Journal of Theoretical and Applied Information Technology
15th March 2018. Vol.96. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MUSIC RECOMMENDATION SYSTEM BASED ON GENRE DISTANCE AND USER PREFERENCE CLASSIFICATION

JONGSEOL LEE[1], KYOUNGRO YOON[2*], DALWON JANG[3], SEI-JIN JANG[4], SAIM SHIN[5], JI-HWAN KIM[6]

[1,2] Department of Computer Science and Engineering Konkuk University, 05029, Seoul, Korea

[1,3,4,5] Communications & Media R&D Korea Electronics Technology Institute, 13509, Gyeong-gi, Korea

[5,6] Department of Computer Science and Engineering Sogang University, 04107, Seoul, Korea

leejs98@konkuk.ac.kr[1], yoonk@konkuk.ac.kr[2], dalwon@keti.re.kr[3], ssjang@keti.re.kr[4], mirror@sogang.ac.uk[5], kimjihwan@sogang.ac.uk[6]

Corresponding author[*]: Phone: +82-2-450-4129

ABSTRACT

Background/Objectives: The personalized music recommendation services can support the user-favorite contents among various multimedia contents. In order to predict user-favorite songs, it is necessary to manage user preferences information and genre classification.

Methods/Statistical analysis: We introduce the mechanism about the automatic management of the user preferences in the personalized music recommendation service. This system automatically extracts the user preference data from the user's brain waves and audio features from music.

Findings: In our study, a very short feature vector, obtained from low dimensional projection and already developed audio features, is used for music genre classification problem. We applied a distance metric learning algorithm in order to reduce the dimensionality of feature vector with a little performance degradation. Proposed user's preference classifier achieved an overall accuracy of 81.07% in the binary preference classification for the KETI AFA2000 music corpus.

Improvements/Applications: we could recognize the user's satisfaction when we use brainwaves. This system can be applied to various audio devices, apps and services.

Keywords: *Music Recommendation, Personalized Service, Genre Distance, Similarity, Genre Classification, EEG Extraction*

1. INTRODUCTION

In accordance with developing digital multimedia environment, it becomes harder and harder to fully enjoy appropriate contents. In order to decrease users' effort - selecting enjoyable contents, there are a lot of needs for the personalized recommendation services. The personalized recommendation services are the agents which predict the favorable contents of users and recommend these to each user.

Commercial music recommendation systems are categorized into two main filtering system groups: content-based and collaborative [1, 2]. Content-based filtering methods are based on the descriptions of music contents and user preference profiles. In a content-based filtering system, constructed keywords and user profiles are used to indicate the types of music content the user likes. In other words, these systems attempt to recommend music contents based on user's content consumption history [3]. This type of recommendation system can apply content analysis techniques, which include various cutting-edge developments in multimedia processing, signal processing, and text analysis processing. However, these techniques have fundamental limitations. Because content-based filtering is focused on the user data generated in the past and the contents offered in the recommendation process, it is impossible to identify the user's real-time reactions. Furthermore, the system requires a significant cost in the manual construction of content information.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Collaborative filtering considers about huge user consumption information, behaviors, and preferences [1]. The user information gathered by the recommendation system can reveal periodic trends in a user's habits. Moreover, these systems cannot account for real-time user responses, as the statistical analysis of user information is executed by the recommendation engine only periodically.

These previous music recommendation systems have technical limitations. First, the current commercial music recommendation systems cannot reflect the user's emotional preference, because the analyzed data do not contain the user's emotional reactions 4. Second, Many systems use a general genre classification based on text such as ID3 tag. Now a day, much music is composed of many kinds of genre.

In, we use brainwave signal classification techniques for user preference management and genre classification based on music features to overcome this limitation. It is very important to classify user's preferences in a recommendation engine. For that, many explicit or implicit way such as usage pattern analyzing and thumb up/down are used. In the past research, we proposed the MusicRecom system based on user's music usage patterns. The preference classification through usage patterns is an implicit analyzing. This system has a cold-start problem because when a user starts recommendation system at first, the system doesn't have his usage history. We didn't use traditional heavy devices for EEG recording. The Brain Machine Interface (BMI) developed for MusicRecom is designed for the user's comfort. The proposed service, for MusicRecom, uses personalized user preference classification models, which recognize positive and negative preference based on real-time brain signals

2. System Architecture

We propose the MusicRecom system which consists of four main functional modules as shown in Figure 1. The user preference classification module collects the user's brainwave signals and classifies them. Through this process, the MusicRecom can get a basic preference music list (10 songs) to use seed music for the recommendation. The music information analyzer parses ID3 tag of the music file or uses an open API for music information for preferred music list. We extract song title, singer, and composer. Singer and composer information will be used in the music recommendation process. This information is managed by the unique content id. The genre classification module extracts the features and classifies genres of 10 seed musics. We utilize the advanced algorithms for the proposed genre classification [5, 6]. The last one is the recommendation engine selecting the music that has the most similar features to user's favorite songs based on genre, singer, and composer.

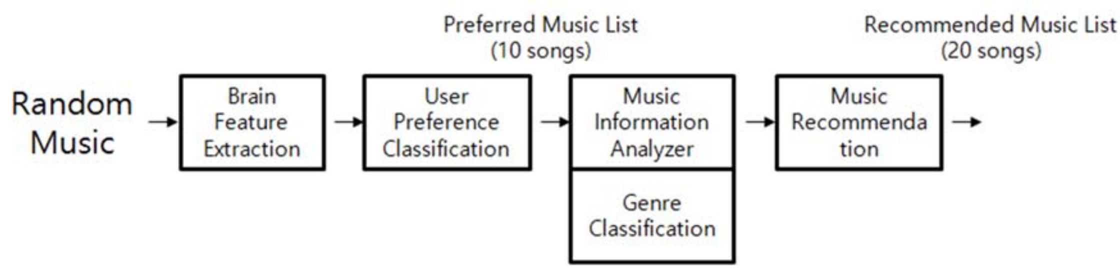

Figure 1: System architecture MusicRecom system based on EEG

2.1. Preference Classification Based on Brainwaves

In past research [8], we developed Brainwave based mood classification algorithm. In this paper, we applied it for preference classification.

Through analyzing the EEC signals when the user hears music, the proposed music recommendation system can recognize the user's emotion. The EEG-based music recommendation system has various advantages as compared with another recommendation techniques. The proposed service is able to respond to the user's real-time emotional state, since EEG reflects the real-time emotional response of the user.

Journal of Theoretical and Applied Information Technology
15th March 2018. Vol.96. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

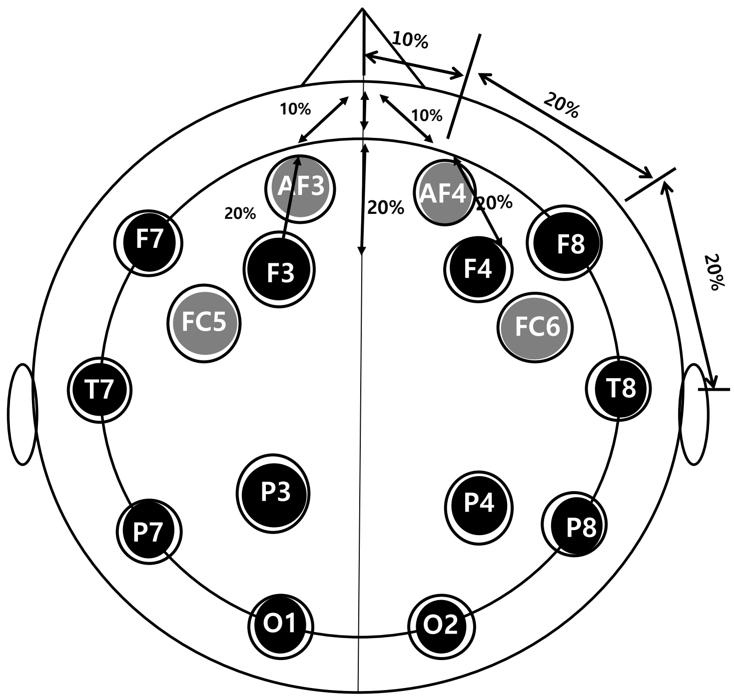

Figure 2: EEG electrode position in EPOC [13]

We use the Emotiv EPOC to collect EEG signals [8]. The EPOC EEG headset has a good reputation among 13 outstanding EEG headsets in the usability test [8]. The 16 EEG acquisition points are illustrated in Figure 2. The system continually acquires signals from 16 pins including 2 reference pins. In order to extract the spectral features from 14-pin EEGs, the brain feature extractor takes the Fast Fourier Transform (FFT) of the 14 EEG signals in a window of 8 seconds and receives the power spectra. The EEG sampling rate of the emotive EPOC is 128 Hz, 9 the window length for the 512-point FFT of the received EEG signals is 8 seconds, and the shift period is 8.7 ms [8, 10]. The power spectra are shown in Table 1 [11, 12].

Table 1: Frequency bands and Frequency ranges [7]

Frequency Band	Frequency Range (Hz)
Delta (δ)	0-4
Theta (θ)	4-7
Alpha (α)	7-13
Beta (β)	13-30
Gamma (γ)	30-50

The process of feature extraction is illustrated in Figure 3. p and t denote a pin index ($1 \le p \le 14$) and a window index ($1 \le t \le T$. T: The total number of windows in an acquired EEG signal), respectively [8]. $pow\_\delta(t, p)$ is defined as the δ band energy in the t-th window from the p-th pin. In a similar way, $pow\_\theta(t, p)$, $pow\_\alpha(t, p)$, $pow\_\beta(t, p)$, $pow\_\gamma(t, p)$ are defined. fv_t' is defined as the set of the five band energies from 14 pins in the t-th window. As a result, the dimension of fv_t' is 70. The set of feature vectors FV for preference classification is presented as:

$$FV = \{ fv_1', \dots, fv_t', \dots, fv_N' \} \quad 1 \le t \le T$$

$$fv_t^i = (pow_{\theta(t,1)}, pow_{\alpha(t,1)}, pow_{\beta(t,1)}, pow_{\gamma(t,1)}, pow_{\delta(t,1)}, \dots, pow_{\theta(t,14)}, pow_{\alpha(t,14)}, pow_{\beta(t,14)}, pow_{\gamma(t,14)}) \tag{2}$$

The MusicRecom constructs a personalized preference model for a single user. In the personalize preference model, a user determines his preference about music clip after listening music clips. Preference depth is classified as the negative preference and positive preference. fv_t' in each music clip are scattered to 70 dimensional spaces with their corresponding preference classes. An SVM with

Journal of Theoretical and Applied Information Technology
15th March 2018. Vol.96. No 5

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

an RBF kernel is applied to classify these two preference classes [13]. In testing, the proposed EEG music recommender gathers fv_t' for 8 seconds while a user listens to a music clip. Each fv_t' classified as the 'positive' or 'negative' preference class using the user's personal preference model. The EEG music recommender resolves the dominant emotional preference as the most frequently counted preference for the music clips.

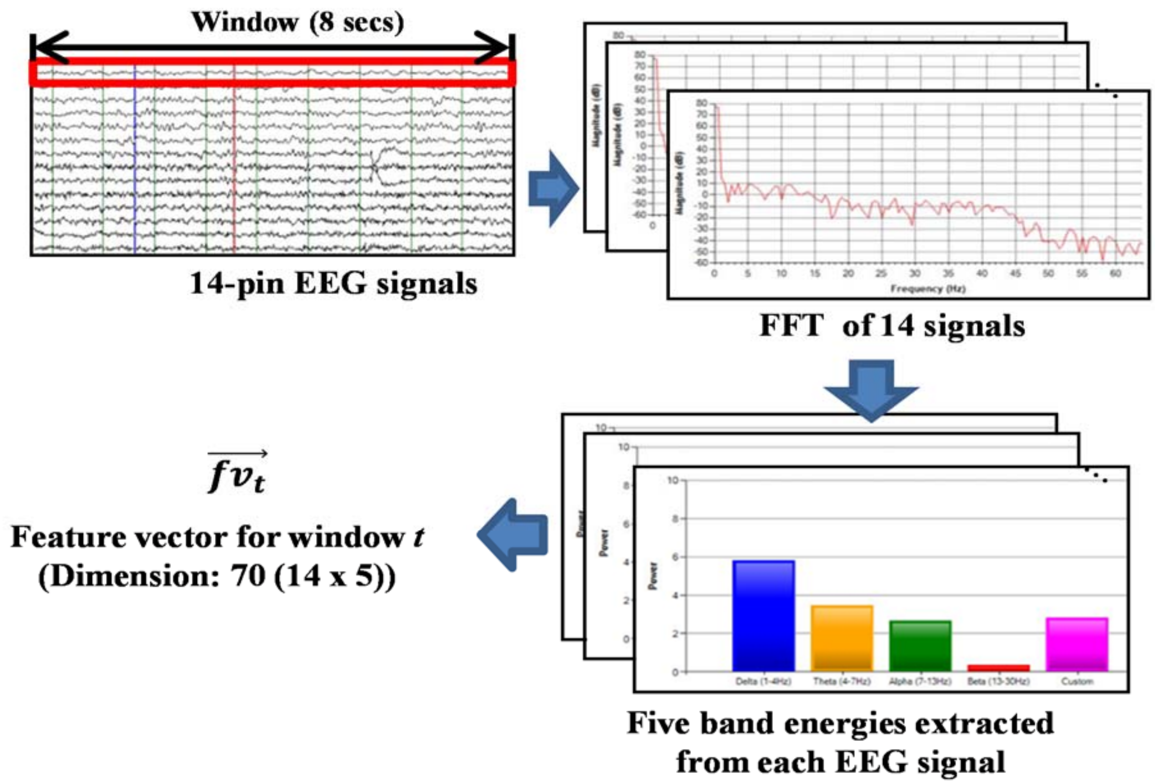

Figure 3: Process of feature extraction from EEG signals [7].

2.2. Genre Feature Classification

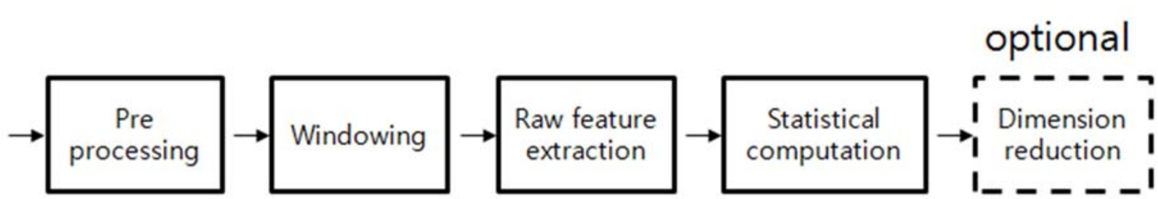

Figure 4: Feature extraction process

The feature extraction process of our system is shown in Figure 4. First, input audio is pre-processed with decoding, down-sampling, and mono-conversion. The pre-processed audio is framed using a Hamming window of about 23ms with 50% overlap. From each window, raw features are obtained. In our system, Mel-frequency cepstral coefficients (MFCC), Decorrelated filter bank (DFB), and Octave-based spectral contrast (OSC) are used. MFCCs represent the spectral characteristics based on Mel-frequency scaling, and they are also used in various music classification systems. The DFB considers the variation of amplitudes between neighboring bands. It is extracted from subtraction of log spectrum in neighboring Mel-scale band. The OSC considers the spectral peak, spectral valley, and spectral contrast in each octave-based subbands. After raw feature

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

extraction, length of the feature vector is 42: 13 for MFCCs, 13 for DFB, and 16 for OSC. After extracting 3 features, their statistical values are computed in order to represent temporal variation.

Table 2: Genre Feature Vector

Raw feature	Statistical value	Dimension
MFCC	Mean	13
	Variance	13
	FMSFM	13
	FMSCM	13
DFB	Mean	13
	Variance	13
	FMSFM	13
	FMSCM	13
OSC	Mean	16
	Variance	16
	FMSFM	16
	FMSCM	16
SUM		168

We used mean, variance, feature-based modulation spectral flatness measure, and feature-based modulation spectral crest measure [14]. Length of the feature vector is quadrupled, and we get a feature vector of dimension 168. This 168-dim vector can be used as it is, but sometimes, the feature vector is used after dimension reduction depending on system design. For an application using low-computational power, short feature vector is necessary. To reduce feature dimension without performance degradation, distance metric learning is applied in our system [15]. Feature dimension is reduced by linear projection, and to get the projection matrix, we used a distance metric learning method – MCML [16]. Generally, distance metric learning algorithm including MCML assumes the general form of Mahalanobis distance.

In the past research[17], with the objective of making short feature vector for music genre classification, we applied low dimensional projection on already developed feature set, and the projection is obtained from a distance metric learning algorithm. We can get over 80% of accuracy with only 10-dimensional feature vector. But if we use genre decision for recommendation, the distance of music is ignored. In this paper recommendation satisfaction is compared when we use distance of music and when we use genre decision. We used 10-dim feature vector using MCML for genre classification and recommendation. The cosine coefficient of feature vector is used to compare to distance of two songs.

2.3. Recommendation Engine

After user preference modeling using brainwaves, the system can know which content is suitable or not. The recommendation engine suggests songs similar with user's preferred songs based on singer, composer and genre. We use the cosine coefficient to compare user's preferred songs and music datasets. The recommendation engine extracts user preference features and dataset features. In Equation (2), we can control recommendation focus through preference weight. If user doesn't want recommendation by singers, then singer preference weight can be set 0. The sum total of preference weight is 1. Default preference weight values of music are 2/10(singer), 1/10(composer), and 7/10(genre). Preference weight has priority order, followed by genre, singer and composer. If PrefVal is same, than genre similarity is considered as first.

$$\mathrm{PrefVal_i} = \sum_{k=0}^{3} \mathrm{w_k} \times \mathrm{Sim_k}(\mathrm{D_i}, \mathrm{T_i}) \qquad (2)$$

$\mathrm{D_i}$ is the music feature vector lists of user prefered music and $\mathrm{T_i}$ is the music feature vector lists of dataset. At last recommendation engine sort candidate songs by preference value and recommend top 20 songs.

3. EVALUATION AND IMPLEMENTATION

3.1. User Interface

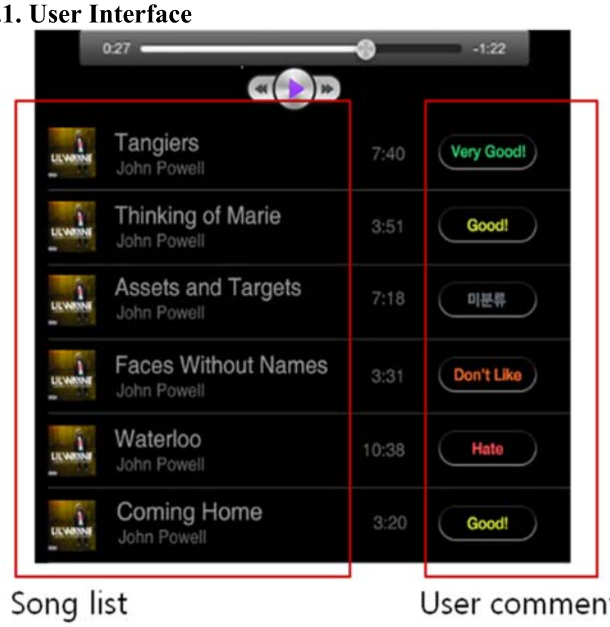

Figure 5: Implementation results of the MusicRecom system

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

For the system evaluation as shown in Figure 5, a music corpus, KETI AFA2000, which contains approximately 2,400 pop mp3 clips, was used [18]. The dataset consists of 8 genres like dance, ballad, trot, children, rock, R&B, pops and carol. Length of each play varies between 4 and 6 minutes except for children and carol. We had gathered usage history for 5 days in the MusicRecom system, and we learned the user preferences for 4 users. The assumption in the evaluation is that music which each user played or commented is each user's answer system, and we compared these answers and the recommended music by the MusicRecom system

3.2. Preference Classification accuracy

To evaluate the proposed system, an EEG dataset is constructed for music recommendation. EEG data and the user favorite music lists were gathered. Each participant reviewed the music list by listening to thirty clips, and selected his or her ten favorite clips and ten least favorite clips from the AFA2000 corpus [8]. The preference modeling module gathered the EEG responses to the selected clips. One-minute EEG signals were extracted for each of the twenty clips. First, the participants listened to their ten favorite music clips; then, they listened to the ten least favorite music clips [8]. The extracted dataset consisted of approximately 260 EEG signals of one minute in duration from the thirteen participants. The age distribution of the participants who contributed to the extraction of EEG data is listed in Table 3.

Table 3: Participant age distribution in EEG data construction

Age	Number of Participants	
	Male	Female
20s	3	2
30s	4	3
40s	1	-
Total	8	5

For EEG acquisition, an Emotiv 14-pin and wireless EEG headset was used [19]. This headset is designed to use 14 specific sensor positions shown in Figure 2: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4 [19]. These fourteen sensors can show in Figure 2. The device acquires multi-channel EEG signals wirelessly. These accuracies were measured by using ten-fold cross validation. The proposed system achieved an 81.07% accuracy rate when the classifier uses all the EEG signals from the fourteen pins. Table 4 summarizes this result.

Table 4: Preference classification accuracy by applying 14 EEG signals using KETI AFA2000

Feature	Accuracy (%)
14-pin EEG signals	81.07

3.3. Recommendation Satisfaction

The MusicRecom suggests 20 songs to participant after analyzing about seed preferred songs. We had gathered usage history for 4 persons in 5 days. In the past research [20], we used usage history for analyzing user preference. In order to evaluate the user's satisfaction about the MusicRecom system, we proposed MRR-based evaluation method given Equation (3) and MAP (Mean Average Precision) in Equation (4) . MRR (Mean Reciprocal Rank) is a standard evaluation measure in web-based information retrieval. This is the related measures, in that they are exactly equivalent for queries with one correct answer.

$$MRR = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{r_i} \quad (3)$$

$$MAP = \frac{\sum_{q=1}^{Q} AveP_{(q)}}{Q} \quad (4)$$

In Equation (3), r_i means the ranking in the recommended list in by the proposed user recommendation engine. Table 5 shows the evaluation results of the MusicRecom recommended contents. In Table 5, Test1 used automatic genre classification (10-dim feature vector) and Test2 used the feature vector distance between seed songs and target songs (10-dim feature vector). If we assumed the web pages in information retrieval are same with the contents in multimedia service, it is possible to check the rank where the first correct answer appeared. If there were no correct answer in the recommended list, Reciprocal Rank is 0. MRR is a mean over n contents. In Equation (4), MAP summarizes rankings from multiple by averaging average precision. Table 5 shows the evaluation results about recommended contents. The average MRR and MAP in Table 5 tells that the newly proposed method increases the users' overall satisfaction about the recommended contents.

Table 5: Evaluation results with MRR

	A		B		C		D		Average	
	MAP	MRR	MAP	MRR	MAP	MRR	MAP	MAP	MAP	MRR
Test 1	0.76	0.83	0.60	0.65	0.60	0.77	0.39	0.31	0.59	0.64
Test 2	0.81	0.88	0.72	0.71	0.70	0.83	0.54	0.44	0.69	0.71

4. CONCLUSION

In this paper, we proposed a recommendation system based on a preference classification using real-time user brainwaves and genre feature classification. Proposed user's preference classifier achieved an overall accuracy of 81.07% in the binary preference classification for the KETI AFA2000 music corpus. And we could recognize the user's satisfaction when we use brainwaves. This system can be applied to various audio devices, apps and services. This paper proposed the advanced method about the automatic management of the user preferences and genre classification based on the song's feature vector. We compared similarity between using decided genre value and using feature vector distance.

5. ACKNOWLEDGEMENT

This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2017

REFERENCES

[1] Ryu Ji-Woong, Bae Bitnara, Kim Mun-Choel, Nam Jae-Ho, Kang Kyung-Ok: Intelligent program guide for digital broadcasting. Conference of the society of broadcasting engineers (2001)

[2] A. Jeyasekar, K. Akshay , Karan. Collaborative Filtering using Euclidean Distance in Recommendation Engine. Indian Journal of Science and Technology. 2016 Oct;9(37):1-5.

[3] Kinam Park. The Potential Knowledge Recommendation System using User's Search Logs. Indian Journal of Science and Technology. 2016 June;9(24):1-7.

[4] Ryu Ji-Woong, Kim Mun-Choel, Nam Jae-Ho, Kang Kyoengok, Kim Jinwoong: User preferences based on intelligent program guide. Journal of the society of broadcasting engineers, Vol.7 NO.2 153-167 (2002)

[5] G. Nam, B. Kang, K. Park," Robustness of Face Recognition to Variations of Illumination on Mobile Devices Based on SVM," Korea Society for Internet Information Trans. on Internet and Information Systems, vol.4, no.1, pp. 25-44, 2010.

[6] C. Song, H. Park, C. Yang, S. Jang, and S. Lee, "Implementation of a practical query-by-singing/humming (QbSH) system and its commercial applications," IEEE Trans. on Consumer Electronics, vol. 59, no. 2, pp. 407-414, 2013.

[7] R. Panda and R. Paiva, "Music emotion classification: analysis of a classifier ensemble approach," in Proc. International Workshop on Music and Machine Learning in conjunction with the International Conference on Machine Learning, Edinburgh, Scotland, 2012.

[8] Saim Shin, Sei-Jin Jang , Donghyun Lee, Unsang Park, Ji-Hwan Kim," Brainwave-based Mood Classification Using Regularized Common Spatial Pattern Filter," KSII Transactions on Internet and Information Systems (TIIS), Vol.10 No.2, 2016.2, 807-824 (18 pages).

[9] Q. Zhang and M. Lee, "Analysis of positive and negative emotions in natural scene using brain activity and GIST," International Journal of Neurocomputing, vol. 72, no. 4-6, pp. 1302-1306, 2009

[10]Z. Wei, "Utilizing EEG signal in music information retrieval," Master Thesis, Department of Computer Science, National University of Singapore, 2010.

[11]S. Koelstra, C. Muehl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: a database for emotion analysis; using physiological ignals," IEEE Trans. Affective Computing, vol. 3, no. 1, pp. 18-31, 2011.

[12]R. Gilmore, "American electroencephalographic society guidelines in electroencephalography, evoked potentials, and polysomnography," Journal of Clinical Neurophysiol, vol. 11, no. 1, pp. 1-147, 1994.

[13]G. Nam, B. Kang, K. Park," Robustness of Face Recognition to Variations of Illumination on Mobile Devices Based on SVM," Korea Society for Internet Information Trans. on Internet and Information Systems, vol.4, no.1, pp. 25-44, 2010.

[14] S.-C. Lim, S.-J. Jang, S.-P. Lee, and M. Y. Kim, "Music genre/mood classification using a feature-based modulation spectrum," in Proc. IEEE Int. Conf. Mobile IT Convergence, 2011, pp. 133-136.

[15]D. Jang and S.-J. Jang, "Very short feature vector for music genre classification based on distance metric learning," in Proc. ICALIP, 2014

[16]A. Globerson and S. Roweis, "Metric learning by collapsing classes," Adv. Neural Inf. Process. Syst., vol. 18, pp. 451–458, 2006.

[17]D. Jang and S.-J. Jang, "Very short feature vector for music genre classification based on distance metric learning," in Proc. ICALIP, 2014

[18]C. Song, H. Park, C. Yang, S. Jang, and S. Lee, "Implementation of a practical query-by-singing/humming (QbSH) system and its commercial applications," IEEE Trans. on Consumer Electronics, vol. 59, no. 2, pp. 407-414, 2013.

[19]"Emotiv software development kit - user manual for release 1.0.0.4," Emotiv, 2011.

[20]Jongseol Lee, Saim Shin, Dalwon Jang, Sei-Jin Jang, Kyoungro Yoon, "Music recommendation system based on usage history and automatic genre classification," in Proc. IEEE International Conference on Consumer Electronics, Las Vegas, USA, pp. 134 - 135, Jan. 2015.