3F3-OS-19a-1

EEG-based Investigation of Music Familiarity and Emotion

Nattapong Thammasan*1 Koichi Moriyama*2 Ken-ichi Fukui*2 Masayuki Numao*2

*1 Graduate School of Information Science and Technology, Osaka University
*2 The Institute of Scientific and Industrial Research (ISIR), Osaka University

Familiarity is a crucial factor in music engagement, but very few study focuses on its neural correlates. We investigated effects of familiarity based on electroencephalogram in music-emotion recognition. Our research focused on self-reporting and continuous annotation based on the hypothesis that emotion in music experiencing is subjective and varies over time. Our methodology allowed subject to select up to 16 MIDI songs, comprising of eight familiar songs and the remaining were unfamiliar songs. We found that prior knowledge of the music indicated by familiarity level affected brain activities reflected by our indicators in frequency domain analysis. Furthermore, using entirely-unfamiliar songs was superior to mixed data in emotion classification. Therefore, unfamiliar songs are recommended to be used in emotion recognition system construction.

1. Introduction

Emotion is conscious experience indicating mental state and psycho-physiological expression. It is well-known that emotion has strong relevance with brain and other physiological reactions, and this makes emotion-brain research become a highly active research. Electroencephalogram (EEG) is a noninvasive tool to capture the electrical signal along the scalp, resulting from activities of neurons of the brain. EEG is frequently used to investigate nature of emotion in the brain due to its prominences in high temporal resolution and significantly low cost. Until now, there are several studies investigating emotion with music using EEG, especially study of differentiating positive and negative emotion in brainwave signal. For instance, there was a work discovering higher activity in left frontal lobes of brain in comparison to the right hemisphere while subjects were in positive emotions, especially in alpha band [Schmidt 2011]. Based on literatures, there are plenty of works proposing emotion recognition algorithms or system construction methods [Lin 2010, Sourina 2012]. Previous researches either used pre-emotion-labeled music clips obtained from standard library, or allowed only single-time emotion annotation in one song. As music is dynamic and changes overtime, it is quite preferable to capture continuous change of emotional state and to evaluate emotion dynamically rather than employing global annotation for the song. Our research focused on self-reporting and continuous annotation based on the assumption that emotion in music experiencing is subjective and varies over time.

As emotions in music listening are subject-dependent, higher level of knowledge in the music stimulus could evoke emotions differed to expected emotions. Thus, familiarity to musical pieces can influence emotion elicitation in real-world music listening. However, familiarity in music is mostly neglected in previous music-emotion recognition research. There are very few study investigating familiarity effect to brain in music listening. Pereira *et al* reported the role of familiarity in the brain correlates of music appreciation using fMRI [Pereira 2011]. The difference in EEG signals of professional musician and non-musical expert were also found in [Mikutta 2014] at the same level of familiarity to specific musical piece. Until now, the difference of brainwave at different level of familiarity to musical excerpt is still unknown. This research is an early work to study of neural correlates of familiarity effects based on EEG. Furthermore, the implication of familiarity affect to music emotion recognition system is also investigated.

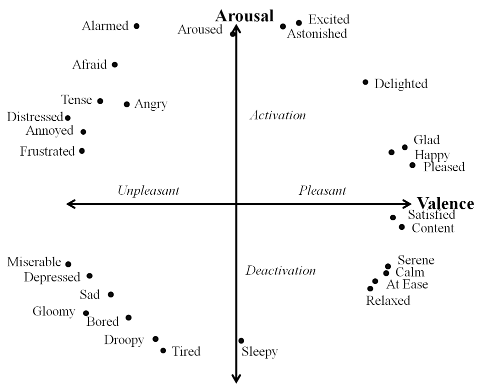

Figure 1: Russell's two-dimensional arousal-valence space [Russell 1980]

To describe human emotion systematically, one of the famous models is arousal-valance emotion model proposed by Russell in 1980 [Russell 1980]. In this bipolar model (Figure 1), valence is represented as horizontal axis indicating positivity or negativity of emotions, while arousal is vertical axis that represent high or low of strengths of emotions. We employed the two-dimensional emotion model to represent emotions as it was effective model in recent works to recognize emotion during music listening, including our preliminary research[Yamano 2012, Thammasan 2014].

Contact: Nattapong Thammasan, Department of Architecture for Intelligence, The Institute of Scientific and Industrial Research (ISIR), 8-1, Mihogaoka, Ibaraki, Osaka 567-0047, Japan, +81-6-6879-8426, E-mail: Nattapong@ai.sanken.osaka-u.ac.jp

2. Research Methodology

2.1 Data Collection Methods

The experiments were conducted in our own developed Java software . The procedure started with questionnaire regarding to personal information and musical preferences. Then, a subject selected 16 music clips to listen from the 40-song MIDI library via the software. By the instruction given to subject, eight of 16 selected songs were the songs that the subject was familiar with, and the others were the unfamiliar songs. Our designed software offered a function to play short sample clip cut from original music clips for comforting subject to indicate familiarity level of each song, where 1-3 referred to low familiarity level (unfamiliar songs) and 4-6 denoted high familiarity level (familiar songs).

The waveguard EEG cap*1 was then prepared to record brain signals. 12 exploited electrodes from total 21 EEG electrodes were named Fp1, Fp2, F3, F4, F7, F8, Fz, C3, C4, T3, T4, and Pz according to their positions in 10-20 international system. Sampling frequency was 250Hz. Impedances of every electrode were control not to exceed 20 kΩ. Notch filter, a band-stop filter with a narrow stop-band, was applied to remove 60Hz power line noise. The subjects were advised to close their eyes and minimize their movement while brainwave were being recorded. EEG signals were amplified by Polymate AP1532*2 amplifier, which was connecting to computer to visualize the signals via amplifier's software, APMonitor*3.

Afterwards, the selected music clips were played for subjects as synthesized sounds by Java Sound API's MIDI package. The lengths of songs were 2 minutes on average. 16-second pauses were included between the end of one song and the beginning of next song in order to allow subject to relax and to avoid effects from previous song. After music listening session ended, subjects listened to the same songs in previous session's list again. This time, subject was instructed to annotate his/her emotions perceived in previous session continuously by clicking on a point in two-dimensional emotion space shown on monitor screen. Arousal values was in the range of -1 to 1, as well as valence values. Finally, subjects were allowed to change familiarity level to any song which they felt that they had known prior to the experiment after listening to full-length song.

2.2 Data Preprocessing

Bandpass filter was applied to extract only 0.5–60 Hz EEG signal. We utilized EEGLAB[Delorme 2011], one of the most popular software/library to handling artifact running under the MATLAB environment, to automatically detect and reject obvious artifact-contaminating data to clean EEG signal. Rejecting portions of continuous EEG data can be performed visually using a function in EEGLAB. In addition, eye-blinking artifact that can not be detected by method in previous step can be removed by applying the computational method in signal processing called independent component analysis (ICA). The ICA technique decomposes multivariate signals into independent non-gaussian subcomponents. We modified technique of artifact removal by using components of Fp1 and Fp2 electrodes, because these two frontal electrodes were closest to eye and eye-blinking artifact had obvious effects on them. After all, we associated EEG signal in each record by emotion data annotated from subjects according to timestamps.

2.3 Feature Extraction Algorithms

We extracted feature from EEG signals by applying two different algorithms. The calculations were performed MATLAB analysis software. We also implemented the algorithm with a sliding window technique to the sequential data to retrieve more data. In addition, sliding window technique enables the possibility to develop real-time emotion classification system. The window size was 1000 samples, equivalent to 4-second length. Note that overlapping of one sliding window to the adjacent window was adjustable.

2.3.1 Fractal Dimension Value

Fractal Dimension (FD) values reveal the complexity of the signals and quantify the concentration level of brain state. It is frequently applied in affective computing researches including EEG-based emotion recognition[Sourina 2012] because of its attractive simplicity and success. In this reseach, we calculated the FD values using the Higuchi algorithm [Higuchi 1988].

In order to obtain the fractal dimension, Higuchi algorithm considers a time series $X(i)$ where $i = 1, ..., N$. From this series, a new series $X_m^k(i)$ can be constructed by the following definition:

$$X_m^k : X(m),\, X(m+k),\, ...,\, X\left(m + \left\lfloor\left(\frac{N-m}{k}\right)\right\rfloor k\right) \quad (1)$$

where $k = 1, 2, .., 2^{(\lfloor \log_2 N \rfloor - 4)}$ is the interval time and $m = 1, 2, \ldots, k$ is the initial time. For example, assuming that the series has $N = 40$ elements and $k = 3$, then the series separated into three series as follow:

$$\begin{aligned} X_1^3 &: X(1),\, X(4),\, X(7),\, ...,\, X(37),\, X(40) \\ X_2^3 &: X(2),\, X(5),\, X(8),\, ...,\, X(38) \\ X_3^3 &: X(3),\, X(6),\, X(9),\, ...,\, X(39) \end{aligned} \quad (2)$$

Then the length of the series X_m^k is defined as:

$$L_m(k) = \frac{1}{k}\left[\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} \frac{\mid X(m+ik) - X(m+(i-1)k) \mid}{\frac{\lfloor (\frac{N-m}{k}) \rfloor k}{N-1}}\right] \quad (3)$$

The average value of the lengths, $\langle L(k) \rangle$, is obtained by averaging all the sub-series lengths $L_m(k)$ for the given k. A relationship exists as follows:

$$\langle L(k) \rangle \propto k^{-\mathrm{FD}} \quad (4)$$

Fractal dimension (FD) is obtained by logarithm plotting between k and associated $\langle L(k) \rangle$and calculating the slope. Higuchi algorithm can be applied even over time series that are not stationary.

*1 http://www.ant-neuro.com/products/waveguard
*2 http://www.teac.co.jp/industry/me/ap1132/
*3 Software developed for Polymate AP1532 by TEAC Corporation.

2.3.2 Power Spectral Density

Recently, power spectral density (PSD) of the EEG were often used to study the relationship of emotional states and brainwave. PSD reveals the power of the signal in specific frequency band. This method is based on the fast Fourier transform (FFT). In this research, each EEG signal is decomposed by PSD method into 5 distinct frequency bands: delta (0–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–40 Hz). The method is applied to EEG data by a library in MATLAB Signal Processing Toolbox[*4]. Note that the integral of the PSD over a given frequency band computes the average power in the signal over that frequency band using toolbox's *avgpower* method. The technique uses a rectangle approximation of the integral of the signal's PSD over specific frequency range.

2.4 Emotion Classification

In this research, emotions were divided to 2 set of 2 classes: high-low arousal and positive-negative valence, corresponding to positions in arousal-valence model. We referred annotated arousal and valence values as "1" class for zero and positive values, and as "0" class of negative values. Since the sliding window technique is applied to EEG data in feature extraction step, all emotion annotation data labeled to each record of EEG signal are unified to label to the window by *majority* method.

For classification, we created two model to recognize arousal and valence class separately by employing three different classifying algorithms: Support Vector Machine, Multilayer Perceptron, and C4.5. The former two algorithms are among most common methods in brain-computer interface, while the C4.5 is dominant in speed of learning. All classification is performed by using WEKA[Hall 2009].

3. Neural Correlates of Familiarity

To investigate neural correlates in music listening, we performed feature extraction by PSD based on 50% on overlapping on sliding window technique to increase number of data for analysis. We examined 5 brain frequencies of signals. From features data extracted by PSD within one subject, we separated data into two groups by level of familiarity: data from familiar song and data from unfamiliar song. We calculated $t - test$ of two set of data in particular bandwave in the electrode, and counted as 1 if there was significant difference between data from familiar songs and from unfamiliar songs at significant level $p<0.01$. We assumed that the difference in specific bandwave and electrode could vary over subject, so we combined the result of significant difference counting across subjects together and examined which bandwave were frequently counted on the difference. The results of counting significant differences are shown in Figure 2. As upper bound number of counting in each bandwave was 180 (12 electrodes x 15 subjects), alpha rhythm was very dominant by this indicator. Plausible reason would be that alpha wave refers to relax state of the brain, thus listening to familiar and unfamiliar songs could differ from each other at the level of relaxation.

*4 http://www.mathworks.com/help/signal/ref/dspdata.psd.html

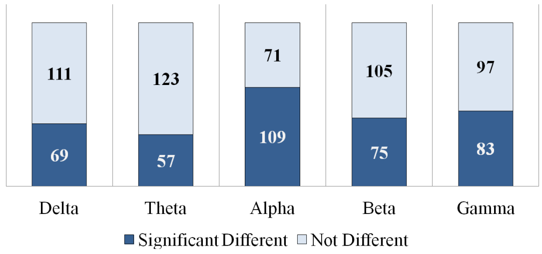

Figure 2: Significant differences of band power between familiar song listening and unfamiliar song listening

4. Familiarity matters Emotion Recognition System

In this section, we focus on familiarity effect to emotion recognition system. As our goal is to construct subject-dependent emotion recognition system, we created model and tested it using data within one subject. Our testing mode was 10-fold cross-validation, which randomly partitions into 10 equal size subsamples. In one trial, a single subsample is used as the validation data for testing the model, and the remaining 9 subsamples are used as training data. The process iterates for 10 times and the results from each fold are averaged to produce overall estimation. To avoid bias of classification due to similar data in training and testing instances, we adjusted overlap between each consecutive sliding windows to zero.

4.1 System Decomposition

EEG data from our experiment were divided in to two groups according to familiarity level; familiar song data and unfamiliar song data. Beyond artifact removal, average percent of data from familiar songs were 52.46 % of entire data, while the most considered unbalance data contained 57.84 % of all data as instances from familiar song session.

4.2 Chance Level

As dataset was unbalance in each subject, we calculated classification results at chance level for each subject as a ground-benchmark to evaluate emotion recognition system. The chance level indexes were calculated by percentage of the maximum between positive and negative instances to total number of instances in each subject.

4.3 Results of Emotion Classification

Due to unbalance of positive and negative instances, we considered averaged emotion classification results above the averaged chance levels in each dataset (familiar/unfamiliar/all songs dataset). In arousal recognition, the relative results are illustrated in Figure 3. It was noticeable that data from unfamiliar song session improved classification over random level more than what data from all sessions did, while results from familiar song session were similar to all song session. The best relative result was from classifying FD value features by SVM using only unfamiliar song data and achieved absolute classification result of 86.6% (SD = 8.1%), while the chance level of unfamiliar song data was 64.2% (SD = 7.0%).

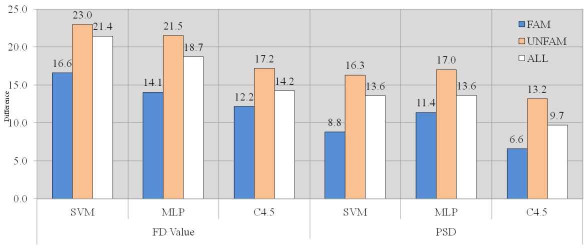

Figure 3: Comparison of arousal classification enhancements above chance level between data from familiar/unfamiliar/all songs

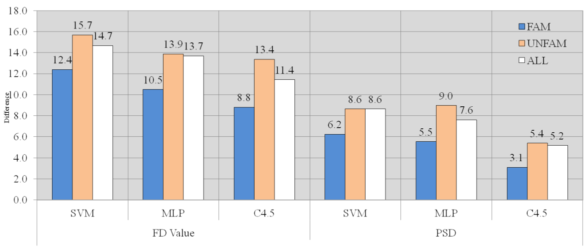

Figure 4: Comparison of valence classification enhancements above chance level between data from familiar/unfamiliar/all songs

In valence recognition, we calculated the performance above chance level in similar way, illustrated in Figure 4. We found that the unfamiliar songs achieved better relative results compared to all data again. Again, FD value classification by SVM achieved highest relative accuracy, while the absolute accuracy was 85.2% ($SD = 8.1\%$) and the chance level in this mode was 69.6% ($SD = 11.2\%$).

5. Discussion

The results suggested that using only unfamiliar songs could give better classification in emotion recognition system using EEG based on continuous self-annotation. It is noticeable that annotation play important role in this research, and it could have direct influence to the aforementioned results. It is possible that familiarity could influence annotation in some way. For instance, even though the musical structure could successfully elicit the same emotion across subjects, which produces similar brain activities, subjects' familiarities to the song could influence self-annotation from subjects. The inconsistency of emotion report is still challenging emotion-music research especially the one that focuses on subjective emotion reports. Another question, how well people can perceive or recognize their own emotional state, still has no clear answer, and it is valuable to examine.

6. Conclusion

Familiarity that influences musical experience is also meaningful to emotion recognition research. Utilizing self-reporting and continuous emotion annotation approach, we concluded from empirical results that data sorely from unfamiliar songs could improve emotion classification compared to ground level indicated by classifying by chance. Therefore, unfamiliar songs are recommended to be utilized in emotion recognition system construction.

References

[Delorme 2011] Delorme, A., Mullen, T., Kothe, C., Akalin Acar, Z., Bigdely-Shamlo, N., Vankov, A., and Makeig, S.: EEGLAB, SIFT, NFT, BCILAB, and ERICA: New Tools for Advanced EEG Processing. Computational Intelligence and Neuroscience (2011),

[Higuchi 1988] Higuchi, T.: Approach to an irregular time series on the basis of the fractal theory, Physica D: Nonlinear Phenomena, 31, 277-283, (1988),

[Hall 2009] Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., and Witten I.H.: The WEKA data mining software: an update, SIGKDD Explor. Newsl. vol. 11, no. 1 (2009),

[Lin 2010] Lin Y.P., Wang C.H., Jung T.P., Wu T.L., Jeng S.K., Duann J.R., and Chen J.H.: EEG-Based Emotion Recognition in Music Listening, Biomedical Engineering, IEEE Transactions on , vol. 57, no. 7, pp. 1798–1806 (2010),

[Mikutta 2014] Mikutta C.A., Maissen G., Altorfer A., Strik W., and Koenig T.: Professional musicians listen differently to music, J. Neuroscience, vol. 268, pp. 102–111 (2014),

[Pereira 2011] Pereira C.S., Teixeira J., Figueiredo P., Xavier J., Castro S.L., and Brattico E.: Music and Emotions in the Brain: Familiarity Matters, PLoS ONE 6(11) (2011),

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

[Schmidt 2011] Schmidt L.A. and Trainor L.J.: Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions, J. Cognition & Emotion, vol. 15, no. 4, pp. 487–500 (2001),

[Sourina 2012] Sourina O., Liu Y., and Nguyen M.K.: Real-time EEG-based emotion recognition for music therapy, J. Multimodal User Interfaces, vol. 5, pp. 27–35 (2012),

[Thammasan 2014] Thammasan N., Fukui K., Moriyama K., and Masayuki Numao M.: EEG-Based Emotion Recognition during Music Listening, Proc. of the 28th Annual Conference of the Japanese Society for Artificial Intelligence (2014),

[Yamano 2012] Yamano Y., Cabredo R., Inventado P.S., Legaspi R., Moriyama K., Fukui K., Kurihara S., and Numao M.: Estimating Emotions on Music Based on Brainwave Analyses, Proc. The 3rd International Workshop on Empathic Computing (2012).