Emotion Profile-Based Music Recommendation

Yu-Hao Chin, Szu-Hsien Lin, Chang-Hong Lin, Ernestasia Siahaan, Aufaclav Frisky, and Jia-Ching Wang

Department of Computer Science and Information Engineering

National Central University, Taiwan, R.O.C.

Abstract- *In this paper, we propose an emotion profile based music recommendation system. In the proposed algorithm, two emotion profiles are constructed using decision value in support vector machine (SVM), and based on short term and long term features respectively. The recognized emotion, emotion profile, and personal historical query results are fed into the recommendation module to generate the recommended music list.*

Keywords- Emotion profile, music emotion recognition, music recommendation system.

I. INTRODUCTION

Recently, there is an amazing growth of music data on the Internet and users are easily to find out various online music data including music sound signals, biographies, discographies, and lyrics. Among the various online music data, the music sound signals which are existed in music files are the most interesting targets to retrieve. To take advantage of the huge amount of music files, an effective music recommendation system is essential. A good music recommendation can be helpful to searching music that fit user's preference. The system can provide users with useful information about the items that might interest them [16]. The function of instantly responding to changes in user's preference is a valuable asset for such systems [3]. Many music recommender systems have been proposed [17]. Existing recommender systems can be classified into three classes: content-based filtering, collaborative filtering, and hybrid approach.

In content-based filtering method, the user profiles are formed by extracting features of music which have been accessed in the past. Then the recommendation system can take user profiles for reference to recommend only the music contents that are highly relevant to the profile. This results in a large variety of artists. For example, various pieces are recommended even when their emotion has not been rated. However, user profile is just one of the many factors characterizing user preferences. It is hard to combine user profile with music database without rating [17]. In the collaborative filtering approach, user's profiles are collected and the relationship between users is analyzed. According the analysis result, the system classifies users to groups. The recommendation system recommends music derived from other profiles in same group. Therefore, each user's profile can be shared to make the system recommend music similar to user's preference [3]. On the other hand, hybrid system tries to combine these two kinds of methods through various ways [12]. For example, Linqi Gao made a hybrid model based on genetic algorithm to improve recommending ability [13].

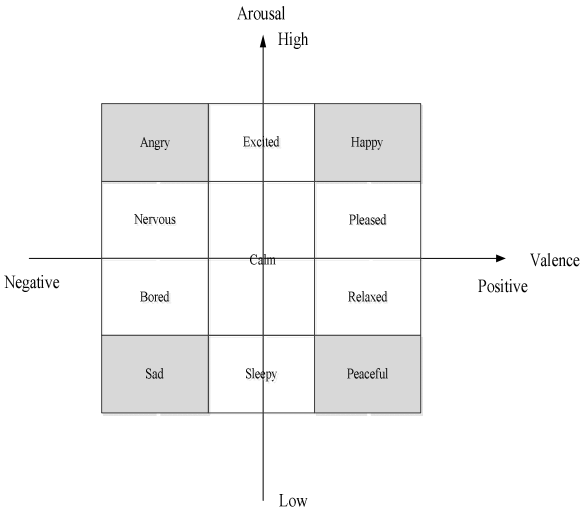

Fig. 1. Modified Thayer's 2-D arousal-valence plane.

To make an emotion based music recommendation system, the emotion content in music should be detected. Many researches have been proposed in music emotion detection [1], [5]-[7], [24]. Existing research method could be divided into two main kinds, dimension approach and categorical approach. Dimension approach maps features to a point on the emotion model plane [4], then regard the coordinate index as features to train regressors. Categorical approach works by directly feeding features into a classifier to recognize the corresponding emotion categories [5]. The recognition method this paper used belongs to the second type. Besides, we consider modified Thayer's arousal-valence plane shown in Fig. 1 [8]. According to the plane, human emotion can be classified into four main classes. We define angry, happy, sad, and peaceful to be emotion class. In Thayer's model, the *x*-axis refers to valence, and the *y*-axis refers to arousal. Valence refers to the grade of positive or negative perceived emotion [7], [9]. If the level of valence is low, it means the song is unpleasant. However, if it is high, the song is considered pleasant. Arousal refers to the intensity of perceived emotion [7], [9].

In this paper, an emotion profile based music recommendation system is presented. Its block diagram is depicted in Fig. 2.

978-1-4799-4266-4/14 $31.00 © 2014 IEEE
DOI 10.1109/U-MEDIA.2014.32

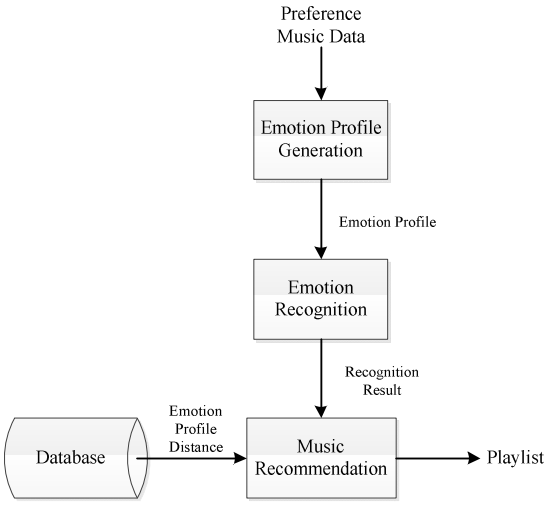

Fig. 2. Block diagram of proposed music recommendation system.

II. Acoustical Feature and Emotion Profile

In reality, perceived emotion is subjective, and it is hard for people to clearly express what they feel when they listen to a song. Previous works shows that we can use acoustical features to present emotion. It is noted that emotion perception is not based on a single feature but a combination of features [4], [18], [25]. There are 8 kinds of acoustical features extracted from music clips in this paper, which are divided into short term feature and long term feature. Short-term features include Eventdensity, Zerocross, Chromagram, MFCC, Spectrum Centroid, Spectrum Spread, and Spectrum Flatness. Long term-feature includes Legendre-based trend coefficients (LBTCs). Acoustical features are extracted from the data by using MIR toolbox [2].

The emotional profiles express the confidence of each of the eight emotion classes [15]. It is an approach to interpret the emotional content of natural human expression by providing multiple probabilistic class labels, rather than a single hard label. For example, happy emotion not only contains happiness content, but also consists of feature properties that are similar to the content of peaceful. The similarity to peaceful may cause data to be recognized as an incorrect class. Therefore, the evidence for happiness and peaceful are conveyed through representation in an emotion profile. The profile can help determine which emotion is most easily be perceived by human. In this paper, emotion content of signal is presented in terms of a set of simple emotion classes: happy, angry, sad, bored, nervous, relaxed, pleased, and peaceful.

In the following, the used acoustical features are described.

A. Eventdensity

Before we extract the eventdensity feature, we extract a feature called Onset first. Onset contains the energy of each pulse's successive pulse. Eventdensity means peaks picked per second on onset curve [2]. Onset curve is shown in Fig. 3.

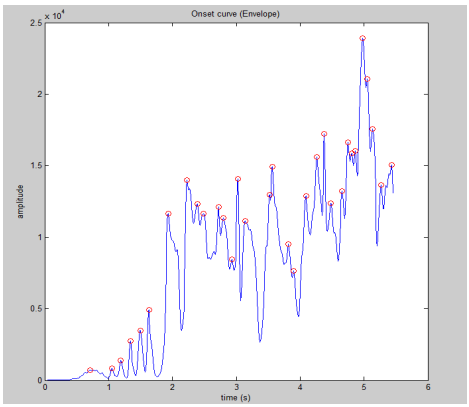

Fig. 3 Onset curve of an angry song

B. Zerocross

Zerocross refers to the number of times a signal changes signs [2].

C. Chromagram

Chromagram is also called Harmonic Pitch Class Profile. Chromagram presents the distribution of energy along the pitches or pitch classes [2].

D. MFCC

Mel-frequency cepstrum coefficients (MFCC) also have good performance on the system. This feature models the human auditory perception system. The origin of MFCCs is the power of Mel windows. The operation is defined as follows:

$$S^k = \sum_{\omega=0}^{F/2-1} W_\omega^k \cdot X(\omega), k = 1,...,M \quad (1)$$

In Eq. (1), $X(\omega)$ means the ω-th power spectral component of an audio stream, and S^k means the power of the k-th Mel window W_ω^k. M means the number of the Mel windows, which often ranges from 20 to 24. The transform operation is defined as follows:

$$C_n = \frac{1}{M}\sum_{k=1}^{N} \log(S^k)\cos\left[(k-0.5)\frac{n\pi}{M}\right], n = 1,...,L \quad (2)$$

In Eq. (2), let L denote the desired order of the MFCCs. We can then find the MFCCs using logarithm and cosine transforms.

E. Spectrum Centroid

Spectrum centroid is an economical description of the shape of the power spectrum [20]-[22]. Additionally, it is correlated with a major perceptual dimension of timbre; i.e. sharpness. Eq. (3) is used to extract the spectrum centroid.

$$ASC = \sum_i \log_2(f_i/1000)p_i \Big/ \sum_i p_i \quad (3)$$

where P_i is the power associated with frequency f_i.

F. Spectrum Spread

Spectrum spread is an economical descriptor of the shape of the power spectrum that indicates whether it is concentrated in the vicinity of its centroid, or else spread out over the spectrum [20]-[22]. It allows differentiating between tone-like and noise-like sounds. Equation (4) is used to extract the spectrum spread.

$$ASS = \sqrt{\sum_i ((\log_2(f_i/1000) - C)^2 p_i) \Big/ \sum_i p_i} \quad (4)$$

G. Spectrum Flatness

This feature describes the flatness properties of the spectrum of an audio signal within a given number of frequency bands. The flatness of a band is defined as the ratio of the geometric mean and the arithmetic mean of the spectral power coefficients within the band [20]-[22]. A high deviation from a flat shape may indicate the presence of tonal components.

H. Legendre-Based Trend Coefficients (LBTCs)

With the MFCCs and the subband powers, this study further uses audio features based on Legendre polynomial [23]. These new features which are called Legendre-based trend coefficients (LBTCs) are derived by computing the trend of each MFCC coefficient and the subband power using Legendre polynomial. To trend of each one dimensional feature, we use Eq. (5) to derive its LBTC.

$$LBTCi(X(n)) = \sum_{n=0}^{\Gamma} X(n) P_i(n), 0 \le i \le 3 \quad (5)$$

$$p_0(n) = (\frac{n}{\Gamma})(\Gamma+1)^{-0.5} \quad (6)$$

$$p_1(n) = \left[\frac{12M}{(\Gamma+1)(\Gamma+2)}\right](\frac{n}{\Gamma} - 0.5) \quad (7)$$

$$p_2(n) = \left[\frac{180\Gamma^3}{(\Gamma-1)(\Gamma+1)(\Gamma+2)(\Gamma+3)}\right]^{0.5} \times \left(\frac{n^2}{\Gamma^2} - \frac{n}{\Gamma} - \frac{\Gamma-1}{6\Gamma}\right) \quad (8)$$

$$p_3(n) = \left[\frac{2800\Gamma^5}{(\Gamma-2)(\Gamma-1)(\Gamma+1)(\Gamma+2)(\Gamma+3)}\right]^{0.5} \times \left[\frac{n^3}{\Gamma^3} - \frac{3n^2}{2\Gamma^2} + \frac{(6\Gamma^3 - 3\Gamma + 2)n}{10\Gamma^3} - \frac{(\Gamma-1)(\Gamma-2)}{20\Gamma^2}\right] \quad (9)$$

where n is the sample index, Γ is the total sample number, P_i (n) and X(n) are the i-th order of Legendre polynomial and the signal to extract its trend.

IV. Emotion Recognition

The emotion recognition system includes two sets of SVM. One is based on short term feature with probability product kernel [19]. The other is based on long term feature with RBF kernel. Each set of classifiers consists of eight SVMs, and classifier outputs are used to construct emotion profiles. We use the emotion profile to train another SVM to perform emotion recognition.

The SVM theory is an effective statistical technique and has drawn much attention in pattern recognition. An SVM is a binary classifier that creates an optimal hyperplane to classify input samples. This optimal hyperplane linearly divides the two classes with the largest margin [10], [14]. Functions that satisfy Mercer's theorem can be used as kernels. and is a kernel function. Using Mercer's theory, we can introduce a mapping function $\varphi(\mathbf{x})$, such that $k(\mathrm{x}_j, \mathbf{x}_i) = \phi(\mathbf{x}_j)\phi(\mathbf{x}_i)$. This provides the ability of handling nonlinear data, by mapping the original input space into some other space.

V. Music Recommendation System

This paper makes an emotion based music recommendation system. In previous work, many music recommendation systems have been proposed. For example, Kyoungro made a recommendation system which asks user to select the emotion class first, and the system recommend a play list [17]. Based on previous music recommendation system, our proposed system can transform music signal to an emotion distribution, and analyze which music might be preferred by user. First, user can choose to upload a music clip that user prefers to listen to, or adjust the emotion profile bar to decide emotion distribution. Our server will classify the data to emotion class. The system then outputs music recommendation lists, and present data's emotion profile through emotion distribution figure. Songs in recommendation list are sorted by emotion profile. For example, music in happiness class is sorted by happiness value. The system is comprised of client and server.

The music database is paced in the music server. For each music file, the music server also stores their corresponding emotion profile and emotion class label. Besides, the emotion class label also makes a taxiomony to structure the music files in the database. To store the emotion profile and emotion class label information, a Gold System is firstly established by training the hierarchical SVMs using the music files with manual emotion-label. Nest, all the music files in the database are feeding into the Gold System to generate their associated emotion profiles and emotion class labels. The strategy also permits an easy way to expand the music database as it avoids the manual labeling of the huge amount of music files as well as the re-training with the continuously growing database.

As mentioned in previous subsection, input query in the form of a user's preference music and emotion profile bars is received at the music server. For a preference music query, the query file is fed into the Gold System to generate its emotion profile and identify its emotion class. Only the music files with emotion-label the same as the identified emotion class database are selected as the recommendation candidates. The final recommendation music files that are

presented and ranked in the play list depends on a similarity measure between the emotion profiles of the input preference music and a database music file. Besides, it also depends on the personal historical query results. It is noted an emotion profile is treated as a mathematical vector to calculate the similarity measure, which Euclidean distance can simply take this role.

V. CONCLUSION

In this paper, we proposed an emotion profile based music recommendation system, which is based on the proposed music emotion recognition method. This system is mainly based on emotion profile generation, emotion recognition, and music recommendation. The first-level SVMs generate emotion profile while second-level SVM performs emotion recognition. The recommendation algorithm outputs the recommended music list in accordance with recognized emotion, emotion profile, and personal historical query results.

REFERENCES

[1] C. H. Yeh, H. H. Lin, and H. T. Chang, "An efficient emotion detection scheme for popular music," *IEEE Trans. Circuits and Systems, 2009. ISCAS 2009. Symposium*, pp. 1799-1802, 24-27 May 2009.

[2] O. Lartillot and P. Toiviainen, "MIR in Matlab (II): A toolbox for musical feature extraction from audio," in Proc. Int. Conf. Music Information Retrieval, pp. 127–130, 2007 http://users.jyu.fi/_lartillo/mirtoolbox/

[3] H. T. Kim, E. Kim, J. H. Lee, and C. W. Ahn, "A recommender system based on genetic algorithm for music data," *Proc. Int. Conf. Computer Engineering and Technology (ICCET)*, 2010 2nd International Conference, April 2010, pp.V6-414, V6-417.

[4] Y. H. Yang and H. H. Chen, "Prediction of the distribution of perceived music emotions using discrete samples," *IEEE Trans. Audio, Speech, and Language Processing,* vol. 19, no. 7, pp. 2184-2196, Sept. 2011.

[5] B. Han, S. Rho, and R. B. Dannenberg, and E. Hwang, "SMERS: Music emotion recognition using support vector regression," in *Proc. Int. Conf. Music Information Retrieval*, Kobe, Japan, 2009.

[6] L. Lie, D. Liu, and H. J. Zhang, "Automatic mood detection and tracking of music audio signals," *IEEE Trans. Audio, Speech, and Language Processing,* vol. 14, pp. 5-18, 2006.

[7] C. Y. Chang, C. Y. Lo, C. J. Wang, and P. C. Chung, "A music recommendation system with consideration of personal emotion," in *Proc. Int. Conf. Computer Symposium* (*ICS*), Dec. 2010, pp. 18-23, 16-18.

[8] R. E. Thayer, *The Biopsychology of Mood and Arousal*. New York: Oxford University Press, 1989.

[9] A. Gabrielsson, "Emotion perceived and emotion felt: Same or different?" *Musicae Scientiae*, pp. 123–147, 2002, special issue.

[10] V. Vapnik, Statistical Learning Theory, New York: Wiley, 1998.

[11] B. Shao, M. Ogihara, D. Wang, and T. Li, "Music recommendation based on acoustic features and user access patterns," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 17, pp. 1602-1611, 2009.

[12] L. Liu, F. Lecue, and N. Mehandjiev, "A Hybrid Approach to Recommending Semantic Software Services," *Proc. Int. Conf. Web Services (ICWS)*, July 2011, pp.379-386.

[13] L. Gao and C. Li, "Hybrid Personalized Recommended Model Based on Genetic Algorithm," *Proc. Int. Conf. Wireless Communications, Networking and Mobile Computing*, Oct 2008, pp.1-4.

[14] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," 2001. Available: http://www.csie.ntu.edu.tw/ ~cjlin/libsvm

[15] E. Mower, M. J Mataric, and S. Narayanan, "A framework for automatic human emotion classification using emotion profiles," *Audio, Speech, and Language Processing, IEEE Transactions,* vol. 19, no. 5, pp. 1057-1070, July 2011.

[16] S. J. Anderson, A. C. M. Fong, and J. Tang, "Robust tri-modal automatic speech recognition for consumer applications," *IEEE Trans. Consumer Electronics*, vol. 59, no. 2, May. 2013.

[17] K. Yoon, J. Lee, and M. U. Kim, "Music recommendation system using emotion triggering low-level features," *IEEE Trans. Consumer Electronics*, vol. 58, no. 2, pp.612-618, May. 2012.

[18] K. Hevner, "Expression in music: A discussion of experimental studies and theories," *Psychological Review*, vol. 48, no. 2, pp. 186–204, 1935.

[19] T. Jebara, R. Kondor, and A. Howard, "Probability product kernels," *Journal of Machine Learning Research*, vol. 5, pp. 819–844, July 2004.

[20] H. G. Kim, N. Moreau, and T. Sikora, MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval. New York: Wiley, 2005.

[21] M. A. Casey, "MPEG-7 sound-recognition tools," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, 737-747, 2001.

[22] ISO-IEC/JTCI SC29 WGII Moving Pictures Expeti Group. Information technology - multimedia content description interface - part 4: Audio. Comittee Draft 15938-4, ISO/IEC, 2000.

[23] C. F. Wu, "Bimodal emotion recognition from speech and facial expression," Master Thesis, Department of Computer Science and Information Engineering, National Cheng Kung University, 2002.

[24] F. C. Hwang, J. S. Wang, P. C. Chung, and C. F. Yang, "Detecting emotional expression of music with feature selection approach," in *Proc. Int. Conf. Orange Technologies* (*ICOT*), March. 2013, pp. 282-286, 12-16.

[25] I. Luengo, E. Navas, and I. Hernáez, "Feature analysis and evaluation for automatic emotion identification in speech," *IEEE Trans. Multimedia,* vol. 12, no. 6, pp. 490-501, Oct. 2010.