Human-centric Music Medical Therapy Exploration System

Baixi Xing
Zhejiang University
Hangzhou, China
sisyxing@gmail.com

Kejun Zhang
Zhejiang University
Hangzhou, China
channy@zju.edu.cn

Lekai Zhang
Zhejiang University
Hangzhou, China
zlkzhang@gmail.com

Eng Keong Lua
Faculty of Information Technology
Monash University
lua.eng.keong@monash.edu

Shouqian Sun
Zhejiang University
Hangzhou, China
ssq@zju.edu.cn

ABSTRACT

Music emotion analytic is useful for many human-centric applications such as medical intervention. Existing studies have shown that music is a low risk, adjunctive and therapeutic medical intervention. However, there is little existing research about the types of music with identified emotions that have the most effect for different medical applications. We would like to discover various music emotions through machine learning analytic so as to identify models of how music conveys emotion features, and determine its effectiveness for medical intervention and treatment. We are developing a Human-centric Music Medical Therapy Exploration System which could recognize music emotion features from Chinese Folk Music Library, and recommend suitable music to playback for medical intervention and treatment. Our networked system is based on Support Vector Machine (SVM) algorithm to construct the models for music emotion recognition and information retrieval. Our main contributions are as follows: Firstly, we built up the Chinese folk music emotions and features library; secondly, we conducted evaluation and comparison with other algorithms such as Back Propagation (BP) and Linear Regression to set up the model construction for music emotion recognition and proved that SVM has the best performance; lastly, we integrated blood pressure and heartbeat data analytic into our system to visualize the emotion fluctuation in different music affection and make recommendation for suitable human-centric music medical therapy for hypertensive patients.

Categories and Subject Descriptors

H.4.4 [**Information Interfaces and Presentation**]: Sound and Music Computing

General Terms

Theory

Keywords

Music information retrieval, Music emotion recognition, Music Therapy

FhMN'13, August 16, 2013, Hong Kong, China.

ACM 978-1-4503-2183-9/13/08…$15.00.

1. INTRODUCTION

Music is commonly considered closely related to human emotion. There is evidence that music could affect physical health parameter, such as heart rate, blood pressure and rehabilitation [12]. Yet the relationship between music, emotion and physiology is as intimate as it is intricate. This study involves music recommendation method based on music emotion recognition. There is plenty of research about effectiveness of music as an intervention for hospital patients [4]. However there is little existing research about the types of music with identified emotions that have the most effect for different medical applications. We are aiming at a Human-centric Music Medical Therapy Exploration System that we can obtain suitable music, providing the music resource for music therapy selection. To begin with a study of Chinese folk music, we find it achieve high scores in positive emotion feature than negative emotion. Since the existing research about music therapy usually applies classical music or instrument music in the graceful and smoothing emotion [12], while Chinese Folk Music has the similar emotion feature, we assume that it could have effectiveness for medical intervention and treatment.

This paper is organized as followings: section 2 generally introduces the related work of music emotion recognition and music information retrieval methods; section 3 presents three kinds of methods to construct the models for music emotion recognition, including Back Propagation (BP), Linear Regression and Support Vector Machine (SVM); section 4 describes the experiments, including the building of Chinese Folk Music emotion database, the experiment of emotion labeling experiment, and principal component analysis and factor analysis based on the experiment results; section 5 presents the efficiency comparison of using BP, Linear Regression and SVM models for music emotion recognition; section 6 briefly introduces the user interface of the Human-centric Music Medical Therapy Exploration System; section 7 we conclude with some directions for future research.

2. RELATED WORK

There is fast growing attention in music information retrieval, especially in the music emotion aspect. Several methods to retrieve emotion information from music features have been proposed. Typically, Yang [13] used Thayer emotion classification model along with Psysound [9] method to extract 15 features from music piece, and built up music emotion recognition model based on fuzzy methods and achieved a recognition rate of 78%. Chen Xiaoou research team from Peking University proposed the multimodal music emotion recognition method based on AdaBoost, and he conducted the test on 2500 Chinese songs [7]. It is demonstrated that the method could improve the music emotion recognition result effectively.

On the music feature selection aspect, the current research is focused on reducing the dimensions of features, more specifically, selecting the related features like Energy, Rhythm, Melody and Timbre. In the latest research, a music feature selection method based on psychoacoustics and music theory was carried out by Schmidt, however the model required human describing data to estimate the emotional content of music [3]. To be mentioned, there is not much research about music feature selection, the typical method, in most cases, relies on tools like MA Toolbox [2], MIR Toolbox, PsySound and Marsyas [6].

On the aspect of music emotion recognition algorithm research, Lin was the first to introduce multi-objective algorithm of support vector machine [15], then Lu proposed AdaBoost method to improve its performance [10]. Besides, the concept of emotion intensity was presented by Yang which made contribution to research of complex music emotion recognition. In the further study, Yang deployed music emotion classification as a regression problem [14], and predict the result by Multiple Linear Regression (MLR), Support Vector Machine (SVM), AdaBoostRT [1] algorithm. Since the regression method requires continuous solution space, there is certain error existing in the experiment.

In this paper, we choose Marsyas to extract 64 types of music features from 500 Chinese folk music clips. Then we apply three kinds of algorithm to tackle music information retrieval task, including BP, Linear Regression and SVM. We would like to find the best algorithm to recognize the music emotion information and provide the basic model for Human-centric Music Medical Therapy Exploration System, and applied it in the new domain of medical use to build up an integrated system.

3. MUSIC EMOTION ANALYTIC METHODOLOGY

In this research, three kinds of algorithms are demonstrated: Back Propagation, Linear Regression and Support Vector Machine.

3.1 BACK PROPAGATION

BP (Back Propagation) has the great strength in non-linear solutions to ill-defined problems, while the only layer nervous system could only solve the question of linear solutions. Multi-layers nervous network including hidden layer should be applied when solving the non-linear questions. BP could successfully solve the problem of reconstruction of non-linear function's multi-layers feed forward nerves system. Currently, this synergistically developed back-propagation architecture is the most popular, effective, and easy to earn model for complex, multi-layered networks.

3.2 LINEAR REGRESSION

Linear regression is a method to answer questions about a scalar dependent variable on one or more predictors, including prediction of future values of a response, discovering which predictors are important, and estimating the impact of changing a predictor on the value of the response [11]. Linear regression is used extensively in practical applications, because it is easier to fit the models which depend linearly on their parameters than the models non-linearly related to the parameters, and linear regression methodology is more transparent.

3.3 SUPPORT VECTOR MACHINES

The Support Vector Machine (SVM) is a supervised classification system that minimizes an upper bound on its expected error. It attempts to find the hyper plane separating two classes of data that will generalize best to future data. Such a hyper plane is the so called maximum margin hyper plane, which maximizes the distance to the closest points from each class. For linearly separable data, only a subset of the ais will be non-zero. Thus, an identical SVM would result from a training set that omitted all of the remaining examples. This makes SVMs an attractive complement to relevance feedback: if the feedback system can accurately identify the critical samples that will become the support vectors, training time and labeling effort can, in the best case, be reduced drastically with no impact on classifier accuracy.

4. EXPERIMENTS

4.1 Chinese Folk Music Database

The emotions within music are usually profound and variable from the start to the end. There will be a fluctuated process expressing emotion rise and fall, especially in Chinese folk music. In this situation, we should cut out a key music clip with a dominated independent emotion feature for the research. According to the method of segments of significant emotional expression (SSEE) stated by Tien-Lin Wu and Shyh-Kang Jeng from National Taiwan University , we proceed the "pre-segment" work for every song to separate a music clip of constant emotion.

In the pre-segment work session, the main emotion recognition and music clip selection are done by 5 music major students. They voted to have the conclusion of each song's dominated emotion feature, if the emotion is hard to define, it would be discard to avoid confusion in the following experiment. Collecting work process is as follows：

- Find 580 Chinese folk songs collection from internet and albums;
- Cut out 30 seconds clip in each song by Kugou software;
- Label each clip with fuzzy music emotion;
- Build up the music emotion feature library;
- Do the modeling with regression algorithm;
- Predict the emotion scores and count the accuracy.

4.2 Music Emotion Exploration Model

Music will be segmented, and the questionnaire results and music features extracted will form the database. Then the music emotion exploring system will be got by a promising algorithm. (Figure 1)

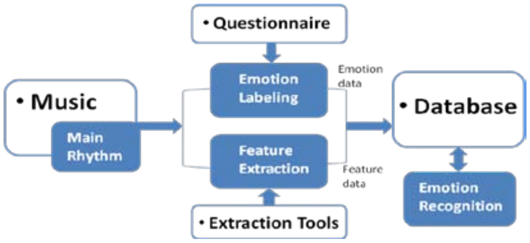

Figure 1. Data analyzing model for music emotion exploration

4.3 Music Material Collection

In this paper, we proceed to build up the music emotion library based on the Chinese folk music collection. Although Chinese folk music has a long history, unfortunately there are not that many preserved and passed down. We explore different ways to gather Chinese folk music collection, such as: timeline of Chinese folk music development, searching for folk music in different Chinese dynasties; Chinese instruments, categorizing folk music according to typical Chinese instruments; land region, appearing that climate and geography difference could nurture different culture which creates music diversity.

We choose the way of land region to make the collection of Chinese folk music for that it is the best way to avoid the repeat song selection situation and the most effective way to achieve the collection's integrity. We spent 2 months on the music searching work, and found out that Chinese folk music is dispersed in different regions which are hard to make a completed collection. Transforming Music to .mp3 format will cause information damage compression, while .wav format will not. When the music is transformed into .mp3 format, the Marsyas obtains several result values "0", which indicates that .mp3 is not suitable for Marsyas system. As a result, we clustere the collection of 580 Chinese folk music songs given as .wav format with the maximal information.

4.4 Music Clip Selection Principles

In terms that Marsyas system supports a length of 30 seconds music clip, we should cut out 30s music clip in every single song. There are some principles of selecting the clips as follows: 1.Only one music clip would be cut out in every single song by experts in music major, the emotion expressed in the music clip could represent the emotion feature of this song; 2.There is a completed melody in every music clip; 3.The main emotion feature is obvious in the music clip. 580 music clips are cut out from 580 songs of 40 Chinese folk music albums and the music source of internet. According to the principles, we select 500 music clips for the following experiment.

4.5 Music Emotion Labeling Experiment

The data collecting work took about six months. The dataset applied for this work consists of 500 music clips of relevant Chinese folk songs. The dataset is divided into 10 groups, each group consists of 50 music clips, for emotion labeling experiment.

- 50 volunteers (aged from 24 to 35) are invited to participate in the experiment. 5 students have music profession background, while 45 students are fond of music but do not have music profession background.
- The volunteer need to score the music clips in the questionnaire.
- They are arranged in separated cubic in a silent environment with no interruption.
- The volunteers wear the same type of headphone and check the music to the same sound volume to ensure the same effect.
- The whole experiment process lasts for one hour and a half, each clips is listened to 2 times.

According to Hevner model [8], the music emotion is set as 8 categories in the questionnaire: Vigorous, Dignified, Sad, Dreaming, Smoothing, Graceful, Joyous, and Exciting. Each volunteer completes the labeling work for one group of 50 music clips, rates each clip from value 0 to 1 for each of the 8 emotion labels. For instance, a certain clip of a song can be described as follows: (0, 0.1, 0, 0.3, 0.2, 0.2, **0.9**, 0.7). The Value above is corresponding to Vigorous, Dignified, Sad, Dreaming, Smoothing, Graceful, Joyous, and Exciting respectively. 0-There is no such kind of emotion expressed in this music clip. 1-There is strong level of such kind of emotion expressed in this music clip. 0.9-the value of Joyous means there is remarkable joyous emotion in this music clip. It is presented that "Joyous" had the highest value in the above, so the dominate emotion feature is joyous, this music clip could be described as "remarkable joyous, and very exciting".There is also some situation that the music clip has no distinct emotion feature. If a music clip has been rated and the main emotion got less than 70% of the total rating result will be considered as hard defined and remove it from the dataset.

5. DATA EVALUATION

5.1 Principal Component Analysis

We assumed that there are some key emotion labels in Chinese folk music, since Hevner model is applied to wider range of music genres. In order to demonstrate the hypothesis, we introduce the principle factor analysis and factor analysis methods to test the results of music emotion labeling experiment and expect to explore fewer factors. The analysis results are as follows: (Table 1, 2, 3).

According to the results (Table 3), 3 principal components can be concluded as 1 "Pleasant component" (representing joyous, graceful and dreaming), 2 "Solemn component" (representing dignified, sad and smoothing), 3 "Arousing component" (representing vigorous and exciting). The component analyzing result will be used in the modeling exploration as a reference.

Table 1. Component

Emotion	Initial	Extraction
Vigorous	1.000	0.821
Dignified	1.000	0.838
Sad	1.000	0.734
Dreaming	1.000	0.778
Smoothing	1.000	0.885
Graceful	1.000	0.875
Joyous	1.000	0.822
Exciting	1.000	0.829

Table 2. Total Variance Explained (V=Variance, C=Cumulative)

Initial Eigenvalues				**Extraction Sums of Squared Loadings**		
No	Total	% of V	C %	Total	% of V	C %
1	3.903	48.782	48.782	3.903	48.782	48.782
2	1.538	19.221	68.003	1.538	19.221	68.003
3	1.414	14.26	82.263	1.414	14.26	82.263
4	0.44	5.498	87.761			
5	0.386	4.83	92.591			
6	0.34	4.246	96.838			
7	0.163	2.031	98.869			
8	0.09	1.131	100			

Table 3. Principal Component Matrix

Emotion	Component		
	1	2	3
Vigorous	0.546	0.034	**0.722**
Dignified	0.149	**0.841**	0.329
Sad	-0.604	**0.547**	0.264
Dreaming	**0.786**	0.348	-0.196
Smoothing	0.614	**0.523**	-0.483
Graceful	**0.918**	0.007	-0.178
Joyous	**0.883**	-0.2	0.051
Exciting	0.774	-0.308	**0.368**

5.2 Chinese Folk Music Emotion Recognition

5.2.1 Music Emotion Recognition Model based on SVM and BP

400 music clips are selected as the training set in each kinds of emotions and the other 100 music clips are left as the testing set. Three steps will be made in achieving the comparison results:

- Step1: Train the classifier based on training set SVM and BP seperately with 400 music clips in the data collection of each emotion;
- Step2: Specify the best value for C and G parameter to build the best model with SVM;
- Step3: Test the best model with the left 100 music clips with SVM and BP seperately.

5.2.2 Discussion

We analyze all the music clips with Marsyas to extract 64 music timbre features from each clip. Timbre features are mainly MFCCs [5]. We framed the signals and calculate the spectral amplitude of each frame. Then we transformed the spectral amplitude's logarithmic to Mel domain and proceeded it with DCT (Discrete Cosine Transform). We select the first 13 dimension features. On the other hand, we transform the spectral amplitude with STFT (Short-time Fourier Transform) and extracted 3 features including centroid spectrum, spectrum attenuation and spectral flux. We got 16 features in total and compute their value of mean value, STD(Standard Deviation), Mean STD and STD STD. At last, we acquired 64 timbre features. Then we implemented the data in BP and LibSVM in WEKA to train the classifier to find the best model. According to crossover algorithm for the training set, we got the best parameters of C and G which were 1 and 0.45 respectively when we applied LibSVM. Finally, we tested the best model and got the results in tables below. (Table 4,5)

Table 4. Music Emotion Recognition Train Result

Algor.	BP		Linear		SVM	
	MSE	CC	MSE	CC	MSE	CC
Digni.	0.0167	0.681	0.0177	0.656	0.0087	0.865
Dream	0.0178	0.553	0.0194	0.492	0.0081	0.863
Excit.	0.0247	0.797	0.0258	0.786	0.0095	0.936
Grace.	0.0176	0.59	0.0200	0.505	0.0089	0.853
Joy	0.0231	0.795	0.0258	0.768	0.0099	0.928
Sad	0.0223	0.745	0.0245	0.719	0.008	0.928
Smoot.	0.021	0.712	0.0244	0.654	0.0093	0.904
Vigor.	0.0199	0.755	0.0230	0.710	0.0099	0.903

Table 5. Music Emotion Recognition Test Result

Algor.	BP		Linear		SVM	
	MSE	CC	MSE	CC	MSE	CC
Digni.	0.0239	0.23	0.0231	0.262	0.0223	0.258
Dream.	0.0338	0.221	0.0322	0.288	0.0331	0.139
Excit.	0.0382	0.646	0.0386	0.644	0.034	0.686
Grace.	0.0452	0.167	0.0436	0.214	0.046	0.17
Joy	0.0346	0.589	0.0325	0.623	0.0314	0.642
Sad	0.0519	0.559	0.0590	0.486	0.0503	0.56
Smoot.	0.0519	0.54	0.0531	0.506	0.0518	0.582
Vigor.	0.0282	0.597	0.0271	0.608	0.0255	0.625

There are two key parameters to evaluate the model: average CC (correlation coefficient), which represents the relative error, while the value of CC stays less than 0.01 will be regarded as a small error rate; MSE (Mean Squared Error), which shows the control force, while the value of MSE reaches above 90% will be seen as having the optimal force. We can see from tables, the recognition results of SVM presented that average CC (correlation coefficient) is 89.75% for training set, 43.39% for the testing set; average MSE (Mean Squared Error) is 0.009 for the training set and 0.037 for the testing set . While the average CC (correlation coefficient) is 70.3% for training set with BP algorithm, 44.36% for the testing set; average MSE (Mean Squared Error) is 0.020 for the training set and 0.038 for the testing set. It is demonstrated that SVM achieves best performance in Chinese Folk Music Emotion recognition, however the average CC for training set and testing set is not outstanding, and especially Dignified, Dreaming and Graceful achieved a relatively low value of CC, which needs further research to improve the test set result.

6. HUMAN-CENTRIC MUSIC MEDICAL THERAPY EXPLORATION SYSTEM

The Human-centric Music Medical Therapy Exploration System could recognize the emotion feature of Chinese Folk Music songs and recommend suitable songs to playback according to preset emotion type and intensity. User could choose the music emotion type and adjust intensity on the system interface, and then the system could search out the corresponding music and playback. At the same time, the system will detect the blood pressure and heartbeat rate by means of the intelligent electronic blood pressure monitor (Figure 2).

Figure 2. The intelligent electronic blood pressure monitor.

The data processing center collects the data (i.e. pulse wave) transmitted from user's blood pressure monitor in real-time, and compares the data with normal blood pressure data.

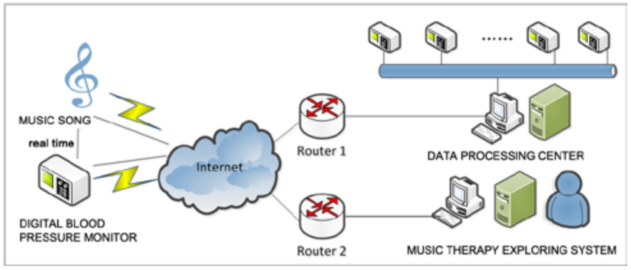

Figure 3. The intelligent blood pressure monitoring system

In the applying process, hypertensive patient's bloodpressure data will be recorded when he is listening to the songs from the Chinese folk music library. If the patient's blood pressure and heartbeat decreases remarkably, then we consider the song might be suitable for hypertensive patients and the system will rate the song with a relatively high score (ranking 0 to 100). We assume that the song with a score of over 80 could be remarked a star to be a recommended song. The score rating is related to the bloodpressure decreasing range. The system score rating mechanism is under further estimation. (Figure 3)

Since music affection differs for different individuals, patient user could have their personal music therapy which included songs that have remarkable performance on reducing their bloodpressure. The songs which get average high scores for large scale of patient users will be collected in the music therapy library for hypertensive patients, and the patient user could search recommended songs from the library. The system could search out more songs from the Chinese folk music library which have similar emotion feature with the recommended songs to enrich the therapy library for patients. (Figure 4)

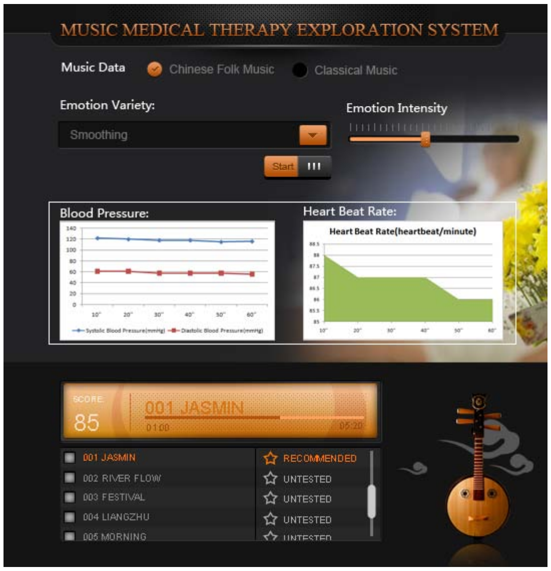

Figure 4. Human-centric Music Medical Therapy Exploration System Interface

7. CONCLUSION

In the study, a music library of Chinese folk music was built up by Marsyas. On the basis of the library, we conducted evaluation and comparison with other algorithms such as BP, Linear Regression and SVM, and proved that SVM has the best performance. Then a Human-centric Music Medical Therapy Exploration System based emotion recognition model was made. We integrated blood pressure and heartbeat data analytic into our system to visualize the emotion fluctuation in different music affection and make recommendation for suitable human-centric music medical therapy for hypertensive patients. It is significant to explore an integrated system which combines music and medical use, and create a new application domain for music emotion recognition and information retrieval. At the present stage, we are studying about whether Chinese folk music could be music therapy for hypertensive patients. Ultimately, the Human-centric Music Medical Therapy Exploration System could make recommendation for different medical use. In the future study, we are planning to evaluate the usability of the system, user's response will be collected to improve the user experiment. Further research also includes large scale data for Chinese Folk music library, accuracy improvement in modeling and multiple label recognition application, and Human-centric Music Medical Therapy Exploration System application evaluation in the hospital environment of different aspect. Specifically, we are going to use other methods to improve the modeling and introduce music therapy sharing into social network to enrich the music data library and explore more potential of music emotion recognition and information retrieval application.

8. ACKNOWLEDGMENT

This study is partly supported by the National Natural Science Foundation of China (61070075, 61004116, 61003147), Postdoctoral Science Foundation of China (2013M531458), Postdoctoral Science Foundation of Zhejiang (Bsh1202055) and Major Program of Science and Technology of Zhejiang Province (2011C14018).

9. REFERENCES

[1] Solomatine, D. P., Shrestha, D. L. 2004. AdaBoost.rt: A Boosting Algorithm for Regression Problems. IEEE International Joint Conference, 1163-1168.

[2] Pampalk, E. A. 2004. Matlab Toolbox to Compute Music Similarity from Audio. In *Proceedings of the International Conference on Music Information Retrieval*, 254-257.

[3] Schmidt. E. M., Turnbull, D., and Kim, Y. E. 2010. Feature Selection for Content-based Time-varying Musical Emotion Regression. *In Proc. ACM Int. Conf. Multimedia Information Retrieval*, 267-274.

[4] Evans, D. 2002. The effectiveness of music as an intervention for hospital patients: a systematic review, *Journal of Advanced Nursing,* 37(1), 8–18.

[5] Wong, E. and Sridharan, S. 2001. Comparison of linear prediction cepstrum coefficients and mel-frequency cepstrum coefficients for language identification. In Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, 95–98.

[6] Tzanetakis, G. and Cook, P. 2002. Musical Genre Classification of Audio Signals. *IEEE Trans. Speech Audio Process*. 10(5): 293-302.

[7] Guan, D. Chen, X. and Yang, D. 2012. Music Emotion Regression Based on Multi-Model Features, *The 9th International Symposium on Computer Music Modeling and Retrieval*, 70-77.

[8] Hevner, K. 1936. Experimental Studies of the Elements of Expression in Music. *American Journal of Psychology*, 48: 246-268.

[9] PsySound, http://members.tripod.com/~densil/.

[10] Lu, Q. Chen, X. Yang, D. and Wang, J. 2010. Boosting for Multi-modal Music Emotion Classification. In *Proceedings of the International Conference on Music Information Retrieval*, 105-110.

[11] S. Weisberg. 2005. *Applied Linear Regression*, John Wiley & Sons.

[12] Teng, X. F. Wong, M. Y. M. Zhang, Y. T. 2007. The Effect of Music on Hypertensive Patients, In *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, 4649-4651.

[13] Yang, Y.H. Liu, C.C. and Chen, H. H. 2006. Music Emotion Classification: A Fuzzy Approach. In *Proceedings of the 14th annual ACM international conference on Multimedia*, 81-84.

[14] Yang, Y. H. and Chen, H. H. 2011. Prediction of the distribution of Perceived Music Emotions Using Discrete Samples. *IEEE Trans. Audio, Speech Lang. Process*, 19(7): 2184-2196.

[15] Lin, Y. P. Wang, C. H. Wu, T. L. Jeng, S. K. and Chen, J. H. 2009. EEG-based Emotion Recognition in Music listening: A Comparison of Schemes for Multiclass Support Vector Machine. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 489-492.