

Detecting Emotions in Classical Music from MIDI Files

Jacek Grekow¹ and Zbigniew W. Ras²

¹ Bialystok Technical University, Faculty of Computer Science, Wiejska 45A,
Bialystok 15-351, Poland
grekowj@wi.pb.edu.pl

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

Abstract. At a time when the quantity of sounds surrounding us is rapidly increasing and the access to different recordings as well as the amount of music files available on the Internet is constantly growing, the problem of building music recommendation systems including systems which can automatically detect emotions contained in music files is of great importance. In this article, a new strategy for emotion detection in classical music pieces which are in MIDI format is presented. A hierarchical model of emotions consisting of two levels, L1 and L2, is used. A collection of harmonic and rhythmic attributes extracted from music files allowed for emotion detection with an average of 83% accuracy at level L1.

Keywords: Music Information Retrieval, Emotion Detection.

1 Introduction

Music has accompanied man for ages in various situations. We hear it in advertisements, in films, at parties, at the philharmonic, in clubs, etc. One of the most important functions of music is its effect on man. Certain pieces of music have a relaxing effect, while others stimulate us to act, and some cause a change in or emphasize our mood. Music is not only a great number of sounds arranged by a composer, it is also the emotion contained within these sounds. At a time when the quantity of sounds surrounding us is rapidly increasing and the access to different recordings as well as the amount of music files available on the Internet is constantly growing, the problem of building music recommendation systems including systems which can automatically detect emotions contained in music files is of great importance.

1.1 Input Data

Many research papers deal with the problem of emotion detection. Some of them rely on audio files [5], [6], [7], [11], [12], [14] and others on MIDI files [1], [8]. In our research, we concentrated on emotion detection in MIDI files containing symbolic

representation of music (key, structure, chords, instrument). The means of representation of music content in MIDI files is much closer to the description which is used by musicians, composers, and musicologists. To describe music, they use key, tempo, scale, sounds, etc. This way, we avoid the difficult stage of extraction of separate notes, tracks, instruments from audio files, and we can concentrate on the deciding element which is the music content.

1.2 Mood Model

There are several models describing emotions contained in music. One of them is the model proposed by Hevner [4]. This model is made up of a list of adjectives grouped in 8 main categories. After modification it was used by Li et al. [5] and Wieczorkowska et al. [12]. This model is quite developed and complex, too complicated to use in our experiment, however, it illustrates the intricacy of describing emotions.

Another model is the two-dimensional Thayer model [9] in which the main elements are Stress and Energy laid out on 2 perpendicular axes. Stress can change from happy to anxious, and Energy varies from calm to energetic. This way, 4 main categories form on the plain: Exuberance, Anxious, Depression and Contentment. This model was used by Liu et al. [7], DiPaola et al. [1], Wang et al. [11], Yang et al. [14].

The model we chose is based on Thayer's model (Fig. 1). Following the example of this model, we created a hierarchical model of emotions consisting of two levels, L1 and L2.

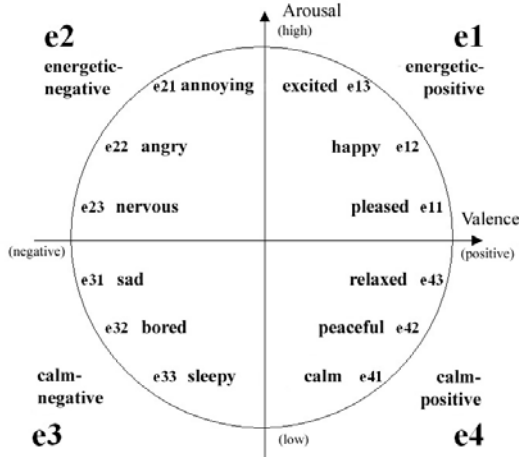


Fig. 1. Thayer's arousal-valence emotion plane

The first level L1 contains 4 emotions. To ease the indexing of files, group names were replaced with compound adjectives referencing Arousal and Valence. Our mood model contains the following groups (Table 1):

Table 1. Description of mood groups in L1, the first level

Abbreviation	Description
e1	energetic-positive
e2	energetic-negative
e3	calm-negative
e4	calm-positive

In the first group (e1), pieces of music can be found which convey positive emotions and have a quite rapid tempo, are happy and arousing (Excited, Happy, Pleased). In the second group (e2), the tempo of the pieces is fast, but the emotions are more negative, expressing Annoying, Angry, Nervous. In the third group (e3), are pieces that have a negative energy and are slow, expressing Sad, Bored, Sleepy. In the last group (e4), are pieces that are calm and positive and express Calm, Peaceful, Relaxed.

The second level is related to the first, and is made up of 12 sub-emotions, 3 emotions for each emotion contained in the first level (Table 2).

Table 2. Description of mood groups in L2, the second level

Abbreviation	Description	Abbreviation	Description
e11	pleased	e31	sad
e12	happy	e32	bored
e13	excited	e33	sleepy
e21	annoying	e41	calm
e22	angry	e42	peaceful
e23	nervous	e43	relaxed

2 Method

2.1 Database

A database with 83 MIDI files of classical music (F. Chopin, R. Shuman, F. Schubert, E. Grieg, F. Mendelssohn-Bartholdy, etc.) was created specifically for the needs of the experiment. Starting from the 5th bar, 16 second segments were isolated from each piece. The shift forward was chosen with the aim of avoiding various, unstable introductions at the beginning of many pieces. Each 16 second segment was divided into 6 subsegments of 6 seconds each with a mutual overlap (overlapping 2/3). There were 498 resulting 6-second subsegments. Overlapping allows for precise tracking of emotion contained within musical segments.

2.2 Indexing

The 498 subsegments were annotated with an emotion by a listener-tester, a person with a formal music education/background, who has professional experience in listening to music.

2.3 Feature Extraction

The next stage was to obtain features describing the files in the database. Specially written software “AKWET simulator - Features Explorer” was used in connection with the program MATLAB, every record in the database was described with 63 features.

Harmony Features




Harmony, along with rhythm and dynamics, is one of the main elements of music upon which emotion in music is dependent. Harmony Features reflect dissonance and consonance of harmony of sounds. They are based on previous work by the author [2], [3]. To calculate the harmony parameters, we used the frequency ratio of simultaneously occurring sounds (Table 3).

A given consonance (interval, chord, polyphone) comprises of simultaneously resonant sounds, the frequency ratio of which can be noted as following:

$$N_{R1}: N_{R2}: \dots : N_{Rk} \tag{1}$$

where k is the number of sounds comprising the consonance.

Table 3. Example consonance sound frequency ratios

k	Musical notation	Consonance sound frequency ratios $N_{R1}: N_{R2}: \dots : N_{Rk}$
2		2:3
3		4:5:6
4		25:30:36:45

From the frequency ratios, we calculated the AkD parameter, which mirrors the degree of dissonance in a single chord. The higher its value, the more dissonant is the consonance; when the AkD value is lower, the consonance is more consonant – more pleasant for the ear.

$$AkD = LCM (N_{R1}, N_{R2}, \dots, N_{Rk}) \tag{2}$$

where k is the number of sounds in a given sample. In the case when $k = 1$, then $AkD = 1$. LCM means Least Common Multiple.

From the sequence of consonance samples collected from a musical segment (Fig. 2), the table can be defined as:

$$AkD_s = (AkD_1, AkD_2, \dots, AkD_p) \tag{3}$$

where p is the number of samples collected from a given segment.

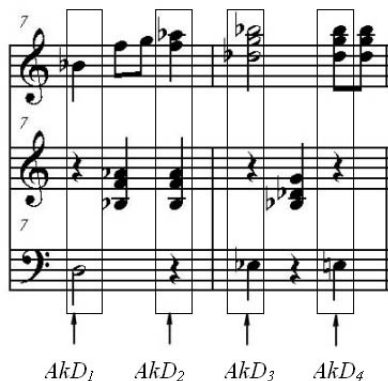


Fig. 2. Process of sample collection from a segment

The moments of sample collection from a segment have been defined according to two criteria. The first is the collection of samples at every eighth, and the second is the collection of samples at every new chord in a segment.

Table 4. Main harmony features

Feature group	Main features
Basic statistical functions	Average AkD_s Standard deviation of AkD_s Number of samples in AkD_s First max in AkD_s Second max in AkD_s Third max in AkD_s
Common values	First common value in AkD_s Second common value in AkD_s Third common value in AkD_s
Chord location	Average amplitude of sound in a chord Standard deviation of sound in a chord

Harmony features describe what kind of harmony occurs in a given segment, which ones dominate, how many of them occur, etc. (Table 4). Below is a presentation of AkD samples (chords) collected at every eighth in a segment from Étude Op.10 No 5 by F. Chopin (Fig. 3).

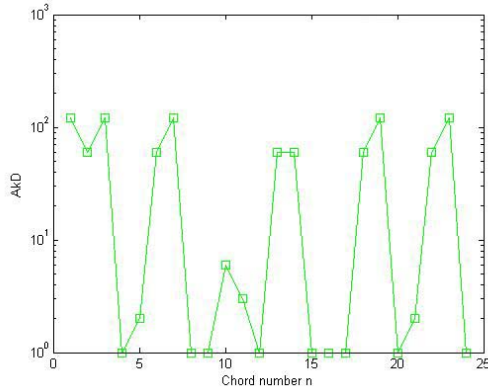


Fig. 3. *AkD* for fragment of F. Chopin Etude Op.10 No 5

Rhythmic Features

Rhythmic features represent rhythmic regularity in a given segment of music. These features were obtained from the beat histogram, which was acquired from the calculation of autocorrelation [10].

$$autocorrelation[lag] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]x[n-lag] \tag{4}$$

where *n* is the input sample index (in MIDI ticks), *N* is total number of MIDI ticks in segment and *lag* is delay in MIDI ticks ($0 < lag < N$). The value of *x[n]* is proportional to the velocity of Note On events.

The histogram was transformed so that each bin corresponded to a periodicity unit of beats per minute (Fig. 4).

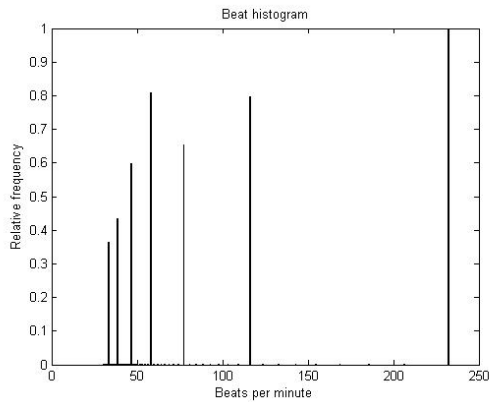


Fig. 4. Beat histogram for fragment of F. Chopin Etude Op.10 No 5

Rhythmic features describe the strongest pulses in the piece, relations between them, their quantity, etc. (Table 5).

Table 5. Main rhythmic features

Feature group	Main features
Strongest Rhythmic Pulses	First Strongest Rhythmic Pulse Second Strongest Rhythmic Pulse Third Strongest Rhythmic Pulse
Pulse Ratios	Ratios of Strongest Pulses
Relatively Strong Pulses	Number of Relatively Strong Pulses Number of beat peaks with a value of 30-50% of relative frequency
Rhythmic Note values	Average Note Duration - Average duration of notes in seconds Note Density - Average number of notes per second

Correlations between Features

Individual features, such as harmony or dynamics, are related to rhythm. They are often correlated. The moment of appearance of a given accent, chord, etc. in the bar is of great significance. The most important and significant parameters were obtained through the correlation of parameters with rhythm.

We created an AkD_B data table. It comprises of AkD samples collected from musical segments at moments of the strongest pulses.

$$AkD_B = (AkD_1, AkD_2, \dots, AkD_b) \tag{5}$$

where b is the number of collected samples at moments of the strongest pulses. All values from the beat histogram which are more than 50% of the strongest (Strongest Rhythmic Pulse) in a beat histogram were accepted as the strongest pulses. Next, statistical features were calculated, similarly as with AkD_s (Table 4).

Dynamic Features

Dynamic features are based on the intensity of sound, the length of sounds, and their development in a segment (Table 6).

Table 6. Main dynamic features

Feature group	Main features
Basic statistical functions	Average of loudness levels of all notes Standard deviation of loudness levels of all notes

The last stage consisted of exporting of the obtained data to Arff format, allowing for data analysis in the WEKA program.

2.4 Mood Detection

Describing emotions contained within a given segment is not always clear-cut. Some segments contain a single emotion, while others can contain several emotions

simultaneously. In order to allow the tester to mark his opinion, the choice of many emotions for a segment was permitted. This allowed the tester to assign not only a single emotion but several to the consecutive examples. Marking an emotion from the lower level, L2, automatically caused the marking of the appropriate emotion from the higher level, L1.

3 The Experiment Results

The program WEKA was used to carry out the experiments, which allowed for testing data utilizing many methods [13].

Because many musical segments were labeled by many labels simultaneously, multi-label classification in emotion detection was used (multi-label decision attribute was replaced by a set of binary decision attributes representing emotions). The same, we transformed data into several two-class types of data and tested one against the rest of the data. For each class, a data set was generated containing a copy of each instance in the original data, but with a modified class value. If the instance had the class associated with the corresponding dataset it was tagged YES, otherwise, it was tagged NO. The classifiers were built for each of these binary data sets. The proposed strategy greatly simplified the process of building classifiers for a decision system with a multi-label decision attribute.

The classification results were calculated using a cross validation evaluation CV-10. We used attribute selection to find the best subset of attributes. The best result was achieved by using Wrapper Subset Evaluator. After testing the data utilizing many methods, one of the best results was achieved with the use of the k-NN classifier (k-nearest neighbors). The use of attribute selection improved the accuracy of classifiers by an average of about 10% (Fig. 5).

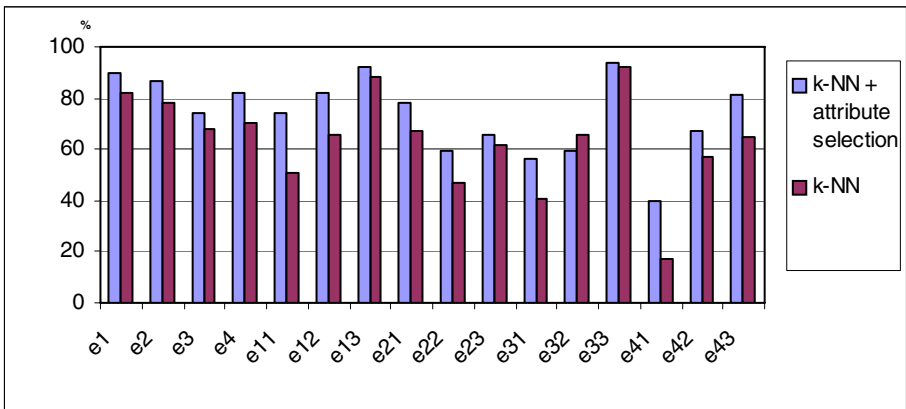


Fig. 5. Comparison of results attained using the k-NN algorithm with and without the use of attribute selection

A classifier was created for each emotion separately. L1 first level classifiers (Table 7) are more accurate than L2 second level classifiers. This is connected with the fact that the groups of examples from the first level are larger as well as the

Table 7. Coverage factor of L1 first level classifiers

Classifier	Emotion	No. of objects	Coverage factor
e1	energetic-positive	151	90%
e2	energetic-negative	172	87%
e3	calm-negative	111	74%
e4	calm-positive	103	82%

Table 8. Coverage factor of L2 second level classifiers

Classifier	Emotion	No. of objects	Coverage factor
e11	pleased	66	74%
e12	happy	69	82%
e13	excited	19	92%
e21	annoying	37	78%
e22	angry	52	59%
e23	nervous	82	66%
e31	sad	47	56%
e32	bored	52	59%
e33	sleepy	12	94%
e41	calm	17	40%
e42	peaceful	30	67%
e43	relaxed	56	81%

emotions are much easier to recognize for the listener. The most accuracy was attained for emotion classifier e1 – energetic-positive (90%), and the least accuracy was attained for emotion classifier e3 – calm-negative (74%).

The accuracy of L2 second level classifiers (Table 8) is somewhat less accurate than L1 first level classifiers, and fluctuates from 40-92%. This is connected with the fact that the example groups for specific emotions are smaller as well as that the recognition of these emotions – on this more precise level – is more difficult for the listener. The least accuracy was attained for emotion classifier e41 – calm. Also, e42 is not high, which is tied to the fact that the division of emotions into groups e41 – calm and e42 – peaceful is not the most apt. These are rather difficult for the listener to distinguish. In the future, for further research, these two groups should be combined into one. The best results (80-90%) were obtained for emotions e12 (happy), e13 (excited), e33 (sleepy), and e43 (relaxed). These are the most easily recognized emotions by the listener, and it is rather difficult to confuse them with other emotions.

4 Conclusion

In this article, we presented emotion detection in pieces of classical music in the form of MIDI files. A hierarchical model of emotions consisting of two levels, L1 and L2, is used. A collection of harmonic and rhythmic attributes extracted from music files allowed for emotion detection with an average of 83% accuracy at level L1.

We plan to find emotional profiles of different users using the music file search system, which searches for files according to emotion. This will be achieved through testing on a larger group and through grouping them according to their responses. This should resolve the problem of subjective emotion assessment by different users. We also plan to expand the collection of attributes as well as enhance the file database by adding other genres of music.

Acknowledgments. This paper is supported by the S/WI/5/08.

References

1. DiPaola, S., Arya, A.: Emotional Remapping of Music to Facial Animation. In: ACM Siggraph 2006 Video Game Symposium Proceedings, Boston (2006)
2. Grekow, J.: An analysis of the harmonic content – main parameters in the AKWET method. In: II Konferencja Technologii Eksploracji i Reprezentacji Wiedzy, TERW 2007, Hołny Mejera (2007)
3. Grekow, J.: Broadening musical perception by AKWEDs technique visualization. In: Proceedings of the 9th International Conference on Music Perception and Cognition, ICMPC9 (2006)
4. Hevner, K.: Experimental studies of the elements of expression in music. *American Journal of Psychology* 48, 246–268 (1936)
5. Li, T., Ogihara, M.: Detecting emotion in music. In: Proceedings of the Fifth International Symposium on Music Information Retrieval (ISMIR 2003), pp. 239–240 (2003)
6. Liu, C., Yang, Y., Wu, P., Chen, H.: Detecting and Classifying Emotion in Popular Music. In: Proceedings of the 9th Joint Conference on Information Sciences (JCIS)/CVPRIP (2006)
7. Liu, D., Lu, L., Zhang, N.: Automatic mood detection from acoustic music data. In: ISMIR (2003)
8. McKay, C., Fujinaga, I.: Automatic genre classification using large high-level musical feature sets. In: Proceedings of the International Conference on Music Information Retrieval, pp. 525–530 (2004)
9. Thayer, R.E.: *The biopsychology arousal*. Oxford University Press, Oxford (1989)
10. Tzanetakis, G., Cook, P.: Musical Genre Classification of Audio Signals. *IEEE Transactions on Speech and Audio Processing* 10(5) (2002)
11. Wang, M., Zhang, N., Zhu, H.: User-adaptive music emotion recognition. In: 7th International Conference on Signal Processing, ICSP (2004)
12. Wiczorkowska, A., Synak, P., Ras, Z.: Multi-label classification of emotions in music. In: Intelligent Information Processing and Web Mining, Advances in Soft Computing, Proceedings of IIS 2006 Symposium, Ustron, Poland, vol. 35, pp. 307–315. Springer, Heidelberg (2006)
13. Witten, I.H., Frank, E.: *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, San Francisco (2005)
14. Yang, Y., Su, Y., Lin, Y., Chen, H.: Music Emotion Recognition: The Role of Individuality. In: Proceedings of the international workshop on Human-centered multimedia ACM MM/HCM 2007 (2007)