

Extraction of User Preference for Video Stimuli Using EEG-Based User Responses

Jinyoung Moon, Youngra Kim, Hyungjik Lee, Changseok Bae, and Wan Chul Yoon

Owing to the large number of video programs available, a method for accessing preferred videos efficiently through personalized video summaries and clips is needed. The automatic recognition of user states when viewing a video is essential for extracting meaningful video segments. Although there have been many studies on emotion recognition using various user responses, electroencephalogram (EEG)-based research on preference recognition of videos is at its very early stages. This paper proposes classification models based on linear and nonlinear classifiers using EEG features of band power (BP) values and asymmetry scores for four preference classes. As a result, the quadratic-discriminant-analysis-based model using BP features achieves a classification accuracy of 97.39% ($\pm 0.73\%$), and the models based on the other nonlinear classifiers using the BP features achieve an accuracy of over 96%, which is superior to that of previous work only for binary preference classification. The result proves that the proposed approach is sufficient for employment in personalized video segmentation with high accuracy and classification power.

Keywords: Preference, video, EEG, classification, feature selection, brain-computer interface.

I. Introduction

There is a flood of video content from the Internet and the many television channels available through terrestrial, cable, and satellite systems. The average length of video viewing per person exceeds twenty hours per week in some European countries and in the United States of America [1]. There is an increasing tendency for viewers to re-watch impressive sections of videos, recall memorable information, or share interesting parts with their acquaintances. Therefore, people require video clips or summaries of video segments that are meaningful to them to efficiently obtain access to the viewed videos. It is cumbersome for the average viewer to extract significant video segments or generate annotations of them by manually using a video editing tool. Therefore, the automatic recognition of user states during video viewing is essential for automatically generating personalized clips and summaries.

There have been some studies on the automatic recognition of user states, such as interest and attention, by analyzing user responses to viewed videos. To measure the degree of user interest in viewed videos, such studies used the fact that users show their current feelings through facial expressions, body movements, and various peripheral responses [2]-[5]. Joho and others [2] proposed facial expression models based on three pronounced levels of facial expressions and their change rate. Although facial expressions are obtained from an observing camera in an unobtrusive way, it is hard for people to accurately interpret facial expressions. In general, people have neutral facial expressions. Peng and others [3] proposed an interest meter for measuring the interest score of a user for home videos by analyzing a captured video containing the user's upper body. Their framework recognized facial expressions for their emotion model and detected head motions,

Manuscript received Feb. 28, 2013; revised Aug. 20, 2013; accepted Sept. 16, 2013.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2010-0028631, 13SS1110).

Jinyoung Moon (phone: +82 42 860 6712, jymoon@etri.re.kr) is with the Software Research Laboratory, ETRI, Daejeon, Rep. of Korea and is also affiliated with KAIST, Daejeon, Rep. of Korea.

Youngra Kim (youngra@etri.re.kr), Hyungjik Lee (leehj@etri.re.kr), and Changseok Bae (csbae@etri.re.kr) are with the Software Research Laboratory, ETRI, Daejeon, Rep. of Korea.

Wan Chul Yoon (corresponding author, wcyoon@kaist.ac.kr) is with the Department of Industrial and System Engineering, KAIST, Daejeon, Rep. of Korea.

blinks, and saccades for their attention model. Because people do not always give significant facial expressions in home videos, they added an analysis of eye and head movements. However, the highs and lows of the interest score could not segregate the entire video into funny/attentive or other segments, although the interest scores were relatively high in the funny/attentive video segments during their experiment. Money and Agius [4], [5] presented an analysis framework for processing peripheral signals, including electrodermal response, respiration amplitude, respiration rate, blood volume pulse, and heart rate. The peripheral signals delicately reflect changes of user states compared to their facial expressions and body movements. The framework clarified the characteristics of each peripheral signal for each genre of film and suggested a rank value representing the significance of the user response. However, the rank values were useful only for summarizing videos in the comedy and horror genres, which arouse noticeably significant user responses.

An electroencephalogram (EEG) is a recorded signal of the electrical activity of the brain influenced by the central nervous system. This type of signal is recorded from the scalp and is different from a peripheral signal, which is influenced by the autonomic nervous system. Because the brain has been recognized as the center for cognitive activities in cognitive science and psychology [6], there have been many studies on analyzing user states through the use of EEG signals [7]-[20]. For emotion recognition, the studies provided subjects with emotion-inducing visual or auditory stimuli, that is, images, sounds, music excerpts, and video clips [7]-[18]. The studies classified the collected EEG signal segments into selected primitive emotion types, which are distributed over the valence-arousal emotion space in the two-dimensional emotion model [21]. The EEG-based methods for emotion recognition achieved similar or better classification accuracy compared to the method based on the peripheral signals [16]-[18]. For attention measurement, the studies classified EEG signals into concentrated and non-concentrated states by having the subjects perform concentration and non-concentration tasks [19] and hold a focused and unfocused gaze [20].

Extracting meaningful video segments requires a method to measure the preference toward videos under the assumption that preferred video segments are more significant to viewers than other video segments. However, the proposed methods for analyzing user states on emotion and attention [7]-[20] are not enough to measure user preference toward videos because induced positive-valence and high-arousal emotions do not always create high preference. For example, many people have been attracted to sad music or sad movies, which evoke negative-valence emotions. In addition, horror movies are popular with some people but not with others. Some people

have preferences for negative emotion in music or videos for various reasons, and all people have their own optimal arousal levels at which they feel comfortable [22], [23].

Compared to many studies on EEG-based emotion recognition, research on EEG-based recognition of user preference is at its very early stage. Aurup [24] targeted product preference induced by product images. Product preference inducing purchases has a close relationship with strong pleasantness corresponding to positive-valence and high-arousal emotions. Hadjidimitriou and Hadjileontiadis [25] performed an EEG-based preference experiment using music as audio stimuli. For the first time, Koelstra and others [17] achieved an average accuracy of 57.9% for like/dislike binary classification using two-minute clips of music video as stimuli and EEG samples collected from five subjects. In their next publication, Yazdani and others [18] achieved an average accuracy of 70.25% using the same stimuli and samples, which was better than that of their previous result. However, their method was designed only for binary preference. In addition, an accuracy of 70.25% is not satisfactory for personalized video segmentation in video applications.

Therefore, we focus on an approach to assess not the emotion aroused but the preference for video stimuli by analyzing EEG signals collected during video viewing. We propose classification models based on linear and nonlinear classifiers using band power (BP) and asymmetry score (AS) features, which are targeted for the following four preference classes: most preferred, preferred, less preferred, and least preferred. All models based on nonlinear classifiers using the BP features of all the frequency bands achieve over a 96% accuracy, which is an outstanding performance compared to previous work [17], [18]. Additionally, the accuracy is maintained in the models using only 43% to 70% of all BP features, which are reduced by a filter method for feature selection. The results prove that the proposed approach is sufficient to be employed for personalized video segmentation with high accuracy and classification power.

II. EEG-Based Methods for Extracting Preference

The proposed methods of this study follow the typical procedure for EEG signal analysis, as shown in Fig. 1.

1. EEG Data Acquisition

EEG data is collected from fifteen healthy right-handed subjects (eight males and seven females) whose ages range from 23 to 43. All of the subjects are graduate students or workers for a research institute.

The video stimuli are designed to establish ground truth

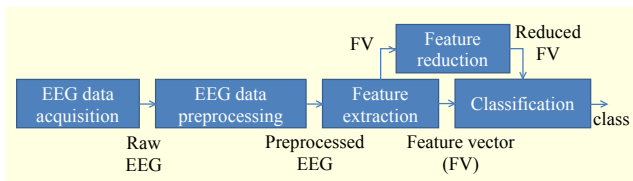


Fig. 1. Typical procedure for classifying user states using EEG signals.

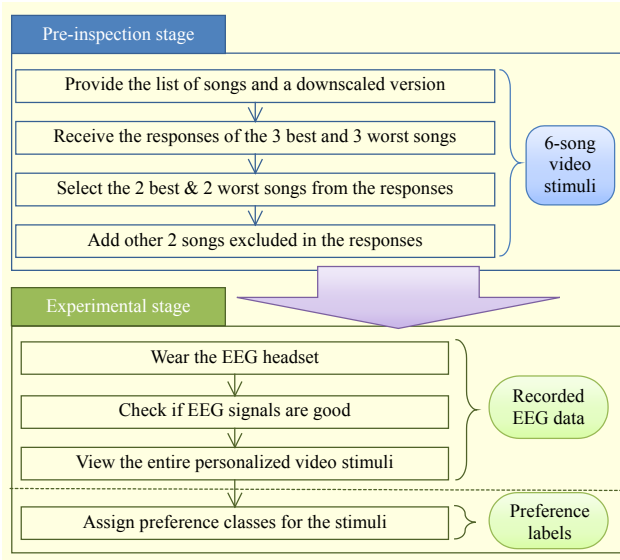


Fig. 2. Experimental procedure consisting of pre-inspection stage for personalized video stimuli and experimental stage for EEG data and preference labels.

labels on video datasets for measuring user preferences. Television music shows are selected as the source of the video dataset because they can be definitively segmented with a song as the unit, and the segmented songs can be clearly ranked by each viewer according to their own preference. Twenty-one Korean pop songs around three minutes long are extracted from music show programs from terrestrial television networks with the full HD resolution (1920×1080). The video dataset contains songs of different genres, including ballads, dance, pop rock, hip hop, electro pop, and swing.

The experimental protocol of this study is divided into pre-inspection and experimental stages, as shown in Fig. 2. The pre-inspection stage is necessary to generate personalized visual stimuli reflecting subject preference for a video dataset. First, subjects receive an explanation for the goal of the pre-inspection and experiment. Second, each subject completes a questionnaire to select the three best and three worst songs out of the twenty-one videos. The list of the twenty-one songs, including their titles and singers, and a downscaled version (480×320) of the twenty-one videos are provided to the subjects. Subjects who are unfamiliar with the latest Korean pop songs can use the downscaled version for a

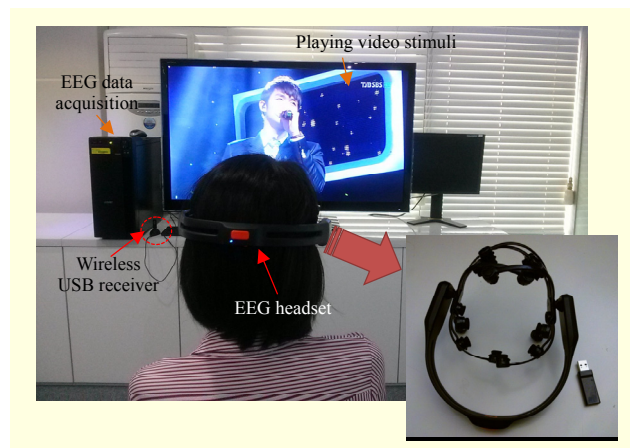


Fig. 3. Experimental environment. While subjects wearing EEG headset view their personalized video stimuli, raw EEG data from EEG headset is accumulated through wireless transmission.

preview of all videos. Third, the video stimuli includes the two best and two worst songs chosen from the questionnaire responses for the well-distributed preference level of the stimuli. Fourth, the video stimuli add two other songs selected randomly from the rest of the fifteen songs excluded in the questionnaire responses. Finally, the personalized video stimuli consist of six songs out of the twenty-one songs of the video dataset.

In the experimental stage, a subject and two operators of the experiment are allowed to access the laboratory. While wearing an EEG headset, the subject receives explanations for the goal of the experiment, the EEG device, the experimental procedure, the time required, and the personalized stimuli. The purpose of the explanations is to reduce the subject's anxiety about the experiment using the unfamiliar EEG device. In addition, the subject is instructed to concentrate on viewing the video stimuli in a nearly static position. After checking whether the EEG signals from all electrodes of the EEG headset are sufficient, the experimenter has the subject watch a 33-second introduction of a music show and then six video segments corresponding to six Korean pop songs, as shown in Fig. 3. As there are no intermissions or neutral videos between the video segments, it takes about 20 minutes for each subject to finish the experiment. The EEG data of the subject is gathered from the EEG headset while the subject views the personalized stimuli.

As shown in Fig. 4, the EEG device collects the subject's EEG signals from the following fourteen channels of wet electrodes placed on the scalp: AF3, AF4, F7, F8, F3, F4, FC5, FC6, T7, T8, P7, P8, O1, and O2. EEG signals are recorded by the EEG device at a sampling rate of 128 Hz. The bandwidth of the recorded EEG signals is from 0.2 Hz to 45 Hz using a

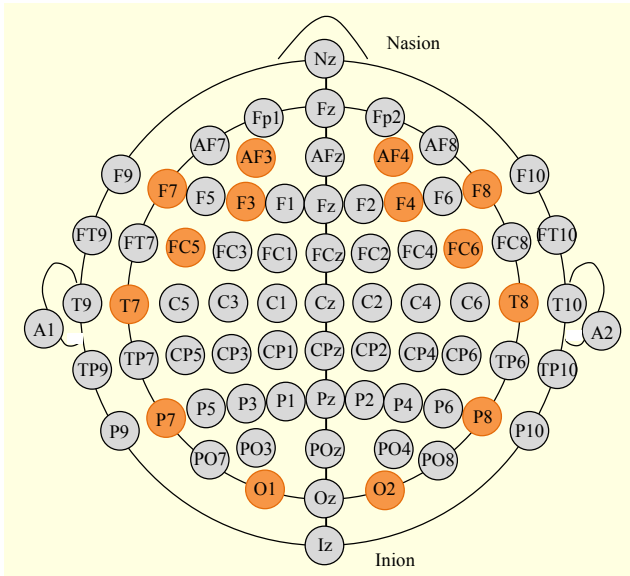


Fig. 4. Placement of fourteen electrodes used in international 10-20 system, which is internationally recognized method to describe location of electrodes on scalp for EEG experiments or applications [26].

0.16-Hz high-pass filter, an 83-Hz low-pass filter, and 50-Hz and 60-Hz notch filters. The notch filters are employed for removing the power line artifacts. The environmental artifacts caused by the amplifier and aliasing are removed from the hardware [27].

To employ the proposed approach when extracting video segments meaningful to users, this study subdivides the binary preference into two-level positive and two-level negative preferences. One of the ground truth labels (most preferred, preferred, less preferred, or least preferred) is assigned to the video stimuli by each subject immediately after viewing all video stimuli. The preference for a video is influenced by both the video itself and the context of videos given together and the sequential order among them in the general environment of video viewing. Accordingly, the subject can accurately assign a preference label for one song according to the degree of preference, not locally but globally after viewing all stimuli. There is no heavy cognitive load to the subject when the subject assigns preference labels because the subject is already familiar with the four songs chosen from the questionnaire responses among the personalized stimuli. Furthermore, the total number of video stimuli does not exceed the capacity of the subject's working memory [28].

2. EEG Data Preprocessing

EEG signal preprocessing means that recorded EEG data is prepared for feature extraction and further analysis. As shown in Fig. 5, the 33-second introduction of a music show and the

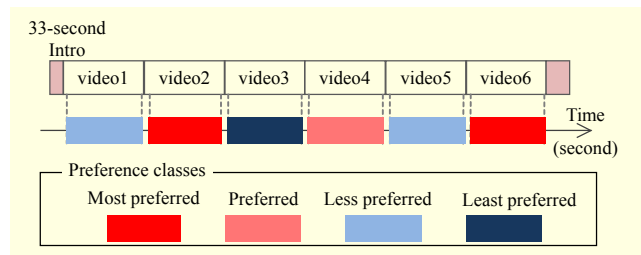


Fig. 5. Preprocessing EEG data. Thirty-three-second introduction of music show and one-second start and end of each video segment are excluded. Each video segment is assigned to one preference class, that is, most preferred, preferred, less preferred, or least preferred.

one-second start and end of each video segment are extracted from the recorded EEG data. The preference classes of the remaining EEG data are assigned according to the results of the questionnaire completed by the subject. In the baseline removal, the mean of the EEG signal from each channel is subtracted from the EEG signal for each channel so that all EEG signal values across the channels are distributed to around zero.

The biological artifacts resulting from blinks and eyeball movements are avoided by excluding artifact-influenced frequency bands in the feature extraction. To prevent artifacts by muscle and eyeball movements, this study instructs subjects to view the video stimuli in front of the television in a nearly static position. Because it is impossible to prevent all biological artifacts from the live subjects, additional effort is necessary to minimize the influence of the biological artifacts after the EEG recording stage. The influence of biological artifacts can be reduced by automatic or manual artifact removal or by the rejection of artifact-related frequency bands. The artifacts by heartbeat and artifacts by blinks and eyeball movements appear at around 1.2 Hz and below 4 Hz, respectively. In contrast, the artifacts by muscle movements are most dominant above 50 Hz [7], [11]. A method to reject the frequency bands affected by such artifacts is employed in this study because the time-consuming artifact removal is inappropriate for a real-time signal analysis.

The EEG values of each channel are filtered by a Butterworth filter with a passband between 4 Hz and 40 Hz, which is an artifact-resilient frequency range. The Butterworth filter is employed because it was designed to have as flat a frequency response as possible in the passband [29].

3. Feature Extraction

One of the most common approaches for investigating EEG data is to analyze the activated and inactivated power values or their changes at significant frequency bands, which are delta (<4 Hz), theta (4 Hz to 7 Hz), alpha (8 Hz to 13 Hz), beta

Table 1. Extracted features and number of features.

Feature	No. of features
BP	56 = 14 channels × 4 bands
AS	28 = 7 channels × 4 bands
BPAS (both BP and AS)	84 = 21 channels × 4 bands

(14 Hz to 30 Hz), and gamma (>30 Hz) [30]. The frequency range of each frequency band varies little in the EEG literature. When the EEG data is separated according to the range of the frequency bands, the frequency bands provide more information on the neural activities of normal people or a sign of psychological disorders. In this study, four frequency bands in the range of 4 Hz to 40 Hz are selected as follows: theta (4 Hz to 7 Hz), alpha (8 Hz to 13 Hz), beta (14 Hz to 30 Hz), and gamma (31 Hz to 40 Hz).

The total power values of each frequency band and each channel are typically extracted as features in the frequency domain by using a discrete Fourier transform (DFT). The DFT coefficient $X(k)$ of a one-second non-overlapping EEG segment, whose length is N , is obtained by (1). The commonly used fast Fourier transform algorithm is adopted. The power spectrum is calculated by the square of the absolute value of $X(k)$ [31].

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot \exp(-j \frac{2\pi}{N} kn), \quad k = 0, 1, \dots, N-1. \quad (1)$$

For four different frequency bands per channel, the total power values are obtained by the sum of the power values within the range of each frequency band. The total power of each band is transformed in a natural log scale because the power values are usually positively skewed [32].

In addition, there are features derived using the asymmetry resulting from the spectral difference between the power values of EEG electrodes that constitute a symmetrical pair. There have been many studies conducted on the relationship between hemispheric asymmetry and emotion since 1979 [33]. In [32], the features were extracted from (2), wherein the right and left EEG power spectra of a symmetric pair are P_R and P_L , respectively.

$$\text{Asymmetry score} = \ln(P_R) - \ln(P_L). \quad (2)$$

For four different frequency bands per channel, the ASs are also extracted as features. All extracted features are listed in Table 1.

The signal smoothing for extracted feature values is aimed at mitigating fluctuations of the power values and reducing the remaining noises. An autoregressive moving average filter with

a five-second window is employed to smooth the power values. Finally, all smoothed features are normalized with the standard score before classification. The normalization by the standard score generates the new feature values of each frequency band from each channel with a mean of 0 and a standard deviation of 1. The normalization is required to reduce the dependency on features with high values or a large margin.

4. Classification

This study employs five classifiers for recognizing four preference classes (most preferred, preferred, less preferred, and least preferred) and comparing the classification accuracies of the classifiers used. The classifiers include three nonlinear classifiers, that is, the k nearest neighbors, a support vector machine (SVM) with a radial basis function (RBF) kernel, and a quadratic discriminant analysis (QDA), and two linear classifiers, that is, an SVM with a linear kernel and a linear discriminant analysis (LDA).

The k nearest neighbor algorithm (k -NN) is an intuitive method for assigning a test instance to the dominant class among its k nearest neighbors in Euclidean distance using the majority vote [11], [14], [31], [34]. This study selects a value of 4 for k , which is a small positive integer, because the maximum number of BPAS features and the curse of dimensionality in k -NN are considered.

An SVM is one of the most popular supervised learning techniques for classification and regression [8]-[11], [13], [15], [27], [31], [34], [35]. An SVM performs classification with an N -dimensional separating hyperplane constructed for minimizing the classification errors on the test set and maximizing its margins. To compare the performance of linear and nonlinear classifiers, this study employs two SVMs with a linear kernel and an RBF kernel. An SVM with a linear kernel requires the value of the cost parameter, and an SVM with an RBF kernel needs both cost and γ parameters. This study uses default values for the parameters provided by the employed SVM library, which are $\text{cost} = 1$ and $\gamma = 1/(\text{data dimension})$, without the tuning of the parameters by cross-validation for determining the best combination of cost and γ parameters because it is a time-consuming process, inadequate for real-time analysis.

LDA is a method to find a linear combination of features to express a categorical independent variable, such as class, which is contrary to a numerical independent variable of a linear regression analysis. Because of its low computational requirement, LDA has commonly been used for systems using EEG signals [10], [14], [31]. QDA [10], [11], [31] is a more general version of LDA using a quadratic surface, which has no assumption that each class is normally distributed. In both

LDA and QDA, there are no parameters to be set.

In the experiment, more than 1,080 instances (six video segments multiplied by about 180 one-second EEG segments per a video segment multiplied by one instance per a one-second EEG segment) are obtained from each subject. All training instances are labeled using preference classes assigned by the subject.

To evaluate the classifiers and features used, the classification accuracy of a model based on each classifier using each feature set is measured by averaging the classification accuracies of subject-dependent classification models from the fifteen subjects. The classification accuracy of a subject-dependent model is obtained from a threefold cross-validation scheme. The k -fold cross-validation scheme partitions instances into k subsets. One subset is used to validate the generated model trained by the other $k-1$ subsets. The process is repeated k times. In the k -fold cross-validation, the accuracy is obtained by averaging the accuracies from k models.

For a classifier evaluation, 65 classification models based on the k -NN, an SVM with a linear kernel, an SVM with an RBF kernel, LDA, and QDA are generated. The k -NN and SVM models for theta, alpha, beta, gamma, and all the frequency bands are generated using BP, AS, and BPAS features. The LDA and QDA models for theta, alpha, beta, gamma, and all the frequency bands are generated using BP and AS features. For LDA and QDA models performing a covariance matrix inversion, this study excludes the BPAS features because BP and AS features are collinear.

5. Feature Selection

To construct efficient classification models maintaining the classification accuracy of the original models before feature selection using the subsets of only the relevant features out of all the BP features, this study employs the minimum redundancy maximum relevance (mRMR) method [36] for feature selection. The mRMR method is one of the filter methods used for proxy measures, such as mutual information and correlation, to score a feature subset. For each subject, the mRMR method generated an ordered feature list for BP features. For the generation of feature subsets, the least relevant feature is removed one by one from the ordered feature list until only two features remain. For each subject, the classification accuracies of models using the generated subsets are measured using threefold cross-validation.

III. Experiment Results

1. Classification

For a classification evaluation, this study thoroughly

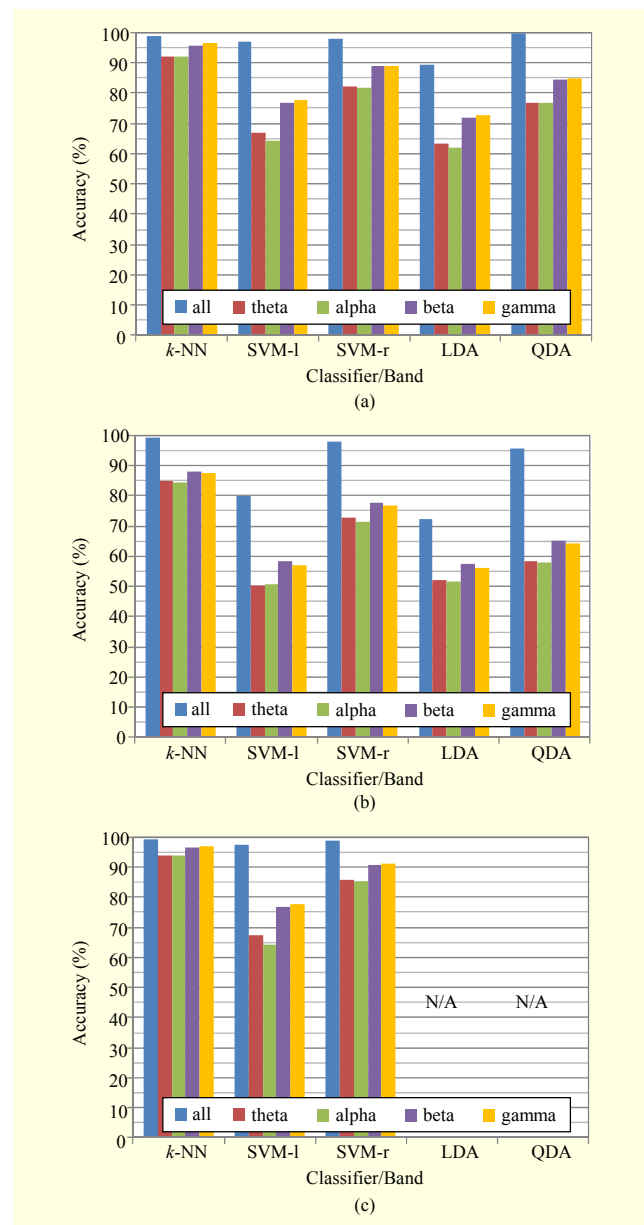


Fig. 6. Average accuracies of all thirteen models. Each “all” bar represents averaged classification accuracy of models using (a) 56 BP, (b) 28 AS, and (c) 84 BPAS features from all frequency bands. In addition, each “theta,” “alpha,” “beta,” and “gamma” bar represents average classification accuracy of models using (a) 14 BP, (b) 7 AS, and (c) 21 BPAS features using its own frequency band. There are no accuracies of LDA and QDA based models using BPAS features because BP and AS features are collinear.

investigates the relationship between classification accuracy and several factors, such as the types of classifiers, features, and frequency bands. The average accuracies of classification models from the fifteen subjects are shown in Fig. 6.

The average classification accuracies of models using nonlinear classifiers, such as the k -NN, an SVM with an RBF

kernel, and QDA, are better than those of models using linear classifiers, such as an SVM with a linear kernel and LDA, irrespective of the types of features, and, therefore, this study mainly focuses on models based on nonlinear classifiers.

Considering the type of features, this study shows that the models based on the nonlinear classifiers using all 56 BP features achieve almost the same average classification accuracy with the models based on nonlinear classifiers using all 84 BPAS features ($p < 0.01$). The best average classification accuracy of 97.99% ($\pm 1.38\%$) is obtained by the k -NN-based model using BPAS features of all the frequency bands, but there is no significant difference ($p < 0.01$) between the two models based on the k -NN and an SVM with an RBF kernel when doing so. The second-best average classification accuracy of 97.39% ($\pm 0.73\%$) is obtained by the QDA-based model using the BP features of all the frequency bands, but there is no significant difference ($p < 0.01$) among the three models based on the nonlinear classifiers using the BP features of all the frequency bands. Although there is no significant difference ($p < 0.01$) among the five models based on the nonlinear classifiers using the 84 BPAS and 56 BP features, there is a significant difference ($p < 0.01$) among the six models based on nonlinear classifiers using the 56 BP and 28 AS features. Therefore, adding AS features to the models using the BP features of all the frequency bands is not influential in enhancing their average classification accuracy.

Considering the frequency bands, this study shows three facts. First, the average classification accuracies of models using the features of the beta and gamma bands are better than those of the models using features of theta and alpha bands with a significant difference ($p < 0.05$) across features. According to the EEG literature [37], [38], gamma-band activity is enhanced by visual and auditory stimuli. The visual and auditory stimuli from the videos are considered the main factors of sufficient accuracy of the models using the gamma band features. Furthermore, brain activities in the beta band are related with busy or anxious thinking and active concentration [39]. The cognitive process of appreciating videos and rating the level of preference might influence the accuracy of models using features of the beta band. Second, there is a big accuracy gap between models using the features of all the frequency bands and each specific frequency band across the features. Finally, the average classification accuracies of models using the features of each specific frequency band are lower than 90%, except those of the k -NN-based models using the BP and BPAS features of the beta and gamma bands. Therefore, this study confirms that an accuracy of over 96% is achieved by models with a combination of one of the three nonlinear classifiers and the BP features of all the frequency bands.

Finally, this study proves the difference between preference

toward video stimuli and emotion evoked by the video stimuli by comparing the average accuracies between two subject groups. One group, P, consists of eight subjects (subject#2, subject#3, subject#4, subject#9, subject#10, subject#12, subject#14, and subject#15), who mapped songs inducing opposite emotions, a sad ballad and an exciting dance, together into the same category of preference class. The other group, E, consists of the other seven subjects. In all the thirteen models using the BP, AS, and BPAS features of all the frequency bands, there is no significant difference ($p < 0.01$) between the accuracy means of groups P and E as determined by one-way analysis of variance. Therefore, this study shows that the proposed models perform preference classification correctly regardless of the emotions induced by the stimuli.

2. Feature Selection

For each subject, this study evaluates the classification accuracies of models using each subset of the most relevant features determined by the mRMR feature selection method to show the feasibility of the reduced models using less features but maintaining sufficient accuracy of the original model.

The evaluation is focused on models based on the nonlinear classifiers using each subset of the BP features, as the BP features showed similar classification accuracy to that of models using the BPAS features. Figure 7 shows the average classification accuracies of the reduced models using each subset including the most relevant features for two subjects, which are for the best and worst cases, respectively. Although this study shows two accuracy plots of the reduced models for the two subjects, the accuracy plots for other subjects also reflect the trend analogous to the plots shown.

The mRMR method enables the reduced models to maintain the classification accuracy of the original models using 43% to 70% of all the BP features irrespective of the type of classifiers. The models for subject#4 are the best case for feature selection because the models using more than 24 of the most relevant features provide similar classification accuracy to that of the original models using all the BP features. Furthermore, the models using more than thirteen of the most relevant features achieve a classification accuracy of over 90%. In contrast, the models for subject#11 are the worst case for feature selection because the models using more than 39 of the most relevant features provide similar classification accuracy to that of the original models using all of the BP features. Furthermore, the models using more than 33 of the most relevant features achieve a classification accuracy of over 90%. Therefore, this study confirms that the proposed models based on the nonlinear classifiers using BP features can be optimized by the mRMR method.

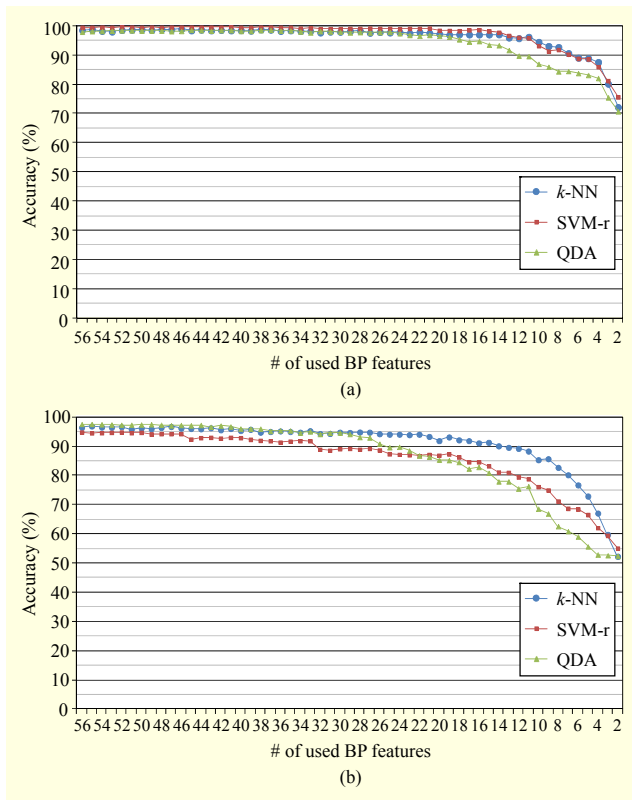


Fig. 7. Classification accuracies of models using each subset of 56 BP features. Number of features used is number of most relevant features selected by mRMR method. To maintain classification accuracy of original model using all BP feature set, reduced models for subject#4 and subject#11 require (a) smallest and (b) largest subsets of relevant BP features, respectively.

IV. Conclusion

Because a large number of videos are available from the Internet and television providers, viewers require their own personalized video summaries or video clips to efficiently access the video segments that are meaningful to them. To extract meaningful video segments according to the degree of user preference, we proposed an approach to extract the two-level positive and two-level negative preferences of users using the features extracted from EEG signals collected during video viewing.

For each subject, 65 classification models based on two linear and three nonlinear classifiers were generated using the BP, AS, and BPAS features of four specific frequency bands and all the frequency bands. Although the best and second-best average classification accuracies of 97.99% ($\pm 1.38\%$) and 97.39% ($\pm 0.73\%$) were obtained by the k -NN-based model using 84 BPAS features and the QDA-based model using 56 BP features, respectively, there was no significant difference ($p < 0.01$) between the two models. The other models based on

the nonlinear classifiers using the BP features also achieved an average accuracy of over 96%. Furthermore, the models reduced by the mRMR method showed similar classification accuracy to that of the original models using only 43% to 70% of all the BP features.

The results are significant to employing our approach for extracting video segments that viewers prefer most. First, our approach showed a high accuracy sufficient for use in real-world video applications. Second, our approach can distinguish between “most preferred” and “preferred” video segments among all video segments evoking a positive preference because it was designed to classify four-level preference. Finally, the proposed models of this study can be optimized while maintaining their accuracies using the mRMR feature selection method when it is employed to commercial systems or products. In the near future, we will extend the experiments to various genres of video content, such as TV commercials, entertainment, drama, news, and sports.

References

- [1] Ericsson Consumer Lab, *TV & Video Consumer Trend Report 2011*, Ericsson, Sept. 2011.
- [2] H. Joho et al., “Exploiting Facial Expressions for Affective Video Summarisation,” *Proc. ACM CIVR*, July 2009, pp. 31:1-31:8.
- [3] W.-T. Peng et al., “Editing by Viewing: Automatic Home Video Summarization by Viewing Behavior Analysis,” *IEEE Trans. Multimedia*, vol. 13, no. 3, June 2011, pp. 539-550.
- [4] A.G. Money and H. Agius, “Analysing User Physiological Responses for Affective Video Summarisation,” *Displays*, vol. 30, no. 2, Apr. 2009, pp. 59-70.
- [5] A.G. Money and H. Agius, “ELVIS: Entertainment-Led Video Summaries,” *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 6, no. 3, Aug. 2010, pp. 17:1-17:30.
- [6] S. Karakaş and E. Başar, “Models and Theories of Brain Function in Cognition within a Framework of Behavioral Cognitive Psychology,” *Int. J. Psychophysiol.*, vol. 60, no. 2, May 2006, pp. 186-193.
- [7] D.O. Bos, “EEG-based Emotion Recognition: The Influence of Visual and Auditory Stimuli,” online paper, Department of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, Netherlands, 2008. Available: http://hmi.ewi.utwente.nl/verslagen/capita-selecta/CS-Oude_Bos-Danny.pdf
- [8] R. Horlings, D. Dacu, and L.J. Rothkrantz, “Emotion Recognition Using Brain Activity,” *Proc. CompSysTech PhD Workshop*, June 2008, article no. 6.
- [9] K. Schaaff and T. Schultz, “Towards Emotion Recognition from Electroencephalographic Signals,” *Proc. ACII*, Sept. 2009, pp. 1-6.
- [10] G. Chanel et al., “Short-Term Emotion Assessment in a Recall

- Paradigm,” *Int. J. Hum-Comp St.*, vol. 67, no. 8, Aug. 2009, pp. 607-627.
- [11] P.C. Petrantonakis and L.J. Hadjileontiadis, “Emotion Recognition from Brain Signals Using Hybrid Adaptive Filtering and Higher Order Crossings Analysis,” *IEEE Trans. Affect. Comput.*, vol. 1, no. 2, July-Dec. 2010, pp. 81-97.
- [12] Y. Liu, O. Sourina, and M.K. Nguyen, “Real-Time EEG-Based Human Emotion Recognition and Visualization,” *Proc. CW*, Oct. 2010, pp. 262-269.
- [13] Y.-P. Lin et al., “EEG-Based Emotion Recognition in Music Listening,” *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, July 2010, pp. 1798-1806.
- [14] M. Murugappan, N. Ramachandran, and Y. Sazali, “Classification of Human Emotion from EEG Using Discrete Wavelet Transform,” *J. Biomed. Sci. Eng.*, vol. 3, no. 4, 2010, pp. 390-396.
- [15] D. Nie et al., “EEG-based Emotion Recognition during Watching Movies,” *Proc. IEEE/EMBS NER*, Apr. 27-May 1, 2011, pp. 667-670.
- [16] Z. Khalili and M.H. Moradi, “Emotion Recognition System Using Brain and Peripheral Signals: Using Correlation Dimension to Improve the Results of EEG,” *Proc. IJCNN*, June 2009, pp. 1571-1575.
- [17] S. Koelstra et al., “Single Trial Classification of EEG and Peripheral Physiological Signals for Recognition of Emotions Induced by Music Videos,” *Proc. BI*, Aug. 2010, pp. 89-100.
- [18] A. Yazdani et al., “Affect Recognition Based on Physiological Changes during the Watching of Music Videos,” *ACM Trans. Interact. Intell. Syst.*, vol. 2, no. 1, Mar. 2012, pp. 7:1-7:26.
- [19] B. Hamadicharef et al., “Learning EEG-Based Spectral-Spatial Patterns for Attention Level Measurement,” *Proc. IEEE ISCAS*, May 2009, pp. 1465-1468.
- [20] J.W. Bang, E.C. Lee, and K.R. Park, “New Computer Interface Combining Gaze Tracking and Brainwave Measurements,” *IEEE Trans. Consum. Electron.*, vol. 57, no. 4, Nov. 2011, pp. 1646-1651.
- [21] J.A. Russell, “A Circumplex Model of Affect,” *J. Pers. Soc. Psychol.*, vol. 39, no. 6, Dec. 1980, pp. 1161-1178.
- [22] D. Västfjäll and T. Gärling, “Preference for Negative Emotions,” *Emotion*, vol. 6, no. 2, May 2006, pp. 326-329.
- [23] S. Garrido and E. Schubert, “Negative Emotion in Music: What is the Attraction? A Qualitative Study,” *Empir. Musicol. Rev.*, vol. 6, no. 4, Oct. 2011, pp. 214-230.
- [24] G.M.M. Aurup, *User Preference Extraction from Bio-signals: An Experimental Study*, master’s thesis, Concordia University, 2011.
- [25] S.K. Hadjidimitriou and L.J. Hadjileontiadis, “Toward an EEG-Based Recognition of Music Liking Using Time-Frequency Analysis,” *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, Dec. 2012, pp. 3498-3510.
- [26] J. Malmivou and R. Plonsey, *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*, 1st ed., New York: Oxford University Press, 1995.
- [27] O. Varol, *Raw EEG Data Classification and Applications Using SVM*, master’s thesis, Istanbul Technical University, 2010.
- [28] G.A. Miller, “The Magical Number Seven Plus or Minus Two: Some Limits on Our Capacity for Processing Information,” *Psychol. Rev.*, vol. 63, no. 2, 1956, pp. 81-97.
- [29] D.P. Allen and C.D. MacKinnon, “Time-Frequency Analysis of Movement-Related Spectral Power in EEG during Repetitive Movements: A Comparison of Methods,” *J. Neurosci. Methods*, vol. 186, no. 1, Jan. 2010, pp. 107-115.
- [30] L. Sörnmo and P. Laguna, *Bioelectrical Processing in Cardiac and Neurological Applications*, 1st ed., Waltham, MA: Elsevier Academic Press, 2005.
- [31] Z. Iscana, Z. Dokura, and T. Demiralp, “Classification of Electroencephalogram Signals with Combined Time and Frequency Features,” *Expert Syst. Appl.*, vol. 38, no. 8, Aug. 2011, pp. 10499-10505.
- [32] J.J. Allen et al., “The Stability of Resting Frontal Electroencephalographic Asymmetry in Depression,” *Psychophysiology*, vol. 41, no. 2, Mar. 2004, pp. 269-280.
- [33] J.J. Allen and J.P. Kline, “Frontal EEG Asymmetry, Emotion, and Psychopathology: The First, and the Next 25 Years,” *Biol. Psychol.*, vol. 67, no. 1-2, Oct. 2004, pp. 1-5.
- [34] W. Yun et al., “Disguised-Face Discriminator for Embedded Systems,” *ETRI J.*, vol. 32, no. 5, Oct. 2010, pp. 761-765.
- [35] T. Jabid, M.H. Kabir, and O. Chae, “Robust Facial Expression Recognition Based on Local Directional Pattern,” *ETRI J.*, vol. 32, no. 5, Oct. 2010, pp. 784-794.
- [36] H. Peng, F. Long, and C. Ding, “Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, Aug. 2005, pp. 1226-1238.
- [37] T.R. Schneider et al., “Enhanced EEG Gamma-Band Activity Reflects Multisensory Semantic Matching in Visual-to-Auditory Object Priming,” *NeuroImage*, vol. 42, no. 3, Sept. 2008, pp. 1244-1254.
- [38] M.A. Kiskeya and Z.M. Comwell, “Gamma and Beta Neural Activity Evoked during a Sensory Gating Paradigm: Effects of Auditory, Somatosensory and Cross-modal Stimulation,” *Clin. Neurophysiol.*, vol. 117, no. 11, Nov. 2006, pp. 2549-2563.
- [39] W.J. Ray and H.W. Cole, “EEG Alpha Activity Reflects Attentional Demands, and Beta Activity Reflects Emotional and Cognitive Processes,” *Sci.*, vol. 228, no. 4700, May 1985, pp. 750-752.



Jinyoung Moon received her B.S. degree in computer engineering from Kyungpook National University, Daegu, Rep. of Korea, in 2000. She received her M.S. degree in computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 2002. Since 2002, she has been working at ETRI. She is a senior member of the engineering staff of the Human Computing Research Section at ETRI. She is also currently a Ph.D. student in the Department of Industrial and Systems Engineering at KAIST. Her research interests include human-computer interaction, machine learning, data mining, and big data analysis.



Youngraek Kim received the B.S. and M.S. degrees in computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Rep. of Korea, in 2009 and 2011, respectively. Since 2011, he has been a member of the engineering staff at ETRI, Daejeon, Rep. of Korea. His research interests include machine learning, pattern recognition, and big data analysis.



Hyungjik Lee received his B.S. and M.S. degrees in electronics engineering from Kyungpook National University, Rep. of Korea, in 1998 and 2000, respectively. Since 2000, he has been a senior member of the engineering staff at ETRI, Daejeon, Rep. of Korea. His research interests include information appliances, home digital services, healthcare services, virtualization technologies, and big data analysis.



Changseok Bae received his B.S. and M.S. degrees in electronics engineering from Kyungpook National University, Rep. of Korea, in 1987 and 1989, respectively. He received his Ph.D. degree in electrical and electronics engineering from Yonsei University, Rep. of Korea, in 2003. From 1989 to 1996, he was a senior researcher at the Systems Engineering Research Institute, where he worked on image processing and pattern recognition. From 1997 to 1999, he worked with the Korea Ministry of Information and Communication, where he participated in establishing national software research and development policy. Since 2000, he has been a principal research staff member of the Human Computing Research Section at ETRI. From 2004 to 2005, he was a research fellow at the School of Information Technologies, University of Sydney, Australia. His research interests include image processing, multimedia codec, home server architecture, information hiding, and data mining.



Wan Chul Yoon received his M.S. degree in industrial engineering from KAIST and his Ph.D. in industrial and systems engineering from Georgia Institute of Technology in 1987. He joined the faculty of the Department of Industrial & Systems Engineering at KAIST in 1988, where he has worked as a full professor. He is the founder and current head of the Department of Knowledge Service Engineering at KAIST. His main research areas include cognitive engineering, human-computer interaction, user experience, system safety, applied intelligence, and knowledge processing.