Expert Systems With Applications 115 (2019) 578–592

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Detecting user attention to video segments using interval EEG features

Jinyoung Moon[a], Yongjin Kwon[a], Jongyoul Park[a], Wan Chul Yoon[b,*]

[a] SW•Content Research Laboratory, Electronics and Telecommunications Research Institute, 218 Gajeong-ro, Yuseong-gu, Daejeon, 34129, Republic of Korea
[b] Department of Industrial & Systems Engineering, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

ARTICLE INFO

Article history:
Received 30 November 2017
Revised 20 July 2018
Accepted 10 August 2018
Available online 11 August 2018

Keywords:
Detection
User attention
Video viewing
Video segments
Interval EEG features

ABSTRACT

To manage voluminous viewed videos, which US adults watch at a rate of more than five hours per day on average, an automatic method of detecting highly attended video segments during video viewing is required to access them for fine-grained sharing and rewatching. Most electroencephalography (EEG)-based studies of user state analysis have addressed the recognition of attention-related states in a specific task condition, such as drowsiness during driving, attention during learning, and mental fatigue during task execution. In contrast to attention in a specific task condition, both inattention and normal attention are meaningless to viewers in terms of managing viewed videos, while detecting high attention paid to video segments would make a valuable contribution to an automatic management system of viewed videos based on viewer attention. To the best of our knowledge, this is the first EEG-based study of detecting viewer attention paid to video segments. This study describes how to collect video-induced EEG and attention data for video segments from viewers without bias to specific genres and how to construct a subject-independent detection model for the top 20% of viewer attention. The attention detection model using the proposed interval EEG features from 14 channels achieved the best average F_1 score of 39.79% with an average accuracy of 52.96%. Additionally, this paper proposes a channel-based feature selection method that considers both the performances of single-channel models and their physical locations for investigating the group of channels relevant to attention detection. The attention detection models using the interval EEG features from all four or some of the channels located in the fronto-central, parietal, temporal, and occipital lobes of the left hemisphere achieved the best F_1 score of 39.60% with an average accuracy of 48.70%. It is shown that these models achieve better performance than models using the features from all four or some of their symmetric channels in the right hemisphere and models using the features from six channels located in the anterior-frontal and frontal lobes of the left and right hemispheres. This paper shows the feasibility of subject-independent and genre-independent attention detection models using a wireless EEG headset with optimized channels; these models can be applied to an intelligent video management system based on viewer attention in real-world scenarios.

1. Introduction

Video has become the most dominant form of digital content. People consume an enormous volume of videos, including video clips from the Internet, scheduled and recorded linear television (TV) programs, and downloaded or streamed TV series and movies through mobile personal devices, such as smartphones, tablets, and laptops, and TV screens at home any time of day. The average hours of video viewing through TV screens and other digital devices have steadily increased. Although the number of hours per week of video viewing varies among consumer groups, the average total viewing hours of nearly two-thirds of all users exceed 27 h per week in thirteen countries from Europe, North and South America, and Asia (Ericsson ConsumerLab, 2016). Additionally, YouTube viewers watched 1 billion hours of YouTube videos a day in 2017, which represents a 10-fold increase since 2012, according to a YouTube announcement (YouTube, 2017). According to an eMarketer report (eMarketer, 2015), the average time spent per day with videos by US adults increased from 4 h and 56 min in 2011 to 5 h and 31 min in 2015.

Viewers require an automatic method of detecting valuable video segments in order to share interesting video segments with other people and rewatch enjoyable or informative video segments among voluminous viewed videos. The current commercial video management methods provided by video sharing sites or available as tools for managing notes for videos, however, entirely depend on manual tasks, such as saving videos in specific playlists and generating notes for videos, which are performed by users. Be-

* Corresponding author.
E-mail addresses: jymoon@etri.re.kr (J. Moon), scocso@etri.re.kr (Y. Kwon), jongyoul@etri.re.kr (J. Park), wcyoon@kaist.ac.kr (W.C. Yoon).

https://doi.org/10.1016/j.eswa.2018.08.016
0957-4174/

cause viewers are in a lean-back mode during video viewing, in which they focus on the monitor and minimize movements, the manual tasks for managing videos interfere with viewers' manner of watching videos. In addition, the methods support either whole-video-level management or management of the notes with only a start timestamp, although the ideal unit of video management is the time interval with the start and end timestamps. Therefore, this paper proposes a method of detecting user attention paid to video segments obtained from viewed videos.

In this paper, electroencephalogram (EEG) signals collected during video viewing are used to assess user responses to video contents to achieve a good temporal resolution of the attention detection method. An EEG is a bio-signal indicating the electrical activity of the brain collected from electrodes placed on the scalp. Because EEG signals can be collected with a temporal resolution in the millisecond range using a relatively low-cost hardware device, compared with other functional neuroimaging technologies, such as computed tomography (CT) and functional magnetic resonance imaging (fMRI), EEG signals have been used in many studies on the analysis of attention-related cognitive user states.

Most studies analyzing attention-related user states have focused on attentional or inattentional states in a specific task condition, such as drowsiness, mental fatigue, and attention during performing specific tasks, such as driving, mental tasks, and learning. Drowsiness is an intermediate state between wakefulness and sleep. Drowsiness in drivers is a major cause of traffic accidents. Mental fatigue is a user state induced by prolonged mental load when someone iteratively performs a tedious but mentally demanding task. Such mental fatigue leads to deterioration in task performance with regard to time and precision, and unalleviated mental fatigue threatens the health and safety of workers in the workplace. Therefore, numerous methods for detecting drowsiness and mental fatigue based on classification, regression (Wu, Lawhern, Gordon, Lance, & Lin, 2016), and indices (Charbonnier, Roy, Bonnet, & Campagne, 2016; Silveira, Kozakevicius, & Rodrigues, 2016) have been proposed. Most previous studies of drowsiness recognition (Awais et al., 2017; Chai et al., 2016; Chuang et al., 2014; Hu, 2017; Khushaba et al., 2011; Li et al., 2016, 2017; Myrden & Chau, 2017; Shabani et al., 2016; Zhang et al., 2017) trained and tested their models using datasets consisting of EEG signals collected during driving simulation for both alert and drowsy states to recognize mental fatigue induced by driving or performing mental tasks. Some EEG-based studies of detecting mental fatigue (Chai et al., 2016; Li et al., 2016) classify normal and fatigue states of a subject with eyes closed and without specific actions before and after performing some tasks, such as a long drive or a series of mental tasks. Other studies (Hu, 2017; Mydren & Chau, 2017) classify normal and fatigue states of subjects during driving or performing the tasks. Attention is the behavioral and cognitive process of selectively concentrating on a discrete aspect of information. For better learning achievement in e-learning, methods for recognizing the attention of learners (Djamal, Pangestu, & Dewi, 2016; Li et al., 2011; Liu, Chaing, & Chu, 2013; Hu et al., 2018) have been proposed. However, EEG-based studies of attention recognition have not included the task of video viewing.

With the exception of Awais et al. (2017), the EEG-based studies of recognizing attention-related states proposed between 2016 and 2018 achieved the best average accuracies of at least 80% and just over 95%, but their models are totally or partially dependent on subjects. Some of them trained and tested their personal models by using samples collected from the same subject, although the training and test sets used were partitioned. Others validated their models in k-fold cross-validation, in which the samples from a subject can be divided and included into the training and test dataset. However, commercialization considerations warrant models that are trained and tested in a leave-one-subject-out (LOSO) cross-validation scheme. Moreover, in most cases, the model cannot be trained by using data from real customers after release.

In contrast to the vast amount of EEG-based studies on drowsiness, mental fatigue, and attention in a specific task condition, such as driving, performing mental tasks, or learning, EEG-based research on detecting attention to video contents is in a very early stage. To the best of our knowledge, there has not yet been a study on detecting attention to various genres of video contents during video viewing. Previously proposed methods (Daimi & Saha, 2014; Gauba et al., 2017; Moon, Kim, Lee, Bae, & Yoon, 2013) recognized user preference for video contents, but these methods have limitations on specific video genres, such as music and advertisement. Belle, Hargraves, and Najarian (2012) used video stimuli for detecting general attention, not for detecting attention to video contents. Therefore, they classify attentional and inattentional states, which were induced by interesting videos, such as racing, and uninteresting videos, including repetitive and monotonous scenes, such as a clock ticking and still images without any movements, respectively. In Mehmood, Sajjad, Rho, and Baik (2016) and Salehin & Paul (2017), EEG-based attention curves within a video for extracting keyframes that are included in a video summary have been proposed. Although they can provide attention indices, which are normalized between 0 and 1 or are standardized with a mean of 0 and standard deviation of 1, without training, the indices should be aggregated and sorted for a time period in advance in order to recognize viewer states between attention and inattention for a specific video segment. In addition, the video datasets they used were biased toward adventure and fantasy, which are interest-stimulating genres.

Therefore, we propose a method for detecting user attention paid to video segments that uses EEG signals collected during video viewing in order to automatically manage highly attended video segments among viewed videos. The attentional states of the viewers are divided into attention and inattention. In this paper, attention is defined as an attentional state in which the viewer shows high attention levels in a range corresponding to more than the highest 10% but less than the highest 30% of all attention levels exhibited by that viewer, corresponding to approximately the highest 20% of attention levels across all subjects. To train and test the models for detecting the attention of video viewers, we constructed an EEG dataset collected during the viewing of video stimuli evenly distributed across genres. The presented videos include six interest-stimulating videos in the genres of dramas and entertainment programs and six informative videos in the genres of news and documentaries. The EEG signals were collected from 18 subjects, each of whom wore an EEG headset with 14 electrodes while viewing the twelve video clips. This paper proposes models for attention detection based on five classifiers using interval EEG features. The proposed interval EEG features are formulated as descriptive statistics of typical EEG features in the frequency domain; they include band power (BP), asymmetric score (AS), and band ratio (BR) features. The attention detection model using interval BR features from all 14 channels achieved an average F_1 score of 39.79% (the average precision of 28.13% and the average recall of 72.19%) with a related average accuracy of 52.96%, as assessed via leave-one-subject-out (LOSO) cross-validation. The single-channel models using interval BR features achieved an average F_1 score of 37.47% (with an average precision of 26.23%, an average recall of 72.19%, and an average accuracy of 53.59%).

Additionally, we propose a selected multi-channel model by a channel-based feature selection method that considers both the performances of single-channel models and their physical locations for investigating the group of channels relevant to attention detection. The attention detection models using the interval EEG features from all four or some channels located in the frontocentral, parietal, temporal, and occipital lobes of the left hemi-

sphere achieved the best and second best F_1 score of 39.60% and 38.50% with an average accuracy of 48.70% and 52.86%, respectively. It was shown that the models achieved better performance than models using the features from all four or some of their symmetric channels in the right hemisphere and models using the features from six channels located in the anterior-frontal and frontal lobes of the left and right hemispheres.

Consequently, this paper makes the following contributions:

1. First, to the best of our knowledge, this is the first EEG-based study of detecting viewer attention paid to video segments during video viewing. To manage voluminous viewed videos in the unit of video segments, this paper describes how to collect EEG and attention data for video segments from viewers and construct detection models for the top 20% of viewer attention using the proposed interval EEG features. The proposed models can be applied to automatic video management systems based on viewer attention paid to video segments; these systems manage highly attended video segments without manual annotation by viewers for sharing and rewatching them.
2. Second, to ensure real-world practicability, this paper validates the proposed models for detecting viewer attention using an EEG dataset, which includes EEG signals collected from 18 subjects with a low-cost 14-channel wireless EEG headset while viewing both interest-stimulating and informative videos, with LOSO cross-validation. Although the detection models trained by EEG data collected with a wired EEG cap having more than 20 channels can achieve better performances than the detection models trained by EEG data collected using the wireless EEG headset, this study aims to show the feasibility of the subject-independent and genre-independent models using EEG signals from a wireless headset, considering their commercialization in real-world scenarios.
3. Third, this paper shows that the models using the interval EEG features from all four or some of the channels located in the fronto-central, parietal, temporal, and occipital lobes of the left hemisphere achieve better performance than models using the features from all four or some of their symmetric channels in the right hemisphere and models using the features from six channels located in the anterior-frontal and frontal lobes of the left and right hemispheres. Our findings could be applied to reduce the electrodes of a commercial EEG headset optimized for attention detection.
4. Finally, this paper proves the superiority of the proposed interval EEG features based on descriptive statistics of the typical EEG features during a given time interval by experimentally comparing the models using the proposed features with models using typical BR features with regard to both the average F_1 score and the average accuracy. Compared with the best all-channel model using typical BR features, the performance of the best all-channel model using interval BR features was enhanced by 9.43% (3.43%p) in the average F_1 score and 112.61% (28.05%p) in the average accuracy.

The remainder of this paper is organized as follows. In Section 2, this study provides some background on EEG measurements. Section 3 describes the materials and methods used to train and test subject-independent models for detecting user attention to video segments. Section 4 reports the experimental results for attention detection. Corresponding discussions with implications and future directions are presented in Section 5. Section 6 presents the literature review of EEG-based recognition for attention-related states. Finally, Section 7 concludes our work.

2. Background on EEG measurements

An EEG measures voltage fluctuations resulting from ionic current flows within the neurons of the brain. The brain's electrical charge is maintained by billions of neurons, which constantly exchange ions and molecules only with the fluid in the extracellular space to maintain resting potentials and propagate action potentials. Ions that reach the electrodes on the scalp can push or pull electrons on the metal of the electrodes. Because metal easily conducts such pushed and pulled electrons, the difference in push or pull voltages between any two electrodes can be measured by a voltmeter. EEG signals are bio-signals that record these voltages over time. An EEG reflects the summation of the synchronous activity of thousands or millions of neurons that have similar spatial orientations because the ions of neurons with similar spatial orientations line up and create waves that can be detected. Because voltage fields fall off with the square of the distance, activity from deep sources is more difficult to detect than currents near the skull. Each scalp electrode records electrical activity at very large scales, reflecting the electrical potentials generated in cortical-layer tissue containing 10 million to one billion neurons. The amplitudes of EEG activity on the scalp commonly lie within the range of 10–100 μV (Sanei & Chambers, 2008; Tatum, Husain, & Benbadis, 2008).

Continuous EEG recordings consist of oscillations at different frequencies and amplitudes that fluctuate over time and provide valuable information regarding a subject's brain state. There are five major types of brain waves, distinguished by their frequency ranges: delta (δ; <4 Hz), theta (θ; 4–7 Hz), alpha (α; 8–13 Hz), beta (β; 14–30 Hz), and gamma (γ; >30 Hz) (Sörnmo & Laguna, 2005). There are slight variations in the frequency ranges of each frequency band, as reported in the EEG literature. Each frequency band is associated with different types of mental states and exhibits specific signal characteristics. The delta band is primarily associated with deep sleep but is easily contaminated by artifacts resulting from muscle movements of the neck and jaw. The theta band is related to deep meditation and the level of arousal. The power in the alpha band is reduced by opening the eyes, hearing unfamiliar sounds, anxiety, and mental concentration. The alpha band power has been thought to indicate a state of relaxed awareness without any attention or a state of concentration. The beta band is associated with waking of the brain caused by active thinking and attention or the solving of concrete problems. Beta activity is encountered chiefly over the frontal and central regions. The power in the gamma band is very low, and its occurrence is rare. The gamma band can be used to confirm certain types of brain disease. In addition, the alpha and beta bands are related to sustained attention. In a relaxed state, the power values of the alpha band increase, and those of the beta band decrease. Conversely, in an attentive state, the power values of the alpha band decrease, and those of the beta band increase. For example, opening the eyes induces a decrease in alpha band activity and an increase in beta band activity because of the emergence of visual attention.

The international 10–20 system (10–20 system), which is depicted in Fig. 1, is an internationally recognized method of describing the locations of electrodes on the scalp in an EEG test or experiment. The letter and number of each position represent the lobe and hemisphere location, respectively. The capital letters F, T, C, P and O stand for the frontal, temporal, central, parietal, and occipital lobes, respectively. For the higher-resolution system, AF between Fp and F, FC between F and C, FT between F and T, CP between C and P, TP between T and P, and PO between P and O are added to the 10–20 system. The lower-case z refers to an electrode located on the midline. Electrode positions with even and odd numbers are located on the right and left hemispheres, respectively. In addition,

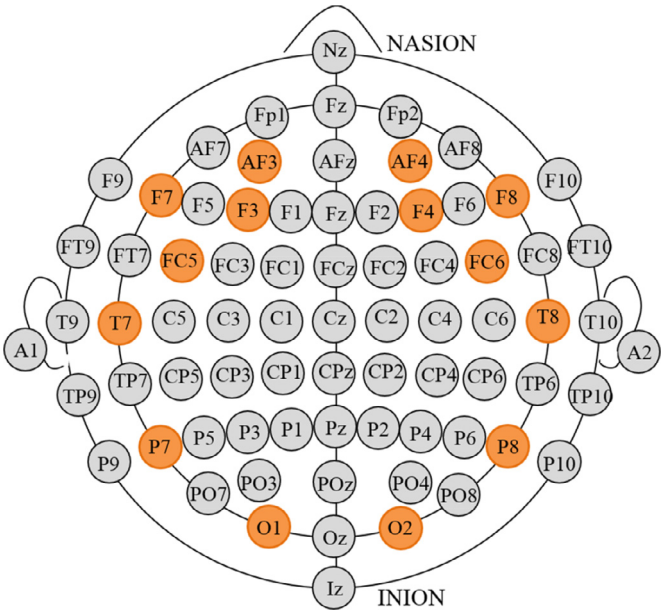

Fig. 1. Placement of the 14 electrodes used in this study in the international 10–20 system, which is an internationally recognized method of describing the locations of electrodes on the scalp for EEG experiments or applications (Malmivou & Plonsey, 1995). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

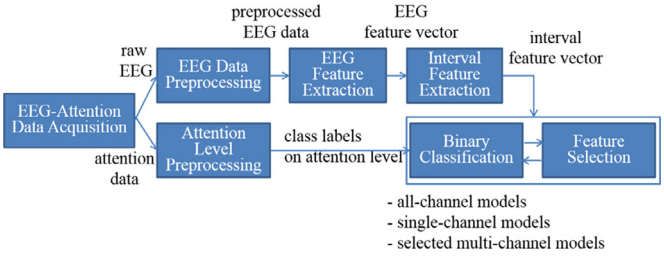

Fig. 2. Procedure for constructing EEG-based attention detection models for video segments.

the letter codes A, Pg and Fp identify earlobe, nasopharyngeal and frontal polar sites, respectively.

3. Materials and methods for EEG-based attention detection

To extract fine-grained indicators of highly attended video segments in all viewed videos, this study proposes a method that detects the top 20% of user attention levels directed toward all video segments. This section describes the method used to construct a genre-independent EEG dataset of attention levels for each video segment and the methods applied to train and test subject-independent models using the proposed interval EEG features; these methods include steps for the preprocessing of the EEG and attention data, feature extraction, classification, and feature selection.

The proposed method follows the procedure shown in Fig. 2. First, raw EEG data were obtained from 18 subjects while they watched twelve video stimuli. After each subject had viewed the videos, the attention levels for video segments were collected from that subject's responses to a questionnaire on the video segments included in the video stimuli. Second, the raw EEG data were preprocessed before feature extraction. The preprocessing step included filtering and segmenting the raw EEG data and generating class labels, namely, attention and inattention, for each unit time interval of each video segment. Third, three types of EEG feature

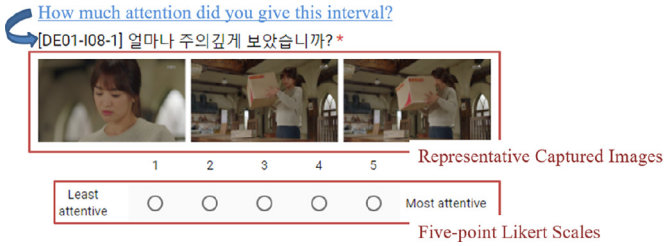

Fig. 3. Screenshot of representative captured images from each video segment, and a question on the attention level for the segment on a scale of one to five.

vectors, namely, BP, AS, and BR feature vectors, in the frequency domain were extracted with a time unit of one second. The BP, AS, and BR features were transformed into interval BP, AS, and BR features by calculating descriptive statistics of the EEG feature vectors, namely, the median, minimum, maximum, and skewness. The developed models include all-channel, single-channel, and selected multi-channel models, which use interval EEG features from all 14 channels, a single channel, or multiple selected channels, respectively. Each step will be explained in the following subsections.

3.1. Experimental setup and EEG attention data acquisition

The video stimuli presented to the subjects were designed to measure the levels of attention paid to video segments independent of genre. The selected video stimuli include two types of video contents: interest-stimulating and informative video clips. The interest-stimulating video stimuli consist of three video clips from dramas and three video clips from entertainment programs. The informative video stimuli consist of three video clips from news programs and three video clips from documentaries. All 160-second video stimuli of at least HD resolution (1080×720) were extracted from different TV programs that were broadcast on Korean TV channels and uploaded to Korean portals and YouTube. The attention levels of each subject for the video stimuli were obtained from the subjects' responses to a questionnaire. To achieve a fine-grained assessment of attention level, each video stimulus was manually segmented into nine intervals on average (between seven and twelve intervals), corresponding to unit time intervals of several seconds in length.

As shown in Fig. 3, the questionnaire included two or three representative captured images from each video segment and presented a question on the attention level for each segment. For the level of attention to a video segment, the questionnaire asked the subject to assign an integer value ranging from one to five points on a five-point Likert scale. The middle point, three, indicates a normal state of attention. The lowest point, one, represents the least attentive level, and the highest point, five, represents the most attentive level.

In accordance with the experimental protocol of this study, as shown in Fig. 4, EEG signals and attention levels for video segments were collected from eighteen healthy right-handed subjects (nine males and nine females) with ages ranging from 21 to 41. Each session consisted of a preparation step for EEG collection and a collection step for EEG and attention data. To reduce biological artifacts in the collected EEG signals, the subject was 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 were of sufficient quality for recording, the experimenter had the subject watch six video stimuli in sequence for approximately sixteen minutes. All video stimuli viewed in a session had the same type of video content.

As shown in Fig. 5, EEG signals were gathered from the EEG device while the subject viewed the video stimuli. After viewing

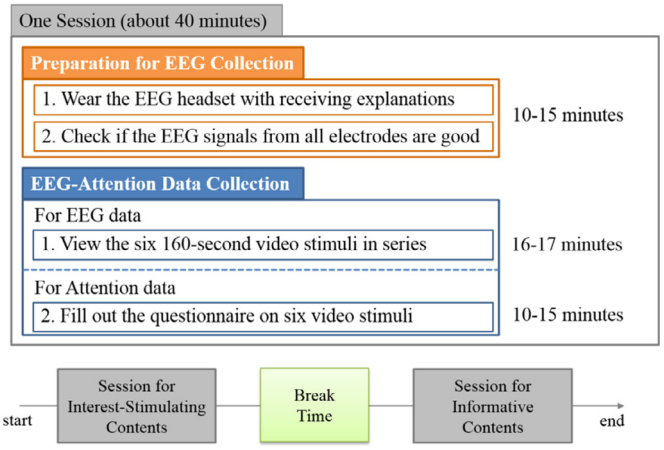

Fig. 4. Protocol of each session with regard to the gathering of EEG and attention data, and the arrangement of the two sessions and the break between them.

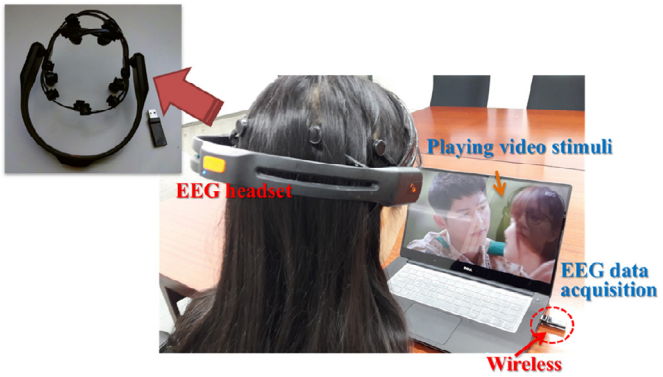

Fig. 5. Experimental environment. The raw EEG data were collected from a subject wearing a mobile EEG headset while the subject viewed the video stimuli.

the video stimuli, the subject completed the questionnaire. Each session took approximately forty minutes for each subject. Each subject participated in two sessions, between which there was a break.

The EEG device collected EEG signals from the following 14 channels corresponding to electrodes placed on the scalp: AF3, AF4, F7, F8, F3, F4, FC5, FC6, T7, T8, P7, P8, O1, and O2, as marked with orange circles in Fig. 1. The EEG signals were recorded at a sampling rate of 128 Hz. The bandwidth of the recorded EEG signals was from 0.2 to 45 Hz; a 0.16-Hz high-pass filter, an 83-Hz low-pass filter, and 50-Hz and 60-Hz notch filters were applied. The notch filters were employed to remove power line artifacts. Environmental artifacts caused by the amplifier and aliasing were removed from the hardware of the EEG device.

3.2. EEG and attention data preprocessing

Data preprocessing was required to segment the EEG signals for feature extraction and to generate attention-related classes synchronized with interval EEG feature vectors.

For each video stimulus, the first two seconds of the EEG signals, corresponding to the hardware and software initialization of the EEG headset, were removed. The remaining 158-s EEG signals for each video contained 20,224 EEG samples (128 EEG samples per second multiplied by 158 s per video). For a video stimulus, 158 EEG features in the frequency domain were derived from the 20,224 well-segmented EEG samples in the feature extraction step.

In this study, biological artifacts caused by blinks and eyeball movements were avoided by excluding the corresponding artifact-influenced frequency bands. Because it is impossible to prevent all biological artifacts in signals collected from live subjects, although the subjects were asked to view the videos in a nearly static position, additional effort was necessary to minimize the influence of biological artifacts after the EEG recording step. The influence of biological artifacts can be decreased by rejecting artifact-related frequency bands. The artifacts caused by heartbeats and the artifacts caused by blinks and eyeball movements appear at frequencies of approximately 1.2 Hz and below 4 Hz, respectively. In contrast, artifacts caused by muscle movements are most dominant above 50 Hz (Petrantonakis & Hadjileontiadis, 2010). Therefore, the models presented in this paper were constructed using interval EEG features from the artifact-resilient frequency range between 4 Hz and 40 Hz.

The EEG data of each channel were filtered through a Butterworth filter with a passband between 1 Hz and 50 Hz. The Butterworth filter was employed because it is designed to have as flat a frequency response as possible in the passband (Allen & MacKinnon, 2010).

For the class labeling of the attention levels, the integer-valued attention level for each interval was assigned a class label of either attention or inattention in accordance with the distribution of the attention levels that were scored by each subject. For each subject, the attention levels corresponding to approximately the top 20% of all attention levels (more than the top 10% and less than the top 30%) were assigned the attention class label, and the remaining attention levels were assigned the inattention class label. In other words, the attention and inattention class labels were normalized in accordance with the personal variations in the scoring of the attention levels.

3.3. EEG and interval feature extraction

In this study, interval EEG features were employed as the basis of the attention detection models. These features are based on descriptive statistics of the typical EEG features in the frequency domain. For each video stimulus, the 20,224 amplitude samples from the raw EEG data (158 s × 128 EEG samples/second) were transformed into 158 BP, AS, and BR features in the frequency domain. Next, for each video segment, the BP, AS, and BR features were transformed into BP-, AS-, and BR-based interval EEG features, respectively. For each video segment, the interval EEG features consisted of four descriptive statistics of the EEG features within the corresponding time interval, namely, the minimum, maximum, median, and skewness.

One of the most common approaches for investigating EEG data is to analyze the activated and inactivated levels of BP, AS, and BR in the delta (<4 Hz), theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (>30 Hz) frequency bands (Sörnmo & Laguna, 2005). These frequency bands provide valuable information on the neural activities of normal people or reveal signs of psychological disorders. In this study, four frequency bands in the range of 4–40 Hz were selected: theta (4–7 Hz), alpha (8–13 Hz), beta (14–30 Hz), and gamma (31–40 Hz).

To extract the power in each frequency band, Welch's method (Welch, 1967) was applied for time-frequency signal decomposition. In Welch's method, time-series EEG signals are partitioned into several disjoint or overlapping blocks, and the EEG signal $x(t)$ in each block is convolved with a windowing function centered at time t and is then transformed into the time-frequency representation $X(f,t)$ by means of the short-time Fourier transform (STFT). The power spectral density (PSD) in the frequency domain is obtained by calculating the squared magnitude of the STFT coefficients. This study employed Welch's method with 128 non-overlapping sampling points, a sampling rate of 128 Hz, and a Hann window function. The absolute BP value for a given frequency band is obtained

Table 1
Three types of interval features and their dimensions for the case in which all 14 channels and all four frequency bands are used.

Interval feature type	EEG feature type	# of dimensions
Interval AS	AS	5 × 7 pairs of channels × 4 bands = 140
Interval BP	BP	5 × 14 channels × 4 bands = 280
Interval BR	BR	5 × 14 channels × 6 pairs of bands = 420

as the sum of all PSD values in a specific frequency range of that band from Eq. (1). In this study, the absolute BP features were transformed into a natural log scale because they are usually positively skewed (Allen & MacKinnon, 2010).

$$BP(channel, band) = ln\left(\sum_{freq=band.freq\_start}^{band.freq\_end} PSD_{channel}(freq)\right) \quad (1)$$

In addition, AS features can be derived from the difference between the BP values of two EEG electrodes that constitute a spatially symmetrical pair in the 10–20 system. Many studies have investigated the relationship between hemispheric asymmetry and emotion since 1979 (Allen & Kline, 2004). Considering the hemispheric asymmetry, the features were extracted using Eq. (2) below, wherein the band powers of the left and right EEG channels of a symmetric pair are denoted by BP_L and BP_R, respectively. Because the BP values are log-scaled, the AS values are obtained by taking the difference between the BP values of each pair of channels.

$$AS_{band}(channel_L, channel_R) = BP(channel_L, band) - BP(channel_R, band) \quad (2)$$

The BR features can be derived from the power ratios between pairs of bands selected from among the four considered bands: theta, alpha, beta, and gamma. The features were extracted using Eq. (3) below, wherein the EEG band powers of the two given frequency bands $band_m$ and $band_n$ are denoted by BP_m and BP_n, respectively, as shown in Eq. (2).

$$BR_{channel}(band_m, band_n) = BP(channel, band_m) - BP(channel, band_n) \quad (3)$$

All EEG features, including the AS, BP, and BR features, were smoothed with a 5-second moving average filter before interval feature extraction to remove fluctuations and highlight their overall trends.

To extract the interval EEG features used in this study, descriptive statistics were calculated for the time-series EEG features within the time interval corresponding to each video segment. Among the most common descriptive statistics for times-series data, such as the minimum, maximum, average, standard deviation, median, skewness, and kurtosis, as well as several types of entropy and energy, changes, and peaks, four measures, namely, the minimum, maximum, median, and skewness, were selected via a model-based approach for the selection of relevant features in accordance with the F_1 scores of the models. All extracted interval features are listed in Table 1.

All interval EEG features were normalized to a standard score before the training and testing of the attention detection models. This normalization generated new feature values with a mean of 0 and a standard deviation of 1. Such normalization is required to reduce the dependency on features with high values or a large margin.

3.4. Classification

For the training and testing of subject-independent classification models, this study used LOSO cross-validation, which is a special case of leave-one-group-out cross-validation in which instances from each subject are grouped. In each trial, 107 instances from one subject (corresponding to all 107 video segments from among the 12 video stimuli presented to each subject) were used for testing, and 1,819 instances from the other 17 subjects (107 video segments from among the 12 video stimuli multiplied by 17 subjects) were used for training. To evaluate the classifiers and features used, the average F_1 score, precision, recall, and accuracy were calculated for each type of model based on each possible combination of a classifier and a feature type by averaging the F_1 scores, precisions, recalls, and accuracies of all 18 models obtained using the training and test sets associated with each of the 18 subjects.

For a comparative study of the classifiers and feature types for predictive models of attentional states (either attention or inattention), five types of classifiers (support vector machine (SVM), logistic regression, decision tree, random forest, and AdaBoost) were employed to train and test attention detection models using the extracted interval AS, BP, and BR features. The five classifiers were selected from among nine candidate classifiers, which additionally included a one-class SVM classifier, a K-nearest-neighbors classifier, a linear discriminant analysis classifier, and a quadratic discriminant analysis classifier; the chosen classifiers were selected based on the average F_1 scores and accuracies obtained in a pre-inspection of all-channel classification models based on each of the nine classifiers. The last four classifiers were excluded from the experiment because both the best average F_1 scores of the all-channel models based on them and the related average accuracies were less than 25%. In addition, to overcome the class imbalance due to the smaller number of instances with the *attention* class label, the class weight between attention and inattention, which was a common parameter across all classifiers, was set to 5:1. All non-default parameters except the class weight were set as shown in Table 2 for each classifier. The listed parameter values were obtained through parameter tuning based on the average F_1 score.

Regarding the sets of channels selected for model building, the classification performance was examined in this study for three types of attention detection models: models using interval EEG features from all 14 channels, from a single channel, and from multiple channels selected through a channel-based feature selection method. The all-channel and multi-channel models are useful for assessing the best classification performance. The single-channel models are valuable for real-world applications because EEG-based models using a single channel or only a few channels are suitable for implementation in a commercial product.

As listed in Table 3, the best sets of frequency bands for models using certain sets of BP, AS and BR features were selected by evaluating 15 and 11 models corresponding to all possible combinations of bands for interval BP and AS features and for interval BR features, respectively, with regard to their average F_1 scores.

4. Experimental results

This study thoroughly investigated the relationships between classification performance in terms of four performance measures, namely, F_1 score, precision, recall, and accuracy, and several factors of the models, such as the classifiers, feature types, sets of chan-

Table 2
Parameters for each classifier.

Classifier	Parameter settings
SVM	C = 1, gamma = 0.001, kernel = 'rbf'
Logistic Regression	C = 0.001
Decision Tree	criterion = 'gini', max_features = 'log2', min_samples_leaf = 1, min_samples_split = 2, max_depth_options = 5
Random Forest	criterion = 'gini', max_features = 'log2', min_samples_leaf = 3, min_samples_split = 2, max_depth_options = 5, n_estimators = 5
AdaBoost	a baseLeaner with the tuned DecisionTree, learning_rate = 0.02, n_estimators = 5

Table 3
Tested sets of frequency bands for finding the best set of bands for a given set of channels.

# of bands	Set of bands	Interval BP/AS features	Interval BR features
4 bands	θ, α, β, γ	θ, α, β, γ	θ/α, θ/β, θ/γ, α/β, α/γ, β/γ
3 bands	θ, α, β	θ, α, β	θ/α, θ/β, α/β
	θ, α, γ	θ, α, γ	θ/α, θ/γ, α/γ
	θ, β, γ	θ, β, γ	θ/α, θ/γ, β/γ
	α, β, γ	α, β, γ	α/β, α/γ, β/γ
2 bands	θ, α	θ, α	θ/α
	θ, β	θ, β	θ/β
	θ, γ	θ, γ	θ/γ
	α, β	α, β	α/β
	α, γ	α, γ	α/γ
	β, γ	β, γ	β/γ
1 band	θ	θ	N/A
	α	α	N/A
	β	β	N/A
	γ	γ	N/A

Table 4
Classification performance of the best all-channel model for each of the five classifiers. The models are ranked by their F_1 scores and sorted in ascending order of rank.

Rank	Model		F_1 score (%)	Precision (%)	Recall (%)	Accuracy (%)
	Classifier	Best Features				
1	Decision Tree	- interval BR - 4 bands	39.79	28.13	75.77	52.96
2	Logistic Regression	- interval BR - 2 bands: α and γ	37.87	26.04	86.25	40.65
3	SVM	- interval BR - 4 bands	37.71	29.79	72.68	51.14
4	Random Forest	- interval BP - 2 bands: θ and γ	37.00	27.28	63.88	54.83
5	AdaBoost	- interval BR - 3 bands: α, β, and γ	34.49	25.28	64.02	50.31

nels, and sets of frequency bands, for all-channel, single-channel, and selected multi-channel models. Additionally, this study compared all-channel models using interval BR features with all-channel models using typical and intervalized BR features to prove the effectiveness of the proposed interval BR features for attention detection.

4.1. Results of all-channel and single-channel classification models

Table 4 shows the classification performance of the best attention detection model based on each classifier among models using interval BR, BP, and AS features from all channels in the best set of frequency bands optimized by their F_1 scores. The best average F_1 score of 39.79% (the average precision of 28.13% and the average recall of 75.77%) and a related average accuracy of 52.96% were obtained by the all-channel model based on the decision tree classifier using interval BR features from all four bands. Except for the model based on the logistic regression classifier, which achieved an average accuracy of 40.65%, the average accuracies of all of the all-channel models using interval BR features exceeded 50%. Considering both the average F_1 score and the average accuracy, the decision tree and SVM classifiers are the most suitable for all-channel models using interval BR features.

Except for the models based on the AdaBoost classifier, the all-channel models using interval BR features from the best set of bands optimized by the average F_1 score achieved better performance with regard to the average F_1 score and precision than the corresponding models using interval BP and AS features. The differences were statistically significant, as determined by one-way analysis of variance (ANOVA) ($p < 0.05$). The AdaBoost-based models using interval BR, BP and AS features were excluded from this comparison because the average F_1 score of the best all-channel model based on the AdaBoost classifier was considerably lower, by 7.28% (2.51%p), compared with the second-lowest F_1 score of the best all-channel model based on the random forest classifier. In Fig. 6, the classification performances of the all-channel models using interval BR, BP, and AS features are compared across the other four classifiers (excluding AdaBoost). The average F_1 scores of the models using interval BR features were higher by at least 10.34% and at most 19.71% compared with those of the models using interval AS features across the four classifiers. Except for the model based on the random forest classifier, the average F_1 scores of the models using interval BR features were higher by at least 4.04% and at most 17.13% compared with those of the models using interval BP features across the four classifiers. The average precisions of the models using interval BR features were higher by at least

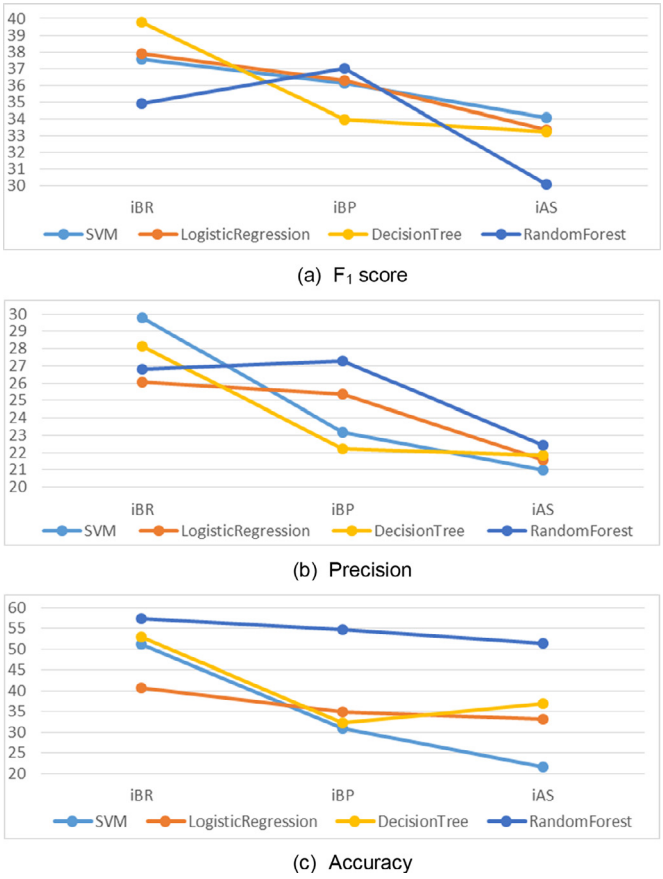

Fig. 6. Comparison of classification performances of all-channel models based on four classifiers using interval BR (iBR), BP (iBP) and AS (iAS) features from all fourteen channels in the best set of frequency bands optimized by the average F_1 score for performance measures of (a) F_1 score, (b) precision, and (c) accuracy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

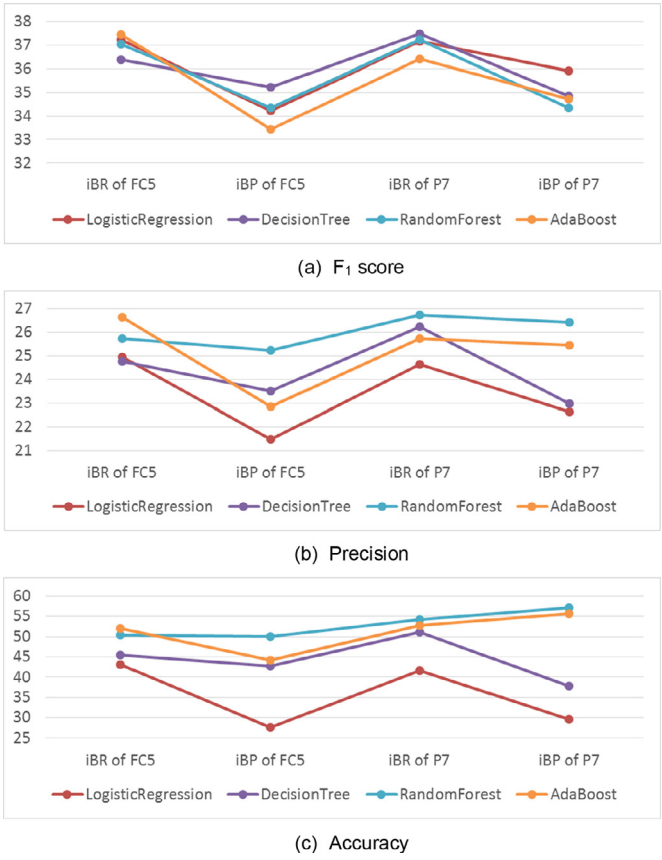

Fig. 7. Comparison of classification performances of single-channel models based on four classifiers using interval BR (iBR), BP (iBP) and AS (iAS) features from the best set of frequency bands optimized by the average F_1 score for performance measures of (a) F_1 score, (b) precision, and (c) accuracy. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

19.68% and at most 28.86% compared with those of the models using interval AS features across the four classifiers. The average accuracies of the models using interval BR features were higher by at least 11.94% and at most 135.67% compared with those of the models using interval AS features across the four classifiers. Meanwhile, the models using interval AS features achieved average F_1 scores of less than 35% and related average accuracies of less than 37%, except for the model based on the random forest classifier.

To find the most significant channels for attention detection among all 14 channels, this study also examined single-channel classification models using interval BR and BP features from each channel in the 11 and 15 sets of frequency bands, respectively, that are listed in Table 3.

Table 5 shows the classification performance of the best single-channel model based on each of the five classifiers among models using interval BR and BP features from the best set of frequency bands optimized by their F_1 scores. Among all single-channel models, the best and second-best average F_1 scores were obtained by models using BR features from P7 and FC5, respectively. The decision-tree-based model using interval BR features from P7 in the θ, β, and γ bands achieved the best average F_1 score of 37.47% (the average precision of 26.23% and the average recall of 72.19%) and a related average accuracy of 51.14%. The AdaBoost-based model using interval BR features from FC5 in the α, β, and γ bands achieved the second-best average F_1 score of 37.45% (the average precision of 26.65% and the average recall of 69.55%) and a related average accuracy of 52.02%. Although the SVM-based single-channel model using interval BR features showed only a mild decrease of 2.52% (0.95%p) in the average F_1 score compared with the SVM-based all-channel model using interval BR features, it exhibited a high average recall of 94.43% and an average accuracy of only 32.04%. These results indicate that this single-channel model predicts the class of most test instances to be the attention class. In contrast, the best single-channel models based on the decision tree, AdaBoost, and random forest classifiers achieved average accuracies exceeding 50%. Considering both the average F_1 score and accuracy, the decision tree and AdaBoost classifiers are the most appropriate for single-channel models using interval BR features.

Similar to the all-channel models, except for the models based on the SVM classifier, the single-channel models using interval BR features from FC5 and P7 achieved better performance in terms of F_1 score and precision than the corresponding models using interval BP features from FC5 and P7. The differences were statistically significant, as determined by the t-test ($p < 0.05$). The SVM-based models using interval BR and BP features from FC5 and P7 were excluded from this comparison because they achieved an average F_1 score below 36% and an average accuracy of 29%. In Fig. 7, the classification performances of the single-channel models using interval BR and BP features from the FC5 and P7 channels are compared across the other four classifiers (excluding SVM). The average F_1 scores of the models using interval BR features from FC5 and P7 were higher by at least 3.29% and at most 11.96% and by at least 3.51% and at most 8.42% compared with those of the models using interval BP features from FC5 and P7, respectively, across

Table 5
Classification performance of the best single-channel model for each of the five classifiers. The models are ranked by their F_1 scores and sorted in ascending order of rank.

Rank	Model		F_1 score (%)	Precision (%)	Recall (%)	Accuracy (%)
	Classifier	Best Features				
1	Decision Tree	- interval BR of P7 - 3 bands: θ, β, and γ	37.47	26.23	72.19	51.14
2	AdaBoost	- interval BR of FC5 - 3 bands: α, β, and γ	37.45	26.65	69.55	52.02
3	Logistic Regression	- interval BR of P7 - 4 bands	37.23	24.96	82.96	43.04
4	Random Forest	- interval BP of FC5 - 3 bands: α, β, and γ	37.23	26.75	68.45	54.15
5	SVM	- interval BR of P8 - 4 bands	36.76	23.43	94.43	32.04

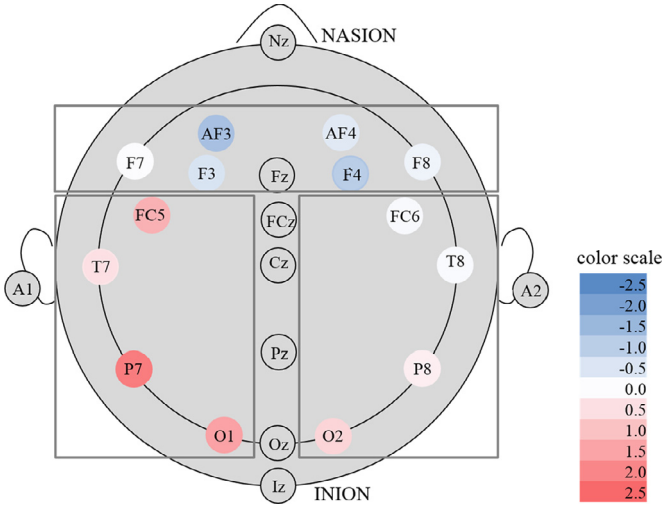

Fig. 8. Color-scaled standardized classification performance of single-channel models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

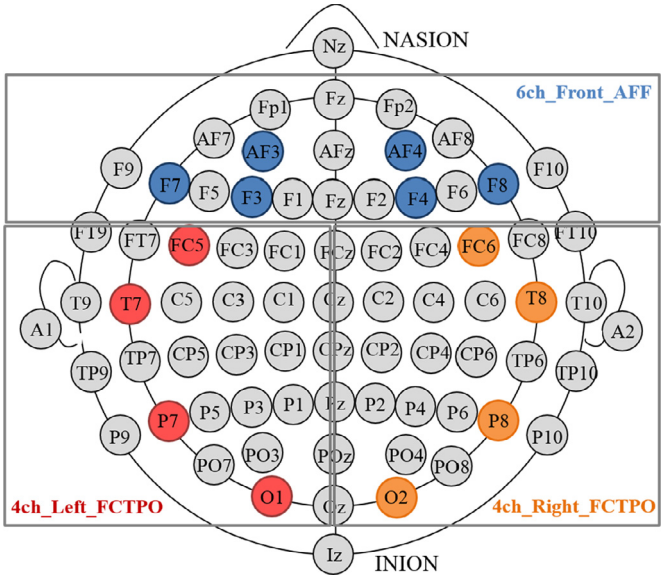

Fig. 9. Channel grouping by the performances of single channel models and physical locations.

the four classifiers. The average precisions of the models using interval BR features from FC5 and P7 were higher by at least 2.06% and at most 16.68% and by at least 1.18% and at most 14.19% compared with those of the models using interval BP features from FC5 and P7, respectively, across the four classifiers. The average accuracies of the logistic-regression-based and decision-tree-based models using interval BR features from FC5 and P7 were higher by 50.85% and 6.08% and by 41.37% and 35.11% compared with those of the models using interval BP features from FC5 and P7, respectively.

To compare the performances of 14 single-channel models considering their physical locations, this study standardizes the F_1 scores of all the single-channel models based on the decision tree using interval BR features of their best frequency bands and then depicts the color-scaled standard score in the 10–20 system, as shown in Fig. 8. The blue-colored and red-colored channels represent their classification performance below and above the average performance on all 14 channels, respectively. All single-channel models whose channel is located in the anterior frontal or frontal lobes achieved below the average performance of single-channel models. The single-channel models whose channel is located in the fronto-central, temporal, parietal, and occipital lobes in the left hemisphere achieved performance above the average performance of single-channel models. In particular, single-channel models using interval BR features from P7, O1, and FC5 achieved the top 3 F1 scores. The single-channel models using interval BR features from FC6, T8, P8, and O2 achieved performances between the performances of the previous two model groups.

4.2. Results of selected multi-channel models

Because it is important to reduce the number of electrodes in an EEG headset for commercial use, channel-based feature selection was performed in this study. In accordance with the average F_1 scores of the single-channel models using interval BR features, the 14 channels were grouped into three channel groups, listed in Table 6 and shown in Fig. 9.

There were statistically significant differences between the average F1 scores of the models using channels from the 4ch_Left_FCTPO, 4ch_Right_FCTPO, and 6ch_Front_AFF channel groups, as determined by one-way ANOVA ($p < 0.05$). To find the best combination of channels and frequency bands for multi-channel models, this study examined 2,420 multi-channel models using interval BR and BP features from channels in the 4ch_Left_FCTPO and 4ch_Right_FCTPO groups (5 classifiers x 2 interval EEG features x 2 channel groups x 11 sets of channels x 11 sets of frequency bands). To reduce the number of candidates for the best multi-channel model, we excluded the channels in 6ch_Front_AFF when constructing the selected multi-channel models since the single-channel models using interval features from the channels in 6ch_Front_AFF achieved the lowest F1 scores.

Table 6
Channel groups and the channel sets for testing multi-channel models. The channels are grouped into three groups in accordance with the average F_1 scores of the single-channel models and each channel group, except 6ch_Front_AFF, have 11 sets of channels.

Channel group	Component channels	Channel sets
4ch_Left_FCTPO	FC5, T7, P7, O1	4ch_Left_{FCTPO} 3ch_Left_{FCTP, FCTO, FCPO, TPO} 2ch_Left_{FCT, FCP, FCO, TP, TO, PO}
4ch_Right_FCTPO	FC6, T8 P8, O2	4ch_Right_{FCTPO} 3ch_Right_{FCTP, FCTO, FCPO, TPO} 2ch_Right_{FCT, FCP, FCO, TP, TO, PO}
6ch_Front_AFF	AF3, AF4, F3, F4, F7, F8	

Table 7
Classification performance of the best selected multi-channel model for each of the five classifiers. The models are ranked by their F_1 scores and sorted in ascending order of rank.

Rank	Model		F_1 score (%)	Precision (%)	Recall (%)	Accuracy (%)
	Classifier	Best Features	- comparison with the best average F1 score of 1ch-model			
1	Logistic Regression	- Interval BR - 3ch_Left_FCTO - 4 bands	39.60 − 6.37% (+2.37%p)	27.14	81.17	48.70
2	Decision Tree	- Interval BR - 2ch_Left_TP - 2 bands: θ and γ	38.50 − 2.75% (+1.03%p)	26.89	73.67	52.86
3	Random Forest	- Interval BR - 4ch_Left_FCTPO - 2 bands: θ and γ	38.28 – 2.82% (+1.05%p)	28.81	62.80	59.71
4	SVM	- Interval BR - 4ch_Left_FCTPO - 4 bands	38.06 – 6.76% (+2.41%p)	25.01	88.59	40.45
5	AdaBoost	- Interval BR - 2ch_Left_FCO - 2 bands: α and γ	37.80 – 0.93% (+0.35%p)	26.95	73.98	48.39

As shown in Table 7, among all of the examined multi-channel models, the best average F_1 score of 39.60% was obtained by the logistic-regression-based model using interval BR features from FC5, T7, and O1 (a three-channel set from the 4ch_Left_FCTPO group) in all four frequency bands. Among the best models based on each of the five classifiers, the second-best average F_1 score of 38.50% was obtained by the decision-tree-based model using interval BR features from T7 and P7 (a two-channel set from the 4ch_Left_FCTPO group) in the corresponding best set of bands; this result is the fifth-best F_1 score among all of the examined multi-channel models. Among the best models based on each of the five classifiers, only two average accuracies of the multi-channel models based on the decision tree and random forest classifiers using interval BR features were above 52%. Considering both the average F_1 score and the average accuracy, the decision tree and random forest classifiers are the most suitable for multi-channel models using interval BR features.

The majority of the top 65 average F_1 scores, all of 37% or higher, were achieved by multi-channel models using interval BR features from all or some of the channels in the 4ch_Left_FCTPO group. There were no models using interval BP features among the 65 models with the highest average F_1 scores. The best model using interval BP features achieved the 91st highest average F_1 score among all of the examined multi-channel models. There were only eight models using the interval BR features from channels in the 4ch_Right_FCTPO group among the 65 multi-channel models with the highest average F_1 scores. The best model using the interval BR features from channels in the 4ch_Right_FCTPO group achieved the 30th average F_1 score among the top 65 average F_1 scores.

4.3. Results on the effectiveness of interval BR features

To prove the superiority of interval EEG features, this study compared the average F_1 scores and accuracies of all-channel models using the proposed interval BR features with those of all-channel models using both typical and intervalized BR features across four of the classifiers. Because it is difficult to train and test an SVM classifier implemented with libsvm using more than 10,000 instances (scikit-learn 0.19.1), the SVM classifier was excluded from this comparison. The intervalized BR features were set to the same values obtained for the descriptive statistics of the typical BR features during each interval for each second of the corresponding interval, whereas the typical BR features consisted of the different values measured for each second during each interval of the same attention level class. The time unit for both the typical and intervalized BR features was one second. As shown in Table 8, the best average F_1 score among models using the proposed interval BR features was higher by 9.43% and 7.66% compared with those of models using the typical and intervalized BR features, respectively. In addition, the best average accuracy among models using the proposed interval BR features was higher by 112.61% and 30.51% higher than those of models using the typical and intervalized BR features, respectively.

5. Discussion

In contrast to the numerous previous studies on attention, which have mainly addressed attention during the performance of a specific task, the present work focused on the top 20% of levels of user attention to video segments among all attentional states during video viewing. The research objectives of this study and

Table 8
Comparison of the average F_1 scores and related average accuracies of all-channel models using typical, intervalized, and interval BR features across four classifiers.

Model	Model Performance - F_1 score (%) - Accuracy (%)			Comparison - Increase in F_1 score as %p (%) - Increase in Accuracy as %p (%)	
	Typical BR	Intervalized BR	Interval BR	with Typical BR	with Intervalized BR
Logistic Regression	36.36	36.32	37.87	4.15% (1.51%p)	4.27% (1.55%p)
	24.91	29.57	40.65	63.19% (15.74%p)	37.47% (11.08%p)
Decision Tree	36.17	36.47	39.79	10.01% (3.62%p)	9.10% (3.32%p)
	28.23	36.59	52.96	87.60% (24.73%p)	44.74% (16.37%p)
Random Forest	36.02	36.96	34.88	−3.16% (−1.14%p)	−5.63% (−2.08%p)
	27.42	40.58	57.48	109.63% (30.06%p)	41.65% (16.90%p)
AdaBoost	35.90	36.30	34.49	−4.21% (−1.51%p)	−4.98% (−1.81%p)
	24.37	43.80	50.31	106.44% (25.94%p)	14.86% (6.51%p)
Best models across classifiers	36.36	36.96	39.79	9.43% (3.43%p)	7.66% (2.83%p)
	24.91	40.58	52.96	112.61% (28.05%p)	30.51% (12.38%p)

those of previous studies are different in three important aspects in terms of the methods of data collection and the preprocessing of the EEG and attention data. First, the difference between attention and inattention in this paper is less distinct and explicit than that considered in previous studies. In this study, the states of both attention and inattention that were induced by video stimuli obtained from real TV programs were divided not by the authors but by each subject. Each subject could assign a different value of attention level to the same video segment, and the video stimuli were real video contents that are available for people to view in a real-world setting. In Belle et al. (2012), states of inattention were induced by predefined uninteresting videos containing repetitive and monotonous scenes or still images without any changes, and these states were compared with the states of attention induced by predefined interesting videos containing scenes from documentaries, movies, and car chases; this approach was adopted because the authors were focused on general attention and inattention. Similarly, in Li et al. (2011), states of inattention, which were compared with states of attention during the performance of specific mental tasks, were induced by asking the subjects to relaxing with their eyes closed while not performing any tasks. Second, in the present study, imbalanced attention and inattention labels were intentionally generated by assigning integer-valued attention levels on a five-point Likert scale to binary states of attention and inattention in accordance with the distribution of the attention levels assigned by each subject, with the intention of detecting the top 20% of user attention levels. In most previous studies, except for Myrden and Chau (2017), equally spaced attention levels collected from each subject either have been used directly as class labels or have been categorized into a few groups in accordance with predefined static guidelines (for example, by mapping attention levels of 1 and 2 to inattention and attention levels of 3, 4, and 5 to attention). In contrast, Myrden and Chau (2017) generated balanced class labels by excluding scarce levels and combining nearest class labels. Finally, the video stimuli presented to our subjects included video clips from four genres (drama and entertainment for arousing interest and news and documentaries for presenting information) to prevent any genre dependency of the trained models. In Daimi and Saha (2014), Moon et al. (2013) and Gauba et al. (2017), the presented video stimuli were restricted to a specific genre, such as music or advertisements.

Among all-channel models, the best average F_1 score of 39.79% and a related average accuracy of 52.96%, as determined through LOSO cross-validation, were achieved by a model based on a decision tree classifier using interval BR features. Among single-channel models, the best and second-best average F_1 scores of 37.47% and 37.45%, with related average accuracies of 51.14% and 52.02%, were achieved by a decision-tree-based model using interval BR features from channel P7 and an AdaBoost-based model using interval BR features from channel FC5, respectively. Through model-based feature selection with the aim of reducing the necessary number of physical channels, selected multi-channel models were developed; among these models, the best average F_1 score of 39.60% and a related average accuracy of 48.70% were achieved by a model using interval BR features from channels FC5, T7, and O1. The second best average F_1 score of 38.50% and a related average accuracy of 52.86% were obtained by the decision-tree-based model using interval BR features from T7 and P7 in the corresponding best set of bands; this result is the fifth-best F_1 score among all of the examined multi-channel models. This three-channel model achieved an average F_1 score equal to 94% of that of the best all-14-channel model based on the decision tree classifier using interval BR features.

In addition to the results summarized above concerning the best-performing attention detection models, the current work provides three valuable findings regarding the factors used to construct such models, which are based on an investigation of evaluated models. The models are generated using various combinations of several factors, such as the feature type, classifier, and set of channels. First, among the considered interval features (BR, BP, and AS), interval BR features are the most suitable for all considered types of attention detection models, including all-channel, single-channel, and selected multi-channel models. For both all-channel models and single-channel models using interval features from FC5 and P7, the best models using interval BR features achieve better performances in terms of average F_1 score and average precision than the best models using BP and AS features, with statistically significant differences as determined by one-way ANOVA ($p < 0.05$), across all investigated classifiers. Second, the superiority of interval EEG features was proven experimentally by comparing the performances of models using typical, intervalized, and interval EEG features. Compared with the performance of the best all-channel model using typical EEG features, that of the best model using the proposed interval EEG features was higher by 9.43% in the average F_1 score and by 112.61% in the average accuracy. Finally, all-channel, single-channel, and selected multi-channel models, which consider different volumes of EEG signals, were shown to require different classifiers by considering both the average F_1 score and the related average accuracy. SVM and decision tree classifiers are suitable for all-channel models, which use not only features from the most informative channels but also features from irrelevant channels. However, the average accuracies of the single-channel and multi-channel models based on the SVM classifier were decreased to below 41%. Decision tree and random forest classifiers are appropriate both for single-channel and selected multi-channel models, achieving average F_1 scores of greater

than 37% and average accuracies of greater than 50%. Decision-tree-based models using interval BR features achieved satisfactory performance in terms of both the average F_1 score and the average accuracy across all-channel, single-channel, and selected multi-channel models.

5.1. Implications of the selected multi-channel models

This study can provide valuable information for EEG-based intelligent systems using viewer attention. The EEG device used for the systems can be optimized with the subset of electrodes located in the fronto-central, parietal, temporal, and occipital lobes of the left hemisphere, which can be seen from the superiority of classification performance of the detection models that used the interval BR features from FC5, P7, T7, and O1 of viewers. In contrast to consumer-grade general-purpose EEG headsets, whose electrodes are mainly located in the anterior frontal (AF) and front (F) lobes, the consumer-grade EEG headset for detecting viewer attention can be made lightweight by excluding channels from the anterior frontal and front lobes and including the electrodes located in the fronto-central, parietal, temporal, and occipital lobes of the left hemisphere.

5.2. Future research directions

However, the proposed method of attention detection does not include the automatic segmentation of an entire video. In this study, all video segments were specified manually to enable the collection of precise attention levels for different video segments from the subjects because good-quality video segments with interval EEG features and attention class labels were necessary to train the proposed models. For the real-time localization of the attended video segments in newly tested videos, the present work could be combined with automatic video segmentation.

Although this paper provides some contributions as the first study of detecting attention paid to during video viewing, there are some weaknesses and limitations of our method. We enumerate four points to enhance and extend the proposed method for real-world intelligent applications.

1 First, the performance of the proposed method should be improved for real-world applications. The best all-channel and multi-channel models achieve a recall of better than 75% and a precision of worse than 30%. The possible use of the proposed model is limited to real-world scenarios requiring high recall and allowing low precision, for example, gathering highly attended video segments with minimized undetected detection results. Considering commercialization for the proposed method, we expect that the proposed method can be enhanced by adding or replacing some entropy EEG features, which have been recently applied for attention, as mentioned in Section 6, and by investigating other classifiers, such as deep neural networks, but not by using more channels from EEG caps.
2 Second, the proposed method can be combined with an automatic video segmentation method in order to be applied to an automatic video management system. Because good-quality video segments with interval EEG features and attention class labels were necessary to train the proposed models, we obtained video segments manually. To test our models of detecting viewer attention for other videos, the videos should be segmented in advance. Investigation of automatic video segmentation methods that are suitable for the proposed method will be required to determine the best combination of them.
3 Third, the proposed method is an off-line method that focuses on its classification performance and does not consider its time complexity. For real-time or online services for detecting viewer attention, the required time for feature extraction and prediction of the proposed method should be investigated, and a search can be performed for alternatives for features or classifiers if needed.
4 Finally, the proposed method can be applied to intelligent applications for rating videos based on viewer attention. The use of the proposed method is not limited to applications for customers, for example, a video management system. Applying the proposed method, video producers or providers can obtain trustworthy results of video ratings for some test videos at a low cost before their release through beta tests because the responses from participants through questionnaires or interviews can include user deception.

6. Related work

This chapter provides the literature review of EEG-based studies of recognizing attention-related user states, such as drowsiness, mental fatigue, and attention, proposed between 2016 and 2018.

As described in Table 9, the EEG-based studies of recognizing attention-related states, such as drowsiness, mental fatigue, and attention in learning, have mainly addressed the recognition of attentional states during the performance of specific tasks, such as driving, cognitive tasks, and learning. As described in Table 9, most EEG-based studies of recognizing detection (Awais et al., 2017; Shabani et al., 2016; Zhang et al., 2017) trained and tested their models using datasets collected during driving simulation because the studies aimed to classify drowsy and alert states while subjects were driving. To recognize mental fatigue induced by driving or performing mental tasks, Hu (2017) and Mydren and Chau (2017) classified normal and fatigue states of subjects while driving and performing some mental tasks, but Chai et al. (2016) and Li et al. (2016) classify the two states of subjects in a relaxed condition with eyes closed before and after performing a long driving task or mental tasks. The EEG-based studies of attention recognition (Djamal et al., 2016; Hu et al., 2018) classify the attention and inattention of subjects while driving and performing mental tasks. The studies did not include the task of video viewing.

Most previous studies classify equally spaced levels of drowsiness, fatigue, and attention, especially two in most cases. The drowsy and alert states are originally divided into two states, which can be determined by observers. The studies distinguished the levels of fatigue and attention by evenly dividing scores, which were rated by subjects, or clustering them into two groups (Mydren & Chau, 2017). However, the equally spaced levels of attention are not suitable for managing viewed videos because viewers need to localize not only attended video segments but also highly attended ones, with approximately the top 20% to 30% of attention, for sharing and rewatching them.

Although most of them achieved the best average accuracies of at least 80% and just over 95%, with the exception of Awais et al. (2017), their models are totally or partially dependent on subjects. In some of them, their personal models were trained and tested using instances collected from the same subject, although instances from the training and test sets used were partitioned. The other studies validated their models in k-fold cross-validation, and the instances from a subject could be divided into two and included in both the training and test datasets. However, the models applied to a product cannot be trained by using data from real customers after purchasing the product if the product does not allow customers to train their personal models. When considered for the commercialization of a product, the model should be validated in the LOSO cross-validation scheme.

The most common features used in the studies are features in the frequency domain using band power based on power spectral

Table 9
Comparisons between studies of drowsiness recognition.

	Tasks	Classes	# of ch	Feature and Classifiers	Classification Performance
Shabani et al. (2016)	Driving simulation Before the experiment, all subjects were instructed to be at least 22 hours awake	alert, drowsy - weighted KSS (Karolinska Sleepiness Scale) score by a subject and two supervisors	15	Feature: Determinism (DET) by Recurrence quantification analysis (RQA) Classifier: SVM	Average accuracy: 90.6% (subject-dependent, 10-fold cv for the selected subject)
Awais et al. (2017)	Driving simulation - 80 min driving on a highway, maintaining a maximum vehicle speed of 80 km/h	alert, drowsy - determined by using observance videos - alert/drowsy- 5-min period before/after a drowsiness event	20	Features: statistical features (per second), relative BP Classifiers: SVM	Average accuracy: 76.36% (subject-independent using LOSO, only using EEG)
Zhang et al. (2017)	Driving Simulation	alert, drowsy drowsy- between 4 to 6 a.m. after sleep deprivation alert- between 9 to 11 a.m. after normal sleep	8	BP of theta, alpha, beta bands Classifier: SVM	Average accuracy: 90.70% (subject-dependent, two-thirds for training, the others for testing)
Chai et al. (2016)	30–sec pre-task for alert 90-min-task for mental workload 30–sec post-task for fatigue	alert (pre-task), fatigue (post-task)	26	Feature: PSD of theta, alpha, beta, and gamma Classifier: Bayesian Neural Network	Average accuracy: 76% (partially subject-dependent: 2-fold cv in total)
Li et al. (2016)	5-min restful states with eyes closed and no action - during four time period (08:30–09:30, 11:00–12:00, 17:00–18:00, 22:00–23:00)	alert (FS = 0), slight fatigue (between 1 and 3), severe fatigue (between 4 and 6) - FS (Fatigue Score) rated by subjects	1	Features: Gravity Frequency, Power Percentage, and Sample Entropy of the temporal and frequency sequence Classifier: Deep Belief Network	Average Accuracy 98.86% (partially subject-dependent: two-thirds for training, the others for testing)
Hu (2017)	Driving simulation - the first 20-min driving - the next driving within 60–120 min.	normal, fatigue Normal - the last 5 min of the first driving Fatigue - the last 5 in of the next driving	32	Features: fuzzy entropy feature Classifier: SVM	Average F1-score: 97.4% (partially subject-dependent: 10-fold cv in total)
Myrden & Chau (2017)	Performing mental tasks - mental arithmetic, anagram solution, and a short-term spatial memory task	low, high - self-report levels of by subjects for fatigue and attention Labeling by grouping into l two classes	15	Features: PSD between 1 Hz to 30 Hz Classifier: LDA	Average accuracy: 74.8% for fatigue 84.8% for attention (subject-dependent: 10-fold cv)
Djamal et al. (2016)	Driving simulation with/without blue light alert	attention, inattention - driving with/without blue light alert	6	Features: PSD between 5 Hz to 30 Hz Classifier: SVM	Average accuracy: 80% (subject-dependent)
Hu et al. (2018)	Learning in a simulated distance learning environment during a 10 min. period	valence dimension: High/Neutral/Low - self-assessed level of attention	6	Features: Features from BPs (max, peak,sum power), entropy (Approximate) Entropy, Kolmogorov entropy, etc. Classifier: kNN	Average accuracy: 80.84% (subject-dependent: for each subject, 3-fold cv)

densities within a specific frequency range corresponding to a frequency, such as abstract band power, relative band power, the ratio of two band powers between two frequency bands, and the difference between two band powers between symmetric channels. In addition, some entropies, such as approximate entropy and sample entropy, of both the temporal and frequency EEG sequences were used.

There have been studies on generating an attention curve from the normalized PSD of EEG signals in order to extract keyframes for a video summary (Mehmood et al., 2016; Salehin & Paul, 2017). Although they provide the values of attention indices without training, the indices should be aggregated and sorted for a time period in advance in order to determine viewer states between attention and inattention for a specific video segment. In addition, the video datasets they used were biased to interest-inducing video genres, such as action, fantasy, adventure, and SF.

7. Conclusion

For the intelligent management of viewed video contents, a method for detecting user attention to video contents is required. At present, the average time spent per day with videos by US adults exceeds five hours. For the fine-grained sharing and rewatching of specific video contents with a time resolution of several seconds, viewers need a method of accessing the video segments to which they have paid the most attention. A fine-grained method detecting the levels of attention paid to video contents could be applied in various intelligent video applications, such as video summarization and fine-grained video bookmarking. The previous literature, however, has mainly addressed the detection of attention-related states during specific tasks. By contrast, research on user attention to video contents is in a very early stage.

To detect attention levels in the top 20% of viewer attention to video segments, a genre-independent dataset was constructed in this study by collecting EEG signals from 18 subjects while they viewed twelve videos of four different genres, and novel models were developed using interval EEG features based on descriptive statistics of typical EEG features. To ensure subject independence, all models were trained and tested through LOSO cross-validation. The best average F_1 score of 39.79% and a related average accuracy of 52.96% were achieved by a model based on a decision tree classifier using interval BR features from all 14 measured channels. Among single-channel models, a logistic-regression-based model using interval BR features from channel P7 achieved the best average F_1 score of 37.47% and a related average accuracy of 53.95%. Through model-based feature selection with the aim of reducing the necessary number of physical channels, the selected multichannel models using the interval EEG features from all four or some channels located in the fronto-central, parietal, temporal, and occipital lobes of the left hemisphere achieved the best and second best F_1 score of 39.60% and 38.50% with an average accuracy of 48.70% and 52.86%, respectively. It was shown that the models achieved better performance than models using the features from all four or some of their symmetric channels in the right hemisphere and models using the features from six channels located in the anterior-frontal and frontal lobes of the left and right hemispheres.

This paper presents an experimental demonstration of the superiority of the proposed interval EEG features and the importance of interval features in units of not seconds but intervals. Compared with the best all-channel model using typical BR features, the best all-channel model using interval BR features achieved improvements of 9.43% in the average F_1 score and 112.61% in the average accuracy. In addition, compared with the best all-channel model using intervalized BR features, in which the feature values for each second within a given time interval were set to the same statistical values calculated as the interval BR features for that time interval, the best all-channel model using interval BR features achieved improvements of 7.66% in the average F_1 score and 30.57% in the average accuracy.

It is expected that the proposed interval EEG features can be applied in place of typical EEG features to enhance previously developed models for detecting attention in a time interval, such as attention during learning and mental fatigue during the performance of metal tasks. In addition, the proposed method can be applied to measure user engagement with video-related contents, such as games, movie trailers, and advertisements, because the predicted attentional states can provide relevant information for both content producers and providers.

Acknowledgements

This work was supported by the ICT R&D program of MSIT/IITP. [B0101-15-0266, Development of High Performance Visual BigData Discovery Platform for Large-Scale Realtime Data Analysis].

References

Allen, J. J., & Kline, J. P. (2004). Frontal EEG asymmetry, emotion, and psychopathology: the first, and the next 25 years. *Biological Psychology, 67*(October (1–2)), 1–5.

Allen, D. P., & MacKinnon, C. D. (2010). Time-frequency analysis of movement-related spectral power in EEG during repetitive movements: a comparison of methods. *Journal of Neuroscience Methods, 186*(January (1)), 107–115.

Awais, M., Badruddin, N., & Drieberg, M. (2017). A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability. *Sensors, 17*(9).

Belle, A., Hargraves, R. H., & Najarian, K. (2012). An automated optimal engagement and attention detection system using electrocardiogram. *Computational and Mathematical Methods in Medicine, 2012*, 528781 1–12. doi:10.1155/2012/528781.

Chai, R., Tran, Y., Naik, G. R., Nguyen, T. N., Ling, S. H., & Craig, A. ,&Nguyen,H. T. (2016). Classification of EEG based-mental fatigue using principal. In *Proc. 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*.

Charbonnier, S., Roy, R. N., Bonnet, S., & Campagne, A. (2016). EEG index for control operators' mental fatigue monitoring using interactions between brain regions. *Expert Systems with Applications, 52*, 91–98.

Chuang, C. H., Ko, L. W., Lin, Y. P., Jung, T. P., & Lin, C. T. (2014). Independent component ensemble of EEG for brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 22*(2), 230–238.

Daimi, S. N., & Saha, G. (2014). Classification of emotions induced by music videos and correlation with participants' rating. *Expert Systems with Applications, 41*(13), 6057–6065.

Djamal, E. C., Pangestu, D. P., & Dewi, D. A. (2016). EEG-based recognition of attention state using wavelet and support vector machine. In *Proc. of the international seminar on intelligent technology and its applications (ISITIA)* (pp. 139–144).

eMarketer. (2015,. April 16). US Adults Spend 5.5 Hours with Video Content Each Day. Retrieved June 5, 2018, from https://www.emarketer.com/Article/US-Adults-Spend-55-Hours-with-Video-Content-Each-Day/1012362

Ericsson Consumerlab. (2016). *TV and media 2016: The evolving role of TV and media in consumers' everyday lives* November. An Ericsson Consumer and Industry Insight Report.

Gauba, H., Kumara, P., Roya, P. P., Singha, P., Dograb, D. P., & Ramana, B. (2017). Prediction of advertisement preference by fusing EEG response and sentiment analysis. *Neural Networks, 92*, 77–88.

Hu, J. (2017). Automated detection of driver fatigue based on AdaBoost classifier with EEG signals. *Frontiers in Computational Neuroscience, 11*(72).

Hu, B., Li, X., Sun, S., & Ratcliffe, M. (2018). Attention recognition in EEG-based affective learning research using CFS+KNN algorithm. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 15*(1), 38–45.

Khushaba, R. N., Kodagoda, S., Lal, S., & Dissanayake, S. (2011). Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *IEEE Transactions on Biomedical Engineering, 58*(1).

Li, Y., Li, X., Ratcliffe, M., Liu, L., Qi, Y., & Liu, Q. (2011). A real-time EEG-based bci system for attention recognition in ubiquitous environment. In *Int. workshop on ubiquitous affective awareness and intelligent interaction (UAAII)* (pp. 33–40).

Li, P., Jiang, W., & Su, F. (2016). Single-channel EEG-based mental fatigue detection based on deep belief network. In *Proc. 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*.

Li, Z., Chen, L., Peng, J., & Wu, Y. (2017). Automatic detection of driver fatigue using driving operation information for transportation safety. *Sensors, 17*(6).

Liu, N.-H., Chaing, C.-Y., & Chu, H.-C. (2013). Recognizing the degree of human attention using EEG signals from mobile sensors. *Sensors, 13*(August (8)), 10273–10286.

Malmivou, J., & Plonsey, R. (1995). *Bioelectromagnetism: Principles and applications of bioelectric and biomagnetic fields* (1st ed.). New York: Oxford University Press.

Mehmood, I., Sajjad, M., Rho, S., & Baik, S. W. (2016). Divide-and-conquer based summarization framework for extracting affective video content. *Neurocomputing, 174*(PartA), 393–403.
Moon, J., Kim, R., Lee, H., Bae, C., & Yoon, W. C. (2013). Extraction of user preference for video stimuli using EEG-based user responses. *ETRI Journal, 35*(6), 1105–1114.
Myrden, A., & Chau, T. (2017). A passive EEG-BCI for single-trial detection of changes in mental state. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25*(4), 345–356.
Petrantonakis, P. C., & Hadjileontiadis, L. J. (2010). Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on Affective Computing, 1*(2), 81–97.
Salehin, M. M., & Paul, M. (2017). Affective video events summarization using EMD Decomposed EEG signals (EDES). In *Proc. 2017 international conference on digital image computing: Techniques and applications (DICTA)*.
Sanei, S., & Chambers, J. A. (2008). *EEG signal processing*. Wiley.
Shabani, H., Mikaili, H., & Noori, S. M. R. (2016). Assessment of recurrence quantification analysis (RQA) of EEG for development of a novel drowsiness detection system. *Biomedical Engineering Letters, 6*(August (3)), 196–204.
Silveira, T. L. T., Kozakevicius, A. J., & Rodrigues, C. R. (2016). Automated drowsiness detection through wavelet packet analysis of a single EEG channel. *Expert Systems with Applications, 55*, 559–565.
Sörnmo, L., & Laguna, P. (2005). *Bioelectrical Processing in Cardiac and Neurological Applications* (1st ed.). Waltham, MA: Elsevier Academic Press.
Tatum, W. O., Husain, A. M., & Benbadis, S. R. (2008). *Handbook of EEG interpretation*. Demos Medical Publishing.
Welch, P. D. (1967). The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics, AU-15*(2), 70–73.
Wu, D., Lawhern, V. J., Gordon, S., Lance, B. J., & Lin, C.-T. (2016). Offline EEG-based driver drowsiness estimation using enhanced batch-mode active learning (EBMAL) for regression. *IEEE Trans. Fuzzy Systems, 25*(December (6)), 1522–1535.
YouTube. (2017,. February 27). You know what's cool? A billion hours, YouTube Official Blog, Retrieved June 5, 2018, from https://youtube.googleblog.com/2017/02/you-know-whats-cool-billion-hours.html
Zhang, X., Li, J., Liu, Y., Zhang, Z., Wang, Z., Luo, D., et al. (2017). Design of a fatigue detection system for high-speed trains based on driver vigilance using a wireless wearable EEG. *Sensors, 17*(March (3)).