This is a pre-copyedited, author-produced version of an article accepted for publication in Lecture Notes on Data Engineering and Communications Technologies following peer review. The final authenticated version is available online at https://doi.org/10.1007/978-3-030-70665-4_103

Music Emotion Recognition based on Chord Identification

Fan Zhang[1], Hongying Meng[1], Maozhen Li[1], Ruixia Cui[2], and Chang Liu[2]

[1]Department of ECE, Brunel University London, London, UK
chanvan@sina.com, Hongying.Meng@brunel.ac.uk, Maozhen.Li@brunel.ac.uk
[2]First Affiliated Hospital, Xi'an Jiaotong University, Xi'an, China
raerui1202@163.com, liuchangdoctor@163.com

Abstract. As one of the most classic human inventions, music can be seen as another language, used to express the author's thoughts and emotion. Music Emotion Recognition (MER) is an interesting research topic in artificial intelligence field for recognising emotions from music. The recognition methods and tools for music signals are growing fast recently. However the attentions are paid mainly on music signal feature extraction and machine learning methods. The theoretical knowledge of music on emotion is usually ignored. This paper propose a new method through important music units – chord to recognize emotion from music. We firstly build a new chord database with chord segments based on selected emotions. Then FFT and statistics (STAT) features are extracted from the music signals in the database. Then one-shot learning are used for chord identification and emotion recognition using Euclidean distance and correlation. The proposed method can recognise 6 emotions from music signals through chord identification with 76.6% accuracy of chord identification. This idea can be developed further using more advanced feature extraction and one-shot learning methods.

Keywords: music emotion recognition, chord, music theory

1 Introduction

With the fast increase in the number of music recordings, it is a challenging task to find a high quality recording with a specific emotion requirement in the music market. Music Emotion Recognition (MER) [3] aims to build a system that can accurately and efficiently identify the emotion embedded inside a music using Artificial Intelligence (AI) methods.

In current MER research, most of the works focus on feature extractions from the frequency domain, time domain and continuity of music signals, most of the signal's own attribute information has been basically mined. However, music as an artificially created sound signal, the information it displays must be derived from the thoughts and emotions that the creator wants to express. Just as human speaking and writing, music as a tool for transmitting emotional information must contain its own as a "language" in the law or structure to help it show the emotions that it wants to express. In the structure of music, chords are known as an important element in turning a dull monosyllabic signal into a rich and colorful piece of music. The chords, like the adjectives in the declarative sentence, set up a display standing on a basic melody skeleton to show the emotions the author wants to express.

The word "chords" are from Greek, the original meaning is the string. In music theory, it refers to the sound of two or more different pitches combined together. In European classical music and the music style influenced by it, more often refers to a combination of three or more pitches, and the combination of two pitches is described by a pitch. In the basic form of chords, the lowest sound is called the "root". The remaining sounds are named according to their pitch relationship with the root note.

This paper aims to develop a new MER system to predict the emotion information of the music through identify its chords as there is strong correlation between chords and emotion.

2 Related Works

In the development history of MER, researchers focused on different areas of a MER system using different AI technologies. Earliest, researchers spent more time in feature selection. Busso et al. [1] used fundamental frequency feature to build a system to recognize emotion before 2010. Markov et al. [5] used the more advanced feature Zero-crossing Rate (ZCR) and Mel-frequency cepstral coefficients (MFCCs) to recognize emotions.

When machine learning was getting popular all over the world, MER researchers moved their attention on it. Meng et al. [6] successfully recognized the emotion with KNN machine learning method for any audio signals. Two year later, Song and Dixon [8] built a MER system with Support Vector Machine (SVM) and Random Forest methods.

Recent years, deep learning methods are shining in the AI field. In 2018, Liu et al. [4] recognized the emotion from 1000 songs through a Convolutional Neural Network (CNN) model. Pellegrini et al. [7] chose Recurrent Neural Network (RNN) model to produce Arousal-Valence emotion prediction from music.

Only few works mentioned MER combination with music theory. Cheng et al. [2] extracted the longest common chord sub-sequence and the chord histogram as the Mid-level features according to the musical knowledge. Other one is that Zhou et al. [11] used Restricted Boltzmann Machines (RBMs) and Gibbs sampling to deal with the chord detection.

However, for all of these related researches, nobody focus on the relation between chords and emotions directly. In this paper, based on the book "Music and Emotions" [9], a new MER system is proposed. The proposed system will directly map music and emotion based on the therotical results between emotion and chords [9] and detect the chords for emotion recognition as there is strong correlation between chords and emotion. To our knowledge, this is the first MER system based on "Music and Emotions" theory.

3 Methodology

In our system, based on "Theory of Musical Equilibration" in the book "Music and Emotions" which designed and conducted by Daniela and Bernd Willimek [9], the different kinds of chords are the key for the emotion recognition system. Six different chords are selected for this research as they have clear emotion information related to them. The 6 selected chords are described and illustrated in Table 1 with their related emotion information. In the book, the example data are selected from the piano music teaching audio by using Adobe Audititon. The music segments have 6 kinds of chords. Each chord corresponds one emotion.

In our system, the selected music signals are split to single chord segments. One example chord will be chosen for the training with its emotion label. It is kind of one-shot learning. After that, other chord segments can be input to the system for identification and emotion recognition. In our method, feature extraction methods such as FFT and statistics (STAT) features are used from the music segments firstly. Then, 1 Nearest Neighbour method based on correlation and Euclidean distance are used for identification and emotion recognition.

Table 1. Six selected chords and their definition

Chord	Description	Emotion
C Minor Chord	The root and third notes are major thirds,and the third and fifth notes are small thirds.	Sad
Minor Sixth Chord	The minor sixth chord, the first inversion of the third chord, the chord is com-posed of three to five major three degrees, five to the root of the pure four degrees.	Fear
Natural Minor Chord	A natural minor chord is a diatonic chord that is built by starting on the sixth degree of its relative major chord.	Despair
Neapolitan Sixth Chord	The Neapolitan sixth chord is generally labeled N6.	Amazement
Major subdominant Chord	The subdominant is the fifth pitch of the scale, which is called the subdominant because it is one level lower than the sub-tonic.	Relax
Augmented Chord	Augmented chord is a chord that is superimposed by three tones plus major third degree intervals.The chord name is represented by the root name as augmented chord.	Exciting

3.1 Data Preparation

Based on the study of the book "Music and Emotions", the mapping relationship between chords and emotions is understood as the criterion for emotional classification. A database with enough samples, accurate chord classification, and enough chord type has been built for the experiments.

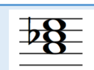
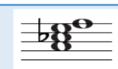
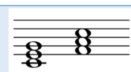
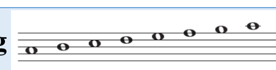
C Minor Chord	**Minor Sixth Chord**
Sad 	**Fear** 
Neapolitan Sixth Chord	**Augmented Chord**
Despair 	**Amazement** 
Subdominant Chord	**Natural Minor Chord**
Relax 	**Exciting** 

Fig. 1. Six kinds of chords and their corresponding emotions.

The six chosen chords are extracted, classified and labeled based on the theory in the book "Music and Emotions" from the piano playing audio by Adobe Audition which is a professional audio processing software. The piano audio was selected according to the professional chord teaching audio. So they would be precise. These six chords correspond to six emotions of amazement, sadness, fear, excitement, despair and relaxation. The database emotions and chords classification is shown in Figure 1 Based on these six chords, the data are divided into 10 different chord segments for each emotion under 78357 average sampling points per segment. A sample database of 66 different chords in 6 class of selected emotions has been extracted from piano audio. After completing the classification and labeling of the six chords, select the most accurate set of 6 different chords as the benchmark chord and other 60 chords for one shot learning.

3.2 Chord Identification

We choose a training data from each chord as the benchmark data. Then, the rest are samples for testing in our one-shot training. When the data has samples and the benchmark, the goal of the study shifts to how the actual labeled benchmark data to match whether the sample data in the database matches. Because chords are a set of one-dimensional audio data, the problem to be solved is to choose what method to find a similar one to the benchmark label in a bunch of audio data. There are many approaches to compare one-dimensional data, the most common of which is the Euclidean distance detection method. In mathematics, the Euclidean distance is the distance between two points in Euclidean space, which can be found in Equation 1.

On the other hand, not every vector has a similar spatial position, so another set of detection methods is needed to match it, thus reducing the error of detection. This method is correlation detection. The correlation coefficient is the first statistical indicator designed by the statistician Carl Pearson and is the amount of linear correlation between the variables studied, usually expressed by the letter r. Due to the different research objects, there are many approaches to define the correlation coefficient. The Pearson correlation coefficient is more commonly used, can be found in Equation 2.

$$d\left(x,y\right)=\sqrt{\sum_{i=1}^{n}\left(x_i-y_i\right)^2} \tag{1}$$

$$r=\frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}} \tag{2}$$

3.3 MER System

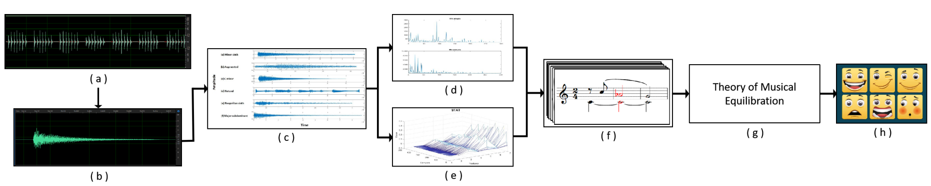


Fig. 2. Chord identification system (a. The piano music; b. The chord segment; c. The chord database d.The FFT features; e. The STAT features; f. The identified chords; h. Theory of Musical Equilibration; f.Emotions).

After successful building the database, the system for chord identification emotion recognition can be set up. The full system is illustrated in Figure 2.

At the beginning of the system building, for analyze the shape, distribution and features of the six different chords, six benchmark labels are simulated for observation. Because the recognition method based on the Theory of Musical Equilibration is first time to be used. Two more basic features and methods were decided to be applied. The spectrum for the chords which based By Discrete Fourier transform (DFT) have been got. It can be easily found from the spectrum, the specific chord have the specific frequency feature. By observing the spectrum, it is found that the frequency characteristics of the audio data can better show the difference between the different chords.

So the FFT of the target music signal was extracted. AT the same time, the statistics feature (STAT) [10] has been extracted for containing more information on frequency domain.

Finally, two mentioned identification methods – Euclidean distance and Correlation have been used to identify the correct chord variety for mapping the emotions from these chords.

4 Experiments

The experiments are started with feature extraction. FFT have been done before interpolated 200 points in chord feature. Then 1000 points features have been extracted from the both chord and samples.

In the next step, the FFT of each benchmark data is calculated.The identification result can be found in Table 2 that the matching results of the three chords of Augmented, C Minor and Minor 6th are much better than the other three.

Table 2. FFT Euclidean distance and correlation confusion matrix.(True: Grand truth; Pred: Prediction)

	Euclidean Distance						Correlation					
Pred \ True	Aug	m	m6	Nat	N6	Sub-d	Aug	m	m6	Nat	N6	Sub-d
Aug	8	0	0	0	2	1	7	0	0	0	3	0
m	0	10	1	0	1	3	2	10	2	0	0	2
m6	0	0	7	0	0	0	0	0	8	0	0	0
Nat	0	0	1	3	2	1	0	0	0	5	0	0
N6	0	0	1	7	5	3	0	0	0	4	6	6
Sub-d	2	0	0	0	0	2	1	0	0	1	1	2

Then STAT feature are extracted and the identification results can be seen from Table 3 that in the case of using the STAT feature selection, the matching accuracy of the latter three chords is greatly improved, but the matching accuracy of the Augmented chord is completely destroyed. This may be due to the principle of extracting STAT features, resulting in the loss of information in many Augmented chords.

The confusion matrix in Table 3 illustrate more details about this results. All the kinds of chord shown the nice identification accuracy except the Augmented one. Especially the C Minor chord and the Subdominant chord shown the perfect results.

Table 3. STAT Euclidean distance correlation confusion matrix.(True: Grand truth; Pred: Prediction)

	Euclidean distance						Correlation					
Pred \ True	Aug	m	m6	Nat	N6	Sub-d	Aug	m	m6	Nat	N6	Sub-d
Aug	2	0	0	0	0	0	2	0	0	0	0	0
m	0	10	0	0	0	0	2	10	3	0	0	0
m6	8	0	9	0	0	0	6	0	7	1	0	0
Nat	0	0	1	8	3	0	0	0	0	9	4	0
N6	0	0	0	2	7	0	0	0	0	0	6	0
Sub-d	0	0	0	0	0	10	0	0	0	0	0	10

At the same time, it can be seen that the results of the correlation method are basically similar to the results of the Euclidean distance. It may be better than the Euclidean distance method in the matching classification of some chords, but it is basically approximate accuracy.

It also can been found in Table 3, the confusion matrix of The Correlation by Pearson correlation coefficient comparison test with STAT features in Chord database.The result of identification accuracy is similar as the Euclidean distance by STAT features.The Augmented Chord identification is terrible but others are effective. Finally, all the chords extracted from the same music signal will be given the label based on majority voting. Then use the Theory of Musical Equilibration, the emotion in this music can be recognized. The results of the identification accuracy is illustrated in Table 4. It can be found that the accuracy of STAT feature reaches 76.6% by using Euclidean distance. However the FFT feature only achieves 63.3% with correlation.

Table 4. The identification accuracy (%) of both STAT and FFT chord feature.

Chords Feature	Euclidean Distance	Correlation	Average
STAT	76.6	73.3	74.9
FFT	58.3	63.3	60.8

In summery, it can be seen from the result that correlation works better on STAT feature, but has opposite effect on FFT. The major subdominant Chord is easier to be identified by STAT feature, which means that the emotion “relax” has better recognition rate when using STAT feature in both Euclidean distance and correlation methods. Exciting is better using FFT feature to recognise. However, Amazement is difficult to recognize by both methods.

5 Conclusion

We propose a new approach to recognise emotions from music signals. The chord theory – Theory of Musical Equilibration is used to successfully building a MER system. The experiments shows that the STAT features work well in most of chords except the Augmented chord. On the other hands, the FFT features provide the best results in Augmented chord, Minor chord and Minor 6th chord. However, it didn't work very well in other three chords. The final result shown that the STAT features achieves 76.6% accuracy which is 13.3% higher than FFT feature. The system used very basic feature extraction and simplest 1-NN classifier. The performance can be improved if better features and identification methods are used.

References

1. Carlos Busso, Sungbok Lee, and Shrikanth Narayanan. Analysis of emotionally salient aspects of fundamental frequency for emotion detection. *IEEE transactions on audio, speech, and language processing*, 17(4):582–596, 2009.
2. Heng-Tze Cheng, Yi-Hsuan Yang, Yu-Ching Lin, I-Bin Liao, Homer H Chen, et al. Automatic chord recognition for music classification and retrieval. In *2008 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1505–1508. IEEE, 2008.
3. Youngmoo E Kim, Erik M Schmidt, Raymond Migneco, Brandon G Morton, Patrick Richardson, Jeffrey Scott, Jacquelin A Speck, and Douglas Turnbull. Music emotion recognition: A state of the art review. In *Proc. ISMIR*, pages 255–266. Citeseer, 2010.
4. Tong Liu, Li Han, Liangkai Ma, and Dongwei Guo. Audio-based deep music emotion recognition. In *AIP Conference Proceedings*, volume 1967, page 040021. AIP Publishing, 2018.
5. Konstantin Markov, Motofumi Iwata, and Tomoko Matsui. Music emotion recognition using gaussian processes. In *Working Notes Proceedings of the MediaEval 2013 Workshop, Barcelona, Spain, October 18-19, CEUR-WS. org, ISSN 1613-0073*, 2013.
6. Hongying Meng and Nadia Bianchi-Berthouze. Naturalistic affective expression classification by a multi-stage approach based on hidden markov models. In *Affective computing and intelligent interaction*, pages 378–387. Springer, 2011.
7. Thomas Pellegrini and Valentin Barrière. Time-continuous estimation of emotion in music with recurrent neural networks. In *MediaEval 2015 Multimedia Benchmark Workshop (MediaEval 2015)*, pages pp–1, 2015.
8. Yading Song and D Simon. How well can a music emotion recognition system predict the emotional responses of participants? In *Sound and Music Computing Conference (SMC)*, pages 387–392, 2015.
9. Bernd Willimek and Daniela Willimek. *Music and Emotions-Research on the Theory of Musical Equilibration (die Strebetendenz-Theorie)*. Bernd Willimek, 2013.
10. Fan Zhang, Hongying Meng, and Maozhen Li. Emotion extraction and recognition from music. In *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pages 1728–1733. IEEE, 2016.
11. Xinquan Zhou and Alexander Lerch. Chord detection using deep learning. In *Proceedings of the 16th ISMIR Conference*, volume 53, 2015.