Emotion-aware Music Information Retrieval Based on Physiological Signals and User Profile

Fanjie Li
Sichuan University
fjlichn@gmail.com

Jeremy Ng
University of Hong Kong
jeremyng@hku.hk

Xiao Hu
University of Hong Kong
xiaoxhu@hku.hk

ABSTRACT

This study aims to explore the possibility of using physiological signals to detect users' emotional responses to music and whether individual differences can affect this process. A user experiment was conducted with 50 participants using a novel MIR system. Data was analyzed using machine learning methods.

1. INTRODUCTION

Music emotion or mood is an important factor in music information retrieval (MIR). However, mood perception varies from user to user, making it difficult to support emotion-aware MIR. With the rapid development of wearable technology, more user data can be collected in unobtrusive ways, which provides a new opportunity to study users' emotional responses to music. This study aims to explore the possibility of using physiological signals to detect users' emotional responses to music and whether individual differences (in personality, music preferences, etc.) can affect this process. A user experiment was conducted to collect users' interactions with a novel MIR system (Moody_v3). During the experiment, users' physiological signals were collected using a wearable wristband. The data were analyzed using machine learning based on features extracted from physiological signals recorded during music listening. The results may contribute to future implementation of emotion-aware intelligent music recommendation in MIR systems.

2. EXPERIMENTAL DESIGN

50 participants (24 male, 26 female) were recruited to join this experiment in a batch manner, with one to five participants in each session. All participants were undergraduate or graduate students in a comprehensive university in Hong Kong, whose majors ranged from Social Sciences, Science, Engineering, Business, Humanities & Arts to Medicine, with various music knowledge background and relatively high frequency of music listening in daily bases.

Each batch of experiment consisted of four main phases: 1) pre-experiment questionnaire; 2) instruction of Moody system and MIR task; 3) participants searching and listening to music in 40 minutes; 4) post-experiment survey. The pre-experiment questionnaire gathered demographic information, music listening behavior and preference, as well as personality and learning style. During phase 3, participants were asked to look for 10 or more songs to make a playlist using the Moody (version 3) system [5]. They were encouraged to search different types of music for more diverse experience. For each music piece listened for more than 30 seconds, participants would answer two questions on their current emotion that were popped up in Moody system. The first question asked participants to give a score of arousal [2] in a scale of 0 (not aroused) to 10 (highly aroused), while the second question was to choose an emotion from a set of options including happy, blessed, excited, sad, melancholic, etc. Besides, one of the participants in each batch was asked to wear an Empatica E4 wrist band [3] when conducting the music searching task. This wrist band recorded physiological signal including electrodermal activity (EDA), blood volume pulse (BVP), inter beat interval (IBI), heart rate (HR) and skin temperature (TEMP) during the whole search process.

After conducting the task, the last phase of the experiment was for the participants to fill a post-experiment questionnaire concerning their emotional states and general experience during the search process. The consent forms were signed at the beginning and a nominal fee was paid at the end.

3. DATA ANALYSIS

During the system interaction (phase 3 in the experiment), Moody system logged user interactions such as searching for songs, playing song, answering question, etc. The physiological data recorded during each single piece of music was aligned with user logs of Moody system based on the timestamps.

Features in physiological data during each music piece were extracted based on time series and spectrum analysis. The features included descriptive statistics such as median, range, standard deviation (stdev), means of the first difference in raw values (MFDR) and in normalized values (MFDN), means of the second difference in raw values (MSDR) and in normalized values (MSDN) [4], low and high frequency in frequency spectrum (LF, HF), and the ration between the two (LF/HF) [1]. Participants' answers on emotion questions were taken as the ground truth labels to be predicted.

A machine learning approach was applied to measure the reliability of using physiological data to recognize users' emotion in MIR systems. Specifically, we compared the performance of k-NN, decision tree, random forest, naïve Bayes and an ensemble vote classifier which combined the probability distributions of 3 base classifiers (i.e. k-NN, random forest and naïve Bayes).

Collected in the pre-experiment questionnaire, participants' personality was analyzed to see if it has a moderation effect on the relationships between physiological signals and emotional responses to music.

4. PRELIMINARY RESULTS

The arousal and mood values rated by participants were each categorized into 3 categories (i.e. positive, negative, neutral) for comparison using ANOVA and t-test as well as classification. Using a one-way ANOVA, we compared physiological features across three arousal/mood categories as well as t-test comparing positive and negative arousal/mood categories. Significant results are shown in Tables 1 and 2.

Across 3 **arousal** categories		Across 3 **mood** categories	
Feature	*p*	Feature	*p*
BVP_median	0.032	BVP_median	0.005
BVP_range	0.094	HR_mean	0.07
BVP_HF	0.044	HR_stdev	0.046
HR_stdev	0.025	HR_median	0.074
HR_range	0.011	HR_range	0.036
HR_MFDN	0.066	HR_MFDR	0.053
HR_MSDN	0.059	HR_MSDR	0.057
HR_LF	0.024	HR_LF	0.056
HR_HF	0.031	HR_HF	0.069
EDA_MFDN	0.004	IBI_mean	0.001
EDA_MSDN	0.005	IBI_median	0.002
EDA_LF/HF	0.037	EDA_ MFDN	0.062
		EDA_ MSDN	0.045
		EDA_LF	0.034
		EDA_HF	0.078

Table 1. Significant results of ANOVA (at $p<0.1$).

arousal		mood	
Feature	*p*	Feature	*p*
BVP_median	0.012	BVP_mean	0.037
BVP_range	0.081	HR_mean	0.073
BVP_LF	0.034	HR_stdev	0.015
BVP_HF	0.017	HR_median	0.074
HR_stdev	0.007	HR_range	0.011
HR_MFDR	0.045	HR_LF	0.019
HR_MSDR	0.046	HR_HF	0.029
HR_LF	0.007	IBI_mean	0.011
HR_HF	0.009	IBI_median	0.013
HR_range	0.003	EDA_MFDN	0.023
IBI_mean	0.055	EDA_MSDN	0.018
IBI_median	0.056	EDA_LF	0.09
EDA_MFDN	0.006		
EDA_MSDN	0.005		
EDA_LF	0.078		

Table 2. Significant t-test results on variables between positive and negative categories.

Based on above analysis, significant features were selected to train classifiers and classify participants' emotions into positive and negative ones. 10-fold cross-validation was applied to evaluate each classifier and the performances were generally not satisfactory, with kappa values ranging between -0.007 to 0.041. To further investigate how individual differences could affect the classification accuracy, the ensemble classifier was then trained with the data partitioned by participants, which showed that the consideration of individual difference improved prediction performances for 13 users on arousal and 12 users on mood. We also trained the ensemble classifier with the data partitioned by participants' personality, particularly on the dimension of extraversion and introversion. Classification results are presented in Table 3.

Personality	Arousal		Mood	
	Accuracy	Kappa	Accuracy	Kappa
Extrovert	83.39%	0.649	92.03%	0.808
Introvert	72.60%	0	76.14%	0.143

Table 3. Classification performances partitioned by participants' personality.

5. REFERENCES

[1] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals", IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5, pp. 293–302, Jul. 2002.

[2] J. A. Russell, "A circumplex model of affect.", Journal of Personality and Social Psychology, vol. 39, no. 6, pp. 1161–1178, 1980.

[3] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti, "Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition", 2014 EAI 4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth). IEEE, pp. 39-42, 2014.

[4] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: analysis of affective physiological state," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 10, pp. 1175–1191, 2001.

[5] X. Hu, J. Lee, D. Bainbridge, K. Choi, P. Organisciak and J. Downie, "The MIREX grand challenge: A framework of holistic user-experience evaluation in music information retrieval", Journal of the Association for Information Science and Technology, vol. 68, no. 1, pp. 97-112, 2015.