Instance-based genre-specific music emotion prediction with an EEG setup

Liu, Xiaoyu; Phyo Wai, Aung Aung; Kumaran, Shastikk; Saravanan, Yukesh Ragavendar; Lin, Zhiping

2018

Liu, X., Phyo Wai, A. A., Kumaran, S., Saravanan, Y. R., & Lin, Z. (2018). Instance-based genre-specific music emotion prediction with an EEG setup. 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2092-2095. doi:10.1109/EMBC.2018.8512630

<https://hdl.handle.net/10356/136687>

<https://doi.org/10.1109/EMBC.2018.8512630>

© 2018 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. The published version is available at:
<https://doi.org/10.1109/EMBC.2018.8512630>

Downloaded on 26 Apr 2023 22:01:59 SGT

Instance-Based Genre-Specific Music Emotion Prediction with An EEG Setup

Xiaoyu Liu, Aung Aung Phyo Wai, Shastikk Kumaran, Yukesh Ragavendar Saravanan, Zhiping Lin

Abstract— This paper explores a novel direction in music-induced emotion (music emotion) analysis – the effects of different genres on the prediction of music emotion. We aim to compare the performance of various classifiers in the prediction of the emotion induced by music, as well as to investigate the adaptation of advanced features (such as asymmetries) in improving classification accuracy. The study is supported by real-world experiments where 10 subjects listened to 20 musical pieces from 5 genres- classical, heavy metal, electronic dance music, pop and rap, during which electroencephalogram (EEG) data were collected. A maximum 10-fold cross-validation accuracy of 98.4% for subject-independent and 99.0% for subject-dependent data were obtained for the classification of short instances of each song. The emotion of popular music was shown to have been most accurately predicted, with a classification accuracy of 99.6%. Further examination was conducted to investigate the effect of music emotion on the relaxation of subjects while listening.

Keywords— Music Emotion, Electroencephalography (EEG), Machine Learning, Music Genres

I. INTRODUCTION

Emotion is an intense response [1], consisting of behavioral, physiological, verbal and neural mechanisms [2], directed at some certain stimuli (which may be an individual or an object). The effects of differing emotional states have been noted in several domains, particularly health and productivity. Due to the omnipresence of music within society, music-induced emotion (*music emotion*) has gained research attention over recent decades. In the neural domain, the features of music-induced cortical responses have been identified [3 - 5]. This achievement has been used for multiple medical purposes, including improving life quality [6], the provision of antenatal training or the provision of therapies on serious illnesses such as cancer and stroke [7][8].

The use of electroencephalography (EEG) has been established as a reliable methodology to study emotional states through the examination of brainwaves in different frequency bands at differing electrode placement locations in the brain. Previous research has applied multiple types of audio-visual stimuli, including pictures, sounds, video and facial expression images [9] [10], but there is comparatively limited EEG research with music, despite the ability of music to induce a deeper emotional experience across cultures [11 - 13]. In particular, the use of popular music, or music across

widely listened to genres, has been heavily limited in EEG research, perhaps due to the reduced effect for differentiation of music of different affects or genres. This has limited the applicability of music emotion prediction research, as the music used in research may be different from the music being listened to across populations. Furthermore, there is a lack of investigation into the variability of music emotion prediction accuracy across genres, which limits application and validation of such techniques.

This paper explores a novel direction in music emotion analysis by investigating the effects of different genres on the prediction of music emotion. We compare the performance of various classifiers in the prediction of the music emotion and apply various advanced features in improving classification accuracy. Besides, we also validate the expert-based annotation process by comparing such a process with a subject-based annotation process through subjective annotation from the selected 10 subjects.

This paper is organized as follows: Section II describes the experiment design, data collected, feature extraction as well as analysis. Section III presents the results and discussions. In Section IV, conclusions are drawn.

II. METHODS AND EXPERIMENT

A. Experiment Design

In this experiment, EEG data were collected from 10 healthy subjects. As shown in Figure 1, the experiment was conducted for a duration of 30 minutes. Surveys were also used to collect basic non-identifiable information about participants and to collect their subjective responses to music for further validation of the expert-based annotation approach. Subjects were given 10-second breaks between each song, and they were played a 10-minute relaxing soundtrack at the end of the experiment to regulate their emotions.

All experiments in the paper involving human subjects are approved by the Institutional Review Board.

Figure 1: Experiment Flow

Xiaoyu Liu and Zhiping Lin are with School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore. (email: liux0073@e.ntu.edu.sg, ezplin@ntu.edu.sg)

Aung Aung Phyo Wai is with school of Computer Science and Engineering, NTU, Singapore. (email: apwaung@ntu.edu.sg)

Shastikk Kumaran and Yukesh Ragavendar Saravanan are with National University of Singapore High School (NUSHS). (email: shastikk@gmail.com, h1310146@nushigh.edu.sg)

Figure 2 Experiment Setting

1) *Emotion Categorization*: Emotion can be viewed through multiple dimensions [14]. In a 2D coordinate map, emotion can be clearly described using 2 parameters: arousal and valence [15]. Therefore, emotion is categorized into four here: High Arousal High Valence (HAHV), High Arousal Low Valence (HALV), Low Arousal High Valence (LAHV), Low Arousal Low Valence (LALV). From the existing valence-arousal model, we can match these states to the following emotional states respectively – happiness, anger, relaxation and sadness.

2) *Music Selection*: We aim to select music that provides adequate balance between various genres of music and various emotional states for the purposes of providing a fair sample of music for the purposes of calibration of the EEG classification model. We therefore select a total of 20 pieces of music – 4 from each genre (classical, electronic dance music, heavy metal, rap and popular music), and 1 from each emotion selected (happiness, sadness, relaxation and anger). 26-second segments of these songs are selected, which are later split into 64 instances of less than a second each for more accurate emotional prediction (we observe that the emotion of a song is able to transit over the period selected). The 10-minute piece played to subjects is of a relaxing nature to prevent adverse effect on experimental results by alternative emotional state/affect from other factors.

3) *Participants*: A total of 10 right handed healthy subjects with mean age of 24.3 ± 6.02 participated in this experiment. All subjects did not have prior medical conditions, were of full hearing ability and were able to express emotion fully, as self-declared in the survey. Most subjects had minimal musical training and were therefore considered as non-musicians, allowing for easier generalization of obtained results for use in the general public for further emotion prediction applications.

4) *Task*: Participants were asked to wear the *Muse* headband and were given required assistance in doing so. They were provided full and extensive instructions on the experimental procedure, before answering a survey which aimed to obtain non-identifiable limited information about the participants for further use in investigation. They were then asked to close their eyes when each individual music piece played. The EEG/Muse set-up was collecting data continuously with time-stamp used to sync between the set-up and *OpenSesame* used to play music. After each piece, a survey was conducted to understand their subjective responses on arousal and valence of the music played for

further use in validation of the expert-based annotation approach.

B. Dataset

1) *Data Collection*: *Muse* (from InteraXon) headband utilized as shown in Fig. 2 was applied to collect EEG data. *Muse* has been widely in brainwave analysis such as meditation [16], fatigue driving simulation [17] and has been proved to be effective in both relaxed and active emotional states. [18]. Four electrode placement locations were utilized for collection of brainwave data- TP9, AF7, AF8, TP10. The reference was placed at the AFZ electrode placement location with 220Hz sampling rate. All results were then stored and further processed after time-based segmentation with the purpose of synchronization and processing for the creation of a classification model.

2) *Features Extraction*: Basic brainwaves and their ratios: For each channel, normalized Alpha (α : 7.5-13 Hz), beta (β : 13-30 Hz), delta (δ : 1-4 Hz) and theta (θ : 4-8 Hz) band powers were extracted as the basic features. Accordingly, the ratios θ/β and δ/β were computed based on the discovery between the ratios and level of comfort, conducted by Tortella et al. [19] in their work. 6 features in total.

Asymmetries: We applied frontal asymmetry which Panagiotis et al. have proved that it can enhance the EEG emotion recognition [9] and is widely used in EEG data analysis. Besides, we have also applied left and right frontal-temporal asymmetry (AF7/TP9 and AF8/TP10), as well as temporal asymmetry (TP9/TP10) for both alpha and beta bands. 8 features in total.

Mellow: For the purposes of examining levels of relaxation, the normalized value of mellowness was computed based on α with additional processing to ensure that the value was more accurate and representative of the true state of relaxedness. This feature was selected for analysis due to its stability and accuracy as shown in the plotted diagram in Fig. 3 and Fig. 4, which shows an example of 1 subjects' mellow plotting for angry and happy labeled music pieces. The value of happy kept steady at 1 during the music piece, which indicates that the subject had a positive and stable emotion. This matches the qualitative understanding that happy music would allow individuals to be mellow. On the other hand, plot for music inducing anger showed significantly lower relaxedness in subjects.

Figure 3 Graph of Mellowness Data - Angry

Figure 4 Graph of Mellowness Data - Happy

C. Classification

We used consumer EEG headset for music classification accuracy and the *Weka* software was utilized in classification to collect the prediction accuracy of the training model obtained from the data. A 10-fold cross-validation test was conducted to obtain prediction accuracy. Feature extraction was conducted (all the features mentioned in in Part B) for each subject and each song on all relevant data prior to the creation of classification models to ensure optimization of selected features for the increase of prediction accuracy. Three classifiers were utilized – the random forest ensemble classifier, the random tree classifier as well as the bagging classifier.

III. RESULT AND DISCUSSION

We obtained the 10-fold cross-validation accuracy for the EEG data cross all 3 classifiers to compare their ability and accuracy in classification. As presented in Table I, we have obtained both subject-independent and subject-dependent classification accuracies, as well as comparison among the analysis of only basic unprocessed features, advanced features and combination of both basic and advanced features. We have also compared accuracy when expert annotation was used as compared to subject-based annotation to validate the expert-based approach.

We obtained a maximum accuracy of 98.42% for subject-independent data from the random forest classifier and 99.16% for subject-dependent data through the same. The random forest ensemble classifier overwhelmingly showed the strongest performance amongst the 3 classifiers used, showing its potential for use in emotion recognition applications. It is worth noting that cross-validation accuracy may be boosted with the use of instances as opposed to classifying for the entire song's emotion, which may affect results especially as the induced emotion for any individual would be considered through the whole song, rather than per individual instance. We generally observe that when basic and advanced features are combined, higher accuracy is observed as compared to when only basic or advanced features are used. Basic features provide higher classification accuracy than advanced features. The expert-based and subject-based annotation processes generally provide similar accuracies, validating the expert-based process for emotional classification of music. Finally, subject-dependent accuracies, as expected, are higher than that of subject-independent data.

Table I: Classification Result

	Feature	Rate & Label	Random Tree (%)	Random Forest (%)	Bagging (%)
SD	BF	EA	89.83	95.28	93.94
		SR	90.47	95.64	93.91
	AF	EA	89.31	95.66	93.50
		SR	90.05	95.72	93.72
	BF & AF	EA	90.78	99.16	95.52
		SR	91.28	99.05	95.77
SI	BF	EA	85.44	94.22	93.33
		SR	85.55	95.16	93.27
	AF	EA	84.13	93.58	93.16
		SR	84.42	93.58	92.02
	BF & AF	EA	86.55	98.42	95.61
		SR	86.03	98.29	95.5

SD – Subject Dependent; SI – Subject Independent; AF – Advanced Feature; BF – Basic Feature; EA – Expert Annotation; SR – Subjective Rating

Table II: Genre – Based Accuracy

	Random Tree (%)	Random Forest (%)	Bagging (%)
Classical	93.05	99.53	97.27
EDM	93.59	99.61	97.03
HM	93.59	99.29	96.88
Pop	94.84	99.61	97.81
Rap	91.88	98.98	95.86

In addition, we collect the classification accuracy for individual genres using the subject-independent set-up with both basic and advanced features and expert annotation, which has already been validated. We obtain the highest classification accuracy from the random forest classifier across all genres. Among genres, pop shows the best performance with the highest accuracy of 99.61% using the ensemble classifier. This may be caused by the general familiarity of the public with pop due to the high prevalence of such music in society. This may provide easy identification and induction of relevant emotion present in the individual music pieces [20]. On the other hand, rap has the lowest classification accuracy among all the genres, with 98.98% classification accuracy. This may be because individuals subjected to the experiment is less aware of rap music and the nuance presented in such music, due to the lack of prominence of such music as compared to other genres within the listening preferences of the populations represented by the subjects. It may also be indicative of the difference of how emotion is portrayed in rap – rap music often portrays music with lyrical cadence as opposed to instrumental music, which means that it is more difficult for listeners to comprehend the emotion of the music. Furthermore, lyrical content is more relevant to rap music, and contributes more heavily to understanding of music emotion. This has been validated, with multiple studies stating that much of the emotion of

music relies heavily on the existence of lyrical content and melodic contextual clues [21].

We find that on average, after normalization, subjects find pop to be the most positive genre valence-wise, followed by dance music, classical music, rap and then heavy metal as shown in Fig. 5. This does match up with existing societal understandings of the genres in question, which treat rap and heavy metal as music which are considered rather angry and aggressive in nature, while EDM and pop music are considered happy, uplifting music in comparison due to the need of music from those genres to reach wider audiences across the world. An example of this is *This Can't Be Life* by Jay-Z, which has an instrumental that would induce happiness but overall sadness of emotion due to the lyrical content. For arousal as shown in Fig. 6, results also match societal understandings, with metal and pop having the highest arousal, followed by rap, EDM and classical, which may be considered to have low arousal due to the relatively little rhythm and vocals present.

IV. CONCLUSION

We have found that in this experiment, music emotion prediction through the construction of classification models based on EEG data is highly efficient and accurate, especially the use of ensemble-based classifiers in such work. We have also found that the splitting of music pieces into multiple instances after time synchronization with EEG data through segmentation can obtain a more accurate result, as well as the creation of a methodology to calculate the overall music emotion of a song. Besides, the combination of both basic and advanced (processed) features during the classification process helps in getting higher accuracy as well.

With the finding of different accuracy of various genres and their effects on valence and arousal, we will look to build therapeutic models for treatment through auditory stimuli as well as examining societal and contextual links to perceived valence and arousal in various genres of music.

Figure 5 Graph of Subject Valence Per Genre

Figure 6 Graph of Subject Arousal Per Genre

ACKNOWLEDGMENTS

The authors would like to thank all the subjects who participated in the survey and experiment.

REFERENCES

- [1] Hume, D. Emotions and Moods. *Organizational Behavior*, 258 – 297.
- [2] Fox Elaine. "Emotion Science: An Integration of Cognitive and Neuroscientific Approaches." Palgrave MacMillan. ISBN 978-0-230-00517-4, pp. 16–17, 2008.
- [3] Koelsch, S. "Neural substrates of processing syntax and semantics in music", *Current opinion in neurobio.* vol. 15, no. 2, pp. 207-212, 2005.
- [4] Koelsch, S. & Friederici, A. D. "Toward the neural basis of processing structure in music". *Annals of the New York Academy of Sciences* vol. 999, no.1, pp. 15-28, 2003.
- [5] Pearce, M. T. et al. "Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation". *NeuroImage* vol. 50, no. 1, pp. 302-313, 2010.
- [6] Nikolay Mirenkov, Kamen Kanev, Hiroshi Takezawa, "Quality of Life Supporters Employing Music Therapy", *22nd Int. Conf. on Advanced Information Networking and Applications - Workshop*, March 2008
- [7] Magee, W. L., & Davidson, J. W. "The effect of music therapy on mood states in neurological patients: a pilot study." *Journal of Music Therapy*, vol. 39, no. 1, pp. 20-29, 2002.
- [8] Nayak, S., Wheeler, B. L., Shiflett, S.C., & Agostinelli, S. "Effect of music therapy on mood and social interaction among individuals with acute traumatic brain injury and stroke." *Rehabilitation Psychology*, vol. 45, no.3, pp. 274-283, August 2000.
- [9] Panagiotis C. Petrantonakis, Leontios J. Hadjileontiadis, "A Novel Emotion Elicitation Index Using Frontal Brain Asymmetry for Enhanced EEG-Based Emotion Recognition", *IEEE trans. On Information Technology in Biomedicine*, vol. 15, no. 5, Sep. 2011.
- [10] Ruoyu Du, Hyo Jong Lee, Ran Wei, "Brain Emotional Oscillatory Activity During the Different Affective Picture and Sound Experience", *10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*, October 2017.
- [11] Hongjian Bo, Lin Ma, Haifeng Li, "Music-evoked emotion classification using EEG correlation-based information," *39th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Sep 2017.
- [12] I. Daly, D. Williams, J. Hallowell, F. Hwang, A. Kirke, A. Malik, J. Weaver, E. Miranda, and S. J. Nasuto, "Music – induced emotions can be predicted from a combination of brain activity and acoustic features," *Brain and Cognition.*, vol. 101, pp. 1-11, November 2015.
- [13] T. Fritz, S. Jentschke, N. Gosselin, D. Sammler, I. Peretz, R. Turner, A. D. Friederici, and S. Koelsch, "Universal recognition of three basic emotions in music," *Current biology*, vol. 19, no. 7, pp. 573-576, April 2009.
- [14] Shah, R.; Lewis, M. B. "Locating the neutral expression in the facial – emotion space" *Visual Cognition.* vol. 10, pp. 540 – 566, 2003.
- [15] Schacter, Daniel L. *Psychology Ed. 2*. 41 Madison Avenue New York, NY 10010: Worth Publishers. ISBN 1-4292-3719-8, 2011.
- [16] Decho Surangsrirat, Apichart Intarapanich, "Analysis of the meditation brainwave from consumer EEG device", *SoutheastCon*, April 2015.
- [17] Ruyi Foong, Kai Keng Ang, Chai Quek, "Corelation of reaction time and EEG log bandpower from dry frontal electrodes in a passive fatigue driving simulation experiment", *39th Annual Int. Conf. of the IEEE EMBC*, Sep 2017.
- [18] Mohan S. Ijjada, et al. "Evaluation of Wearable Head Set Devices in Older Adult Populations for Research", *Int. Conf. on Computational Science and Computation Intelligence*, December 2015.
- [19] Tortella-Feliu M., Morillas-Romero A., Balle M., Llabres J., Bornas X., Putman P. "Spontaneous EEG activity and spontaneous emotion regulation," *Int. Journal of Psycho.* vol. 94(3), pp. 365-372, Dec 2014.
- [20] Yuiko Kumagai, Mahnaz Arvaneh, Haruki Okawa, Tomoya Wada, Toshihisa Tanaka, "Classification of Familiarity Based on Cross -Correlation Features Between EEG and Music," *39th Annual Int. Conf. of the IEEE EMBC*, Sep 2017.
- [21] McKay, C., et al. "Evaluating the genre classification performance of lyrical features relative to audio, symbolic and cultural features," *Proc. of the Int. Society for Music Information Retrieval Conf.* 213-8. 2010.