Emotion of Music: Extraction and Composing

Yizhi Yuan*

School of Computer Science and Information Engineering, Shanghai Institute of Technology, Shanghai, China

* Corresponding Author Email: zhangqingjin@haiwainet.cn

Abstract. Music, the third art that has accompanied the development of human civilization, has played a vital role in human expression of emotions from ancient times to the present. It has played a vital role in human expression of emotions from ancient times to the present. This study describes three different ways of musical emotional extraction within the current research field: a. Extraction by measuring the physiological characteristics of the subject; b. text feature extraction by analyzing lyrics in songs and c. extraction by analyzing audio features. This paper also discusses the research results of three different groups of researchers to use specific emotions for music creation by designing and applying the GA based on KTH rule system, the mLSMN with logistic regression, and the MAgentM framework. The purpose of this research is to provide reference materials for subsequent researchers through the detailed introduction of the MER method, and to provide industry stakeholders with an outlook for the music industry.

Keywords: MER; Emotional composition; Music; Emotion extraction.

1. Introduction

Since ancient times, music has greatly influenced consciousness and power which could impress people's souls. Although people's emotions have been expressed through music since its origin, the study of musical emotions has only recently become popular. Since the 15th century, western musicians have consciously aimed to create works by aiming the purpose of appealing to musical emotions. Therefore, a theory of emotions in music, known as "The Doctrine of the Affections" was developed by German composer Johann Mattheson, based on Descartes' description of the physiological processes that support and determine passion [1]. The concept of music emotion is usually based on the concept of emotion in a broad sense and specifically refers to the subjective emotional response induced by the listener under the action of music stimulation, specifically including subjective feeling, brain activation, psychophysiology, emotional expression, action tendency and emotion regulation [2].

However, the form of current music style has changed greatly compared to the past, and with the booming development of technology, musicians can use different technical means to create music, especially in the field of electronic music. Therefore, the recognition of musical emotions by computers came into sight. There are now many methods that can support the computer to complete the emotional recognition operation of music, and all of them can be generally divided into three steps: the division of emotional space, the extraction and classification of emotional features in music [3]. With the procedures given above, it creates many different algorithms for researchers to complete the analysis work of musical emotion and then compose with it. For example, using Thayer model in BP neural network to measure the eight dimensions of each note in a piece of music with given functions, and the calculation results are highly similar to the musical experts' work. In addition, for music with lyrics, there is also a study of analyzing the mood of the current section through lyrics. Its mechanism is to create a huge vocabulary database by establishing the connection between PAD (pleasure, arousal and dominance) values and words, then computer could determine which mood the current section represents by analyzing its lyrics [4].

The MER (Music Emotion Recognition) method could play a big role in composing, too. As it known to all, music is the carrier and source of the creator's emotions, which means clearer emotional expression reinforces people's emotional empathy for the music they listen to. Software "MusicSculptor", e.g., is a good way for people to compose music. Users can set the features of the

music they want to create, by selecting the desired music style on the interface of the software, and then it will generate music corresponding to the emotional style through Sieve's algorithm [5].

Nowadays, there are many examples of computer composition, and the intention of this article is to conduct a series of discussions on musical emotion recognition and its applications, so as to make the composition of computer music more attractive to listeners. This study will introduce the basic concepts of musical emotion and expand to how to identify musical emotion and practical application cases of musical emotion in composition. In Section 5, this article will discuss the limitation of current studies and possible directions for future work. Finally, conclusions are included in Section 6.

2. Basic Description of Emotion in Music

It's well-known that music can express emotions from different angles. In terms of general cognition, the emotions contained in music exist as both objective and subjective emotions. On this basis, it can be divided into two categories: one is the emotion that the music itself wants to express, and the other is the result of the listener superimposing his own emotions with the emotions expressed by the music [6]. In the first case, according to the various characteristics of the music including pitch, waveform, choice of instruments, and even the lyrics in the song, etc., it will be judged what kind of feelings the composer wants to express to the audience through the piece, and the expression of this emotion is objective and fundamental for the listener's perspective. However, the emotions that a piece of music can express will also have different interpretations based on the subjective feelings of the audience, and the listener will relate to his current subjective experience when enjoying the music, which makes the emotions that the current music wants to express have a richer level, i.e., the second situation-subjective musical emotions.

There are the basic types of emotions that a piece of music wants to express based on the first case. The basic emotions of human beings are divided into joy, anger, sorrow, and happiness, Russell create a two-dimensional emotion plane based on these four emotions [5], as shown in Fig. 1. The X-axis represents the positive and negative of the current emotional energy, which is the valence of the emotion; The Y axis represents the arousal of emotions, it represents how energetic the music is.

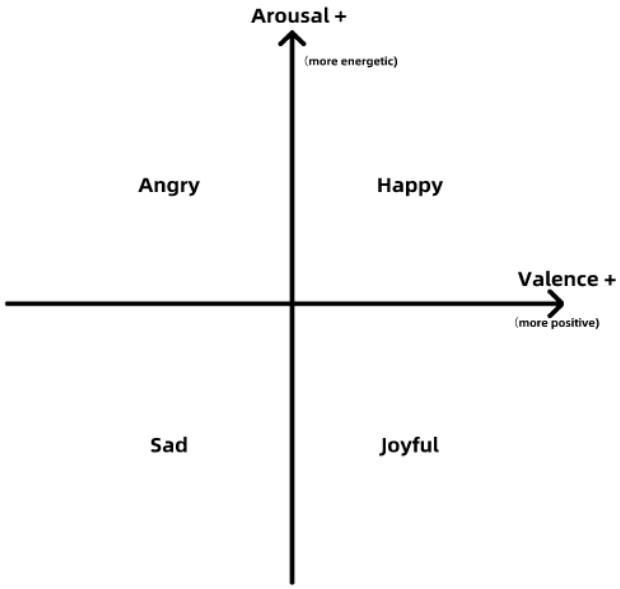

Fig. 1 2D Emotion Plane created by Huang and Lin.

The second case is a bit more complicated. Today's music is not only presented to people as a work of art, it also has many functions. This has to mention what the purpose of modern people listening to music is. According to research by Juslin and Laukka, after surveying 141 people about their basic motivations for listening to music (the statistics are not mutually exclusive), nearly 47% of respondents indicated that their purpose was to "release their emotions", 33% said they listened to

music to "gain peace of mind", and reasons such as "can make me recall a certain memory" and "make me work more efficiently" are also included in the consequences [7]. These results can be a good indication that the emotions that music brings to listeners cannot be roughly summarized into one of the four basic emotions mentioned above. In fact, based on containing basic emotions, music can induce people to stimulate more and more complex emotions. As presented in Fig. 2, it illustrates the relationship between common emotions in real life, emotions usually expressed by music, and emotions usually induced by the music within the full range of human emotions.

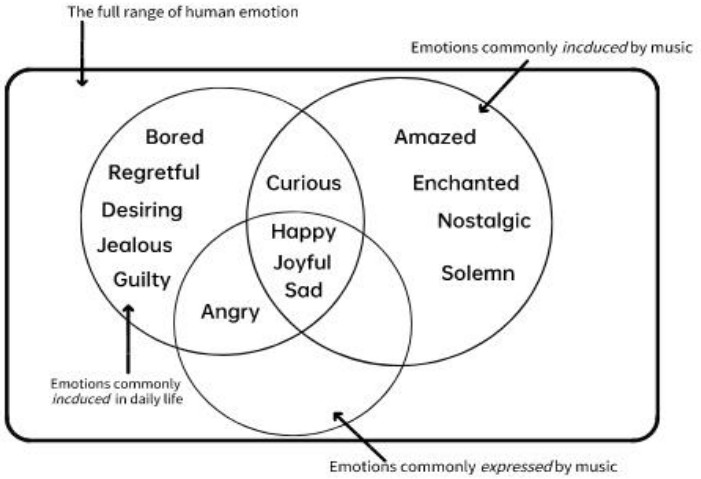

Fig. 2 Hypothesized relationships between three emotions mentioned above.

3. Emotion Extraction

Nowadays, there are many ways to extract emotions from music, and this article will briefly introduce three different methods of music emotion extraction: (1) a method of identifying the emotions evoked by the music currently playing by detecting the physiological state of the listener; (2) a method of emotional extraction by analyzing the features of the given music and (3) A method of recognizing the emotions of music with lyrical features.

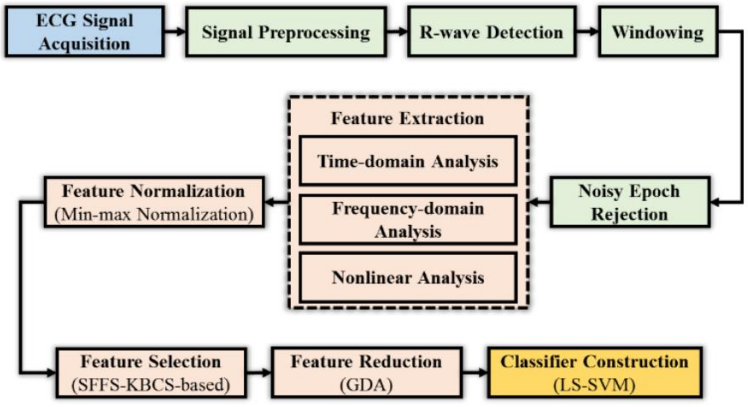

Fig. 3 Diagram of ECG-based emotion recognition algorithm depicted by Hsu et al.

3.1. Method with Physiological Detection

As mentioned earlier in the introduction, the subjective emotions generated by the listeners when listening to music are also a kind of musical emotions. The study of Hsu et al. is based on the detection of the subject's ECG signal to analyze the emotional changes that occur when the subjects listen to music. Fig. 3 shows the block diagram depicted by researchers, it demonstrates each procedure of ECG-based emotion recognition algorithm as follows: the first step is to preprocess the obtained ECG signals, then the researchers use R-wave detection to prepare for the upcoming extraction of ECG-

based features. Subsequently, the researchers use windowing to divide the entire signal into minute-by-minute clip and denoise the ECG signals. After extracting several features from the time-, frequency-domain and nonlinear analysis of ECG signals, the researchers would use the minimum-maximum normalization method to finish feature normalization, then use algorithms based on SFFS-KBCS (sequential forward floating selection-kernel-based class separability-based) to help in selecting appropriate features. Hsu et al. also do GDA (generalized discriminant) to do the final processing of the data, and finally use the LS-SVM (least squares support vector machine) of the one-against-one strategy for emotion recognition [8].

3.2. Method with Audio Features

Relying on the extracted audio features to analyze the emotions expressed by current music is one of the most mainstream music emotion recognition methods. The method introduced in this paper is a musical emotion recognition system with deep Gaussian processing as the core mechanism studied by Chen et al. [9]. The system is divided into two parts: feature extraction and classification. Researchers use computers to extract the audio features (rhythm, dynamics, timbre, pitch and tonality) of the testing songs, compare and match them with the features of training songs that have been classified by emotion type in existing databases (built by music clips from Internet), and achieve the purpose of music emotion recognition. The detailed steps are depicted in Fig 4.

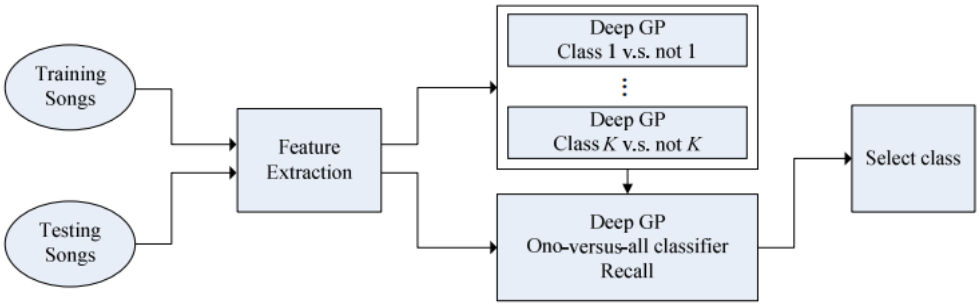

Fig. 4 MER System architecture based on deep Gaussian process.

3.3. Method with Lyrical Features

The third method is to extract musical emotions based on the lyrical features of the lyrics of the testing song. In fact, in most existing studies, this approach cannot be considered as the sole benchmark for music emotion recognition, it often combined with the audio features of the music for researchers to analyze. However, this is only a combination made from the requirement of emotional recognition accuracy, and does not mean that the textual characteristics of the lyrics do not have the ability to extract musical emotions from the clips to which they belong. Following Russell's model, Jamdar et al. roughly estimate and classify what the emotion is intended to be expressed in the testing song by calculating the arousal and the valence of each sentence and whole song with different methods. The overall process is shown in Figure 5.

Fig. 5 Process of extraction of lyrical features.

The steps to accomplish the process are simple. However, due to the ambiguity of the word, the emotions that can be expressed by words in different contexts may be different. Therefore, to avoid this ambiguity, the researchers used Parts-Of-Speech (POS) in this study to tag the whole lyric file. Similarly, the researchers constructed association rules and applied them to each sentence of the lyrics, so that the relationship between verbs, adjectives, nouns, and negativities could be established [10].

4. Composing with Emotion

Music creation usually requires motivation, and part of the trigger for this motivation can be attributed to momentary inspiration, such as the composer inadvertently playing some interesting melody. But more can be attributed to the composer's desire to use music to convey the emotions in his or her heart when composing music. Once upon a time, composers who wanted to complete a piece needed different musicians to play instruments in an ensemble in the studio; Now, with the help of computers and analog signals, a person can become a band. The same change occurred in the composition with emotion. Back in old days, people did not have the conditions to compose anytime and anywhere they want, which led to some works that could have been masterpieces disappeared because composers could not grasp fleeting emotions. But nowadays, people only need to click the mouse, adjust some features on the composition software, and they can quickly generate a piece of music full of their emotions at that time.

There are some examples of emotional music generation. In Liu and Ting's work, they introduce Zhu et al.'s work. Based on the KTH rule system, Zhu et al. created interactive genetic algorithms (GA) capable of composing music for specific emotions [11]. The rules within the KTH system are evaluated by simulating the performance principles of musicians on stage. Similarly, Ferreira and Whitehead have generated several pieces of emotional music on GA by applying their mLSMN (the generative multiplicative long short-term memory network with logistic regression) model [12]. Their work is based on the study of Radford et al. which generates text with emotions. They created emotional music based on valence of the music by tweaking several features of the given sample, such as melody, harmony, tempo, timbre, etc. Then, they invited several human annotators to describe their feeling about music clips generated by utilizing GA.

Casella and Paiva's study, MAgentA (Musical Agent Architecture) [13], is also a powerful agent that automatically generates music based on emotions. It consists of three modules, and related to the extraction and application of musical emotions are two of them: The Perception Module and The Reasoning Module. The Perception Module is used to periodically view the emotional changes caused by notes generated in the current environment, with a minimum probing period of one note per second. The Reasoning Module would receive any event comes from the Perception Module. It would identify whether the changed emotion has already been included in Mood Algorithms Database. Magenta is a customizable and powerful framework integrating emotional data collection and emotional music creation, it allows users to create their own emotion algorithm database and generate music with specific emotions by simply fine-tuning its parameters.

5. Limitations & Prospects

Even if the field of MER has received a lot of attention today, there are still certain shortcomings in MER method and its application. For example, a problem of finer judgment of emotions needs to be solved. According to Thayer's theory and Russell's two-dimensional emotion plane, many of works done by current researchers in MER area are based on the duo's theory. Although they have been tested by time, they are slightly outdated in today's highly technologically advanced social environment. The fundamental reason is that human emotions are extremely complex, although it can be summarized into four basic emotional categories, but if the goal of the MER research is to enable machines to better identify the subjective and objective emotions that music can express, researchers also need to provide richer and more detailed models of human emotions for machines to train and learn. In addition, research on the subjective emotions that music can induce is still rare. Even with the help of machine learning, because of the ambiguity of words, the emotions analyzed by the machine will still be biased from the emotions experienced by actual humans. Therefore, researchers need to take emotional granularity into account so that they can avoid related problems [14].

As for the discussion of the future direction of musical sentiment recognition, in the MER field, there are some uncharted areas waiting for researchers to explore, for example, the musical emotion classification (MEC) area. Although there are some researchers already conceived a idea about it, for

example, the research of Xu et al. [15]. In their conception, researchers should split the testing song into two parts: accompanying music and human vocals, and then perform emotional extraction on these two parts. Because in a normal song, the emotions expressed with the lyrics may be slightly different from those expressed by the accompaniment. Only by allowing researchers to separate these two parts for emotion extraction and analysis can the value of the two emotions be more clearly divided, and then achieve the purpose of more accurate musical emotion recognition.

6. Conclusion

In summary, this study mainly introduces two research directions of musical emotion: musical emotion extraction and musical emotion composition. Based on the description of musical emotions, three cases of musical emotion extraction are introduced, which are monitoring human ECG to determine the emotions induced by the current musical segment, using deep Gaussian processing to extract the features of the audio and classify the emotions corresponding to the obtained results, and finally analyzing the text features of the lyrics of the testing songs and matching the emotions. In addition, this paper gives examples of composing with musical emotions. In the end, limitation of current study and future work of MEC would be discussed. The significance of this paper is to introduce several major music emotion recognition methods and their applications in the current field, so that ordinary people can understand the role of machine learning in the entertainment industry, and point out the future development direction of the music industry for practitioners in music-related industries.

Referencess

[1] Perlovsky L. Musical emotions: Functions, origins, evolution. Physics of life reviews, 2010, 7(1): 2-27.

[2] De Berardinis J, Cangelosi A, Coutinho E. The multiple voices of musical emotions: Source separation for improving music emotion recognition models and their interpretability. Proceedings of the 21st international society for music information retrieval conference. 2020: 310-317.

[3] Liu J, Huang M. Music emotion understanding by computer based on BP neural network. 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). IEEE, 2017: 1-5.

[4] Kim Y E, Schmidt E M, Migneco R, et al. Music emotion recognition: A state of the art review. Proc. ismir. 2010, 86: 937-952.

[5] Russell J A. A circumplex model of affect. Journal of personality and social psychology, 1980, 39(6): 1161.

[6] Juslin P N. What does music express? Basic emotions and beyond. Frontiers in psychology, 2013, 4: 596.

[7] Juslin P N, Laukka P. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. Journal of new music research, 2004, 33(3): 217-238.

[8] Hsu Y L, Wang J S, Chiang W C, et al. Automatic ECG-based emotion recognition in music listening. IEEE Transactions on Affective Computing, 2017, 11(1): 85-99.

[9] Chen S H, Lee Y S, Hsieh W C, et al. Music emotion recognition using deep Gaussian process. 2015 Asia-Pacific signal and information processing association annual summit and conference (APSIPA). IEEE, 2015: 495-498.

[10] Jamdar A, Abraham J, Khanna K, et al. Emotion analysis of songs based on lyrical and audio features. arXiv preprint arXiv:1506.05012, 2015.

[11] Liu C H, Ting C K. Computational intelligence in music composition: A survey. IEEE Transactions on Emerging Topics in Computational Intelligence, 2016, 1(1): 2-15.

[12] Ferreira L N, Whitehead J. Learning to generate music with sentiment. arXiv preprint arXiv:2103.06125, 2021.

[13] Casella P, Paiva A. Magenta: An architecture for real time automatic composition of background music. Intelligent Virtual Agents: Third International Workshop, IVA 2001 Madrid, Spain, September 10–11, 2001 Proceedings 3. Springer Berlin Heidelberg, 2001: 224-232.

[14] Yang X, Dong Y, Li J. Review of data features-based music emotion recognition methods. Multimedia systems, 2018, 24: 365-389.

[15] Xu J, Li X, Hao Y, et al. Source separation improves music emotion recognition. Proceedings of international conference on multimedia retrieval. 2014: 423-426.