramatic, agitated, frustrated</i>	117
5	<i>sacred, spooky</i>	23
6	<i>dark, bluesy</i>	23

3 Collection of Music Data

Gathering the data for such experiment is a time-consuming task, since it requires evaluating a huge amount of songs/music pieces. While collecting data, attention was paid to features of music, which are specific for a given class. Altogether, 303 songs were collected and digitally recorded - initially in

MP3 format, then converted into .snd format for parameterization purposes. For the database the entire songs were recorded.

The main problem in collecting the data was deciding how to classify the songs. The classification was made taking into account various musical features. Harmony is one of such factors. From personal experience (R. Lewis), it is known that a 9th going back to a major compared to a 5th going back to a major chord will make the difference between pathetic and dark. Pathetic is, in one view, the sense one gets when the cowboy loses his dog, wife and home when she leaves him for another (all on 7ths and 9ths for dissonance, and then we go back to a major right at the chorus and there is a sense of relief, almost light hearted relief from this gloomy picture the music has painted). In other view, pathetic is when our army starts an important battle, flags are slating and bravery is in the air (like in Wagner's Walkirie). Dark - is Mozart's Requiem - continuous, drawn out, almost never ending diminishing or going away from the major and almost never going back to the major let alone going even "more major" by augmenting.

Certain scales are known for having a more reliable guarantee to invoke emotions in human being. The most guaranteed scale to evoke emotion is the Blues scale which is a Pentatonic with accentuated flattened III and VIIth. Next comes the minor Pentatonic which is known for being "darker". The Major Pentatonic has "lighter" more "bright" sound and is typically utilized in lighter country, rock or jazz. Other scales such as Mixolydian an Ionian and so forth would diverge into other groups but are not as definitive in extracting a precise emotion from a human being.

There are Minor Pentatonics used primarily in rock and then Major Pentatonic used primarily in Country - which is sweeter. "Penta" means five, but the reason it has six notes is because we also add the lowered 5th of the scale as a passing tone making this a six note scale.

The "Blues scale" is really a slang name the Pentatonic Minor scale that accentuates its flattened III and VIIth's. For instance, when a common musician plays with a trained pianist who is not familiar with common folk slang, if the musician wanted the trained pianist to play with the band while it played a bluesy emotional song, one could simply tell the to play in C6th (notes C E G A C E) over Blues chord progression in the key of A" [root / b3 / 4th / b5th / 5th / b7 back to root) The aforementioned will make any audience from Vermont to Miami, South Africa to Hawaii feel bluesy. But why? What is in this mathematical correlation between the root wave and the flatted fifths and sevenths that guarantees this emotion of sadness?

But getting back to the Pentatonic. It is the staple jazz, Blues, country and bluegrass music. The two different Pentatonic scales are major Pentatonic R - 2 - 3 - 5 - 6 which goes great over major chords. The minor Pentatonic is R - b3 - 4 - 5 - b7 and works well for chord progressions based on minor chords. Now the b3 is where we can bend the emotions of the crowd, it separates country and "nice" music to Blues, Metal and so forth because it sounds horribly out

of place over a major chord. So, we avoid this by playing the b3 with a bend or slide into the 3rd before going to the root - that is Blues. But the twangy country sound uses the major Pentatonic and it keeps returning to the tonic note. The sound that makes the twang sound is produced by bending the second interval. When a person like Stevie Ray Vaughn, or B.B. King leans back so overcome with emotion he is really simply playing these five notes, with that b3 and sometimes the b7 and creating pleasing improvisations over this "Blues scale". Almost all the notes will sound good with almost any basic Blues tune, in a major or minor key so long as the scale is used with the same root name as the key you are playing in.

Another issue to consider when collecting the song was copyright. The copyright issue is a two part test:

1. Did the original means of obtaining the media comply with copyright law? and
2. Is it being used for personal use, or conversely for financial means and/or pier to pier use, i.e. giving away or selling the media without the owner's consent?

In our case, the original music was bought by one of the authors, R. Lewis, through CD's in the store or/and from iTunes. The authors are not selling neither giving the music to others. The authors went to great lengths with UNCC security and legal to make sure that it was password protected. Regarding length, in the past it used to be seven consecutive notes. Recently, a Federal Court in the US issued a ruling that stated that if a jury believes it was stolen off then regardless of the length.

Therefore, our data collection was prepared respecting the copyright law.

4 Features for Audio Data Description

Since the audio data itself are not useful for direct training of a classifier, parameterization of audio data is needed, possibly yielding reasonably limited feature vector. Since the research on automatic recognition of emotions in music signal has started quite recently [5], [13], there is no well established set of features for such a purpose. We decided to use features describing timbre of sound and the structure of harmony. To start with, we apply such parameterization to a signal frame of 32768 samples, for 44100 Hz sampling frequency. The frame is taken after 30 second from the beginning of the recording. The recordings are stored in MP3 format, but for parameterization purposes they are converted to .snd format. The feature vector, calculated for every song in the database, consists of the following 29 parameters [14], [15]:

- *Freq*: dominating pitch in the audio frame
- *Level*: maximal level of sound in the frame

- *Trist1, 2, 3*: Tristimulus parameters for *Freq*, calculated as [7]:

$$Trist1 = \frac{A_1^2}{\sum_{n=1}^N A_n^2} \quad Trist2 = \frac{\sum_{n=2,3,4} A_n^2}{\sum_{n=1}^N A_n^2} \quad Trist3 = \frac{\sum_{n=5}^N A_n^2}{\sum_{n=1}^N A_n^2} \quad (1)$$

where A_n - amplitude of n^{th} harmonic, N - number of harmonics available in spectrum, $M = \lfloor N/2 \rfloor$ and $L = \lfloor N/2 + 1 \rfloor$

- *EvenH* and *OddH*: Contents of even and odd harmonics in the spectrum, defined as

$$EvenH = \frac{\sqrt{\sum_{k=1}^M A_{2k}^2}}{\sqrt{\sum_{n=1}^N A_n^2}} \quad OddH = \frac{\sqrt{\sum_{k=2}^L A_{2k-1}^2}}{\sqrt{\sum_{n=1}^N A_n^2}} \quad (2)$$

- *Bright*: brightness of sound, i.e. gravity center of the spectrum, calculated as follows:

$$Bright = \frac{\sum_{n=1}^N n A_n}{\sum_{n=1}^N A_n} \quad (3)$$

- *Irreg*: irregularity of spectrum, defined as [4], [1]

$$Irreg = \log \left(20 \sum_{k=2}^{N-1} \left| \log \frac{A_k}{\sqrt[3]{A_{k-1} A_k A_{k+1}}} \right| \right) \quad (4)$$

- *Freq1, Ratio1, ..., 9*: for these parameters, 10 most prominent peaks in the spectrum are found. The lowest frequency within this set is chosen as *Freq1*, and proportions of other frequencies to the lowest one are denoted as *Ratio1, ..., 9*
- *Ampl1, Ratio1, ..., 9*: the amplitude of *Freq1* in decibel scale, and differences in decibels between peaks corresponding to *Ratio1, ..., 9* and *Ampl1*.

5 Usefulness of the Data Set: Classification Experiments

We decided to check usefulness of the obtained data set in experiments with automatic classification of emotions in music. K-NN (k nearest neighbors) algorithm was chosen for these tests. In k-NN the class for a tested sample is assigned on the basis of the distances between the vector of parameters for this sample and the majority of k nearest vectors representing known samples. CV-5 standard cross-validation was applied in tests, i.e. 20% of data were removed from the set for training and afterwards used for testing; such an experiment was repeated 5 times. In order to compare results with Li and Ogihara [5], experiments were performed for each class separately, i.e. in each classification experiment, a single class was detected versus all

other classes. The correctness ranged from 62.67% for class no. 4 (dramatic, agitated and frustrated) to 92.33% for classes 5 (sacred, spooky) and 6 (dark, bluesy). Therefore, although our database still needs enlargement, it initially proved its usefulness in these simple experiments.

6 Summary

Although emotions induced by music may depend on cultural background and other contexts, there are still feelings commonly shared by all listeners. Thus, it is possible to label music data with adjectives corresponding to various emotions. The main topic of this paper was preparing a labelled database of music pieces for research on automatic extraction emotions from music.

Gathering of data is not only a time-consuming task. It also requires finding the reasonable number of music pieces representing all classes chosen, i.e. emotions. Labelling is more challenging, and in our research it was performed by a musician (R. Lewis). However, one can always discuss whether other subjects would perceive the same emotions for these same music examples. We plan to continue our experiments, expanding the data set and labelling it by more subjects.

The data we collected contain a few dozens of examples for each of 13 classes, labelled with adjectives. One, two, or three adjectives were used for each class, since such labelling may be more informative for some subjects. The final collection consists of more than 300 pieces (whole songs or other pieces). These audio data were parameterized, and feature vectors calculated for each piece constitute a database, used next in experiments on automatic classification of emotions using k-NN algorithm. Therefore, our work yielded a measurable outcomes thus proving its usefulness.

References

1. Fujinaga, I., McMillan, K. (2000) Realtime recognition of orchestral instruments. Proceedings of the International Computer Music Conference, 141–143
2. Hevner, K. (1936) Experimental studies of the elements of expression in music. American Journal of Psychology **48**, 246–268
3. Jackson, W. H. (1998) Cross-Cultural Perception and Structure of Music. On-line, available at <http://internet.cybermesa.com/~bjackson/Papers/xcmusic.htm>
4. Kostek, B, Wiczorkowska, A. (1997) Parametric Representation Of Musical Sounds. Archives of Acoustics **22**, **1**, 3–26
5. Li, T., Ogihara, M. (2003) Detecting emotion in music, in *4th International Conference on Music Information Retrieval ISMIR 2003*, Washington, D.C., and Baltimore, Maryland. Available at <http://ismir2003.ismir.net/papers/Li.PDF>
6. Marasek, K. (2004) Private communication

7. Pollard, H. F., Jansson, E. V. (1982) A Tristimulus Method for the Specification of Musical Timbre. *Acustica* **51**, 162–171
8. Rickard, N. S. (2004) Intense emotional responses to music: a test of the physiological arousal hypothesis. *Psychology of Music* **32**, (4), 371–388. Available at <http://pom.sagepub.com/cgi/reprint/32/4/371>
9. Sloboda, J. (1996) Music and the Emotions. British Association Festival of Science, The Psychology of Music
10. Smith, H., Ike, S. (2004) Are Emotions Cross-Culturally Intersubjective? A Japanese Test. 21 Century COE "Cultural and Ecological Foundations of the Mind", Hokkaido University. The Internet: <http://lynx.let.hokudai.ac.jp/COE21/>
11. Tato, R., Santos, R., Kompe, R., and Pardo, J. M. (2002) Emotional Space Improves Emotion Recognition. *7th International Conference on Spoken Language Processing ICSLP 2002*, Denver, Colorado, available at <http://lorien.die.upm.es/partners/sony/ICSLP2002.PDF>
12. Vink, A. (2001) Music and Emotion. Living apart together: a relationship between music psychology and music therapy. *Nordic Journal of Music Therapy*, **10(2)**, 144–158
13. Wiczorkowska, A. (2004) Towards Extracting Emotions from Music. International Workshop on Intelligent Media Technology for Communicative Intelligence, Warsaw, Poland, PJIIT - Publishing House, 181–183
14. Wiczorkowska, A., Wroblewski, J., Slezak, D., and Synak, P. (2003) Application of temporal descriptors to musical instrument sound recognition. *Journal of Intelligent Information Systems* **21(1)**, Kluwer, 71–93
15. Wiczorkowska, A., Synak, P., Lewis, R., and Ras, Z. (2005) Extracting Emotions from Music Data. 15th International Symposium on Methodologies for Intelligent Systems ISMIS 2005, Saratoga Springs, NY, USA