Neural Computing and Applications (2023) 35:7925–7941
https://doi.org/10.1007/s00521-021-06591-6

S.I. : TAM-LHR

Neural correlates of affective content: application to perceptual tagging of video

Shanu Sharma[1] · Ashwani Kumar Dubey[1] · Priya Ranjan[2] · Alvaro Rocha[3]

Received: 18 June 2021 / Accepted: 24 September 2021 / Published online: 11 October 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

Over the past years, a digital multimedia uprising has been experienced in every walk of life, due to which the un-annotated or unstructured multimedia content has always been a key issue for research. The multimedia content is usually created with some intended emotions, which the creator wants to induce in viewers. The affectiveness of the multimedia content can be measured by analyzing elicited emotions of its viewers. In this paper, we present a rigorous study of human cognition using EEG signals while watching a video, to analyze the affectiveness of video content. The analysis presented in this paper is done to establish an effective relationship between video content and the human emotional state. For this, the most effective scalp location and frequency ranges are identified for two categories of videos, i.e., excited and sad. Furthermore, a common affective response (CAR) is extracted for finding the distinguishable features for aforementioned categories of videos. The CAR is calculated and tested on the publicly available dataset "AMIGOS," and the results presented here show the utility of cognitive features on extracted scalp locations and frequency ranges for automatic tagging of video content. The current research explores the innovative applicability of neuro-signals for a mouse-free video tagging based on human excitement level to augment a range of brain–computer interface (BCI)-based devices. It can further aid to automatically retrieve the video content which is exciting and interesting to human viewers. With this analysis, we aimed to provide a thorough analysis which can be used to customize a low-cost and mobile EEG system for automatic analysis and retrieval of videos.

Keywords Affective content · Video tagging · EEG signals · Brain–computer interface (BCI)

1 Introduction

Over the past decades, there is a remarkable expansion in the field of video generation, analysis, and display technologies [1]. It showed a huge transformation from analog television to high-definition (HD) displays, where noisy and low-resolution videos have been replaced by large displays called home theaters [1, 2]. In this technological revolution, a massive uprising of digital videos can also be seen in every walk of life, where high-quality digital videos have become ubiquitous, and is present in supermarkets, classrooms, shops, airports, and at other places [2]. Due to this extensive use of multimedia content in our daily life, un-annotated or unstructured multimedia content has always been a key issue for research. With the advancement of technologies, today a lot of progress has been done in the area of computer vision and pattern recognition for the development of reliable systems toward automatic

✉ Ashwani Kumar Dubey
dubey1ak@gmail.com

Shanu Sharma
shanu.sharma3@s.amity.edu

Priya Ranjan
ranjan.p@srmap.edu.in

Alvaro Rocha
amr@iseg.ulisboa.pt

1 Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, UP 201313, India

2 Department of Electronics and Communication Engineering, SRM University, Amaravathi, Andhra Pradesh 522240, India

3 ISEG, University of Lisbon, Rua Do Quelhas, N° 6, 1200-781 Lisbon, Portugal

multimedia processing [2, 3]. As this multimedia content is mostly targeted to human users, researchers are nevertheless working for the incorporation of subjective and cognitive aspects in existing systems to increase their performance. The inclusion of cognitive aspect corresponding to human's perception, reception and processing of information has always been a great interest among researchers [4, 5]. A continuous progress can also be seen in field of cognitive computing to mimic the human brain capabilities for solving variety of tasks such as data analysis, pattern recognition, etc. [2, 4]. Thus, human's cognition for affective content analysis of videos can facilitate the implicit tagging of multimedia content [5].

In recent decades, with the advancement of neuro- and vision science technology, today we have developed a deep understanding of the processing of visual information by the brain [4]. Researchers are now familiar with functional modeling of human neurons, i.e., which portion of the brain is used in the process of recognition, attention, navigation or decision-making [4, 5]. Furthermore, the field of cognitive psychology has provided us a way to understand human behavior by analyzing associated mental processes. Various brain activities can be decoded with the help of neural signal processing and brain–computer interfaces, whereas the understanding of neural correlates with human cognition is always been a topic of interest in neuroscience.

Based on these studies, it is believed that the neural information can also be used for the analysis of various video-based tasks like video quality of experience, the influence of video content on human perception or emotion detection in videos, etc. As the human brain outperforms the semantic interpretation of multimedia data for many applications, this paper aims at developing the video analysis technique using the concept of cognitive psychology which explores the concept of human cognition.

This paper makes the following key contributions as given as follows:

- First, the whole visual processing area of the human brain is analyzed and an attempt is made to extract the most effective scalp location for tagging two categories of videos, i.e., excited and sad.
- Second, we utilize EEG signals to understand "how humans think" employing five main frequency ranges, namely gamma, beta, alpha, theta, and delta of the EEG waves. Our primary focus is to find the most effective frequency band of EEG signals for distinguishing the two different visual stimuli.
- Third, rigorous computer simulations are conducted to analyze participant's EEG signals for aforementioned video categories to extract the participant's common affective response (CAR) w.r.t video stimuli. A detailed analysis is presented on simulation results to find the participant's differentiable CAR with the aim for developing an effective categorization of video stimuli.

The remaining paper is structured as: Sect. 2 describes the background and motivation to conduct this research. Section 3 presents the experimental paradigm of our work including dataset description and computation analysis. Further results achieved through computation analysis are discussed in Sect. 4. Section 5 presents key observations and findings of the proposed work, and Sect. 6 concludes the paper with the applicability of proposed findings.

2 Related work and motivation

Proper assessment of video is usually done by analyzing effective characteristics of the video. Video processing using these characteristics can provide a high correlation with the subjective scores of human observers. Till now manual analysis of video is the predominant method, but it is a slow attention intensive process where the human operator is not able to cope up with a flood of multimedia content generated on daily basis. In general, the manual process of video analysis involves the user's perception, thinking, and action steps in sequence [4]. For the past few decades, the multimedia research society is persistently trying to simulate the human's brain-behavior to provide cognitive abilities to machines [5]. However, deep understanding of this capability is still a formidable challenge.

In recent years due to the easy availability and decrease in the cost of EEG systems, the development of non-invasive brain–computer interfaces (BCI) devices is in great demand. EEG signals have been seen as the communication medium for automatic brain–robot interaction (BRI). It deals with the issues of identifying the human cognition activities by decoding their brain waves captured through an EEG device [6, 7]. EEG-based technology also has direct application in the rehabilitation of psychiatric or differently abled patients [8–11]. Extensive research has been conducted to analyze the emotional state of a person using EEG signals [12–16]. There have been some studies on identifying the human's cognitive state by decoding various physiological parameters of users in response to different stimuli. An emotion analysis related work is presented in [12], where authors have established the relationship between users' EEG responses and emotional judgment in response to different audio, visual and multimodal stimuli. With the recent development in the field of machine learning, researchers are also trying to incorporate deep learning models for automatic EEG signal classification for different emotional states [15, 17].

"Neuromarketing" is another emerging interdisciplinary field of neuroscience and psychology, where authors have tried to model the consumer's sub-conscious preferences toward marketing stimuli. It uses BCI to analyze the user's perception to derive affective information for decision-making [18, 19]. Though BCI has many outstanding applications, its applicability to analyze multimedia content is an emerging area of research. Authors [20–22] presented usability of EEG signal analysis based on rapid serial visual presentation (RSVP) of visual stimuli for human mental analysis while watching images. Lees et al. [23] presented a comprehensive review of RSVP-based BCI devices. Some literature [24–27] reveals the combined application of EEG and computer vision which was primarily focused on images. Mohedano et al. [25] explored the utility of BCI for segmenting object's images where an image is divided into several parts and are presented to the participants. Brain reactions of participants are then recorded as EEG signals. These EEG signals are then used as a response to estimate the probability of the location of an object of interest. It is also used to create the EEG map for the entire image. Furthermore, binarized EEG maps were used to seed the Grabcut object segmentation algorithm. Mohedano et al. [26] extended the work presented in [25] and demonstrated the use of BCI for image retrieval. On the other hand, Healy and Smeaton [27] showed the applicability of EEG for automatic searching of images.

Based on the aforementioned discussion, it is evident that BCI based on EEG technology is an emerging and hot topic of research. Nowadays, the interest of the researchers inclines toward examining multimedia content based on the human experience as characterized by their EEG response. Authors [28] showed a comprehensive result by trying to measure "interestingness," to perform cognitive tagging of videos, etc. Tauscher et al. [29] presented a review of three modalities, namely EEG, eye tracking, and user ratings for assessment of artifacts in videos and images. In another case, authors [30–34] showed some early EEG-based work for gaming, video categorization, and video summarization.

The well-known philosophy for a video to be considered good is as follows: "the video must hit the consumer's mind affectively that can be achieved through affective content within the video and this affectiveness largely depends upon the user's state of mind."

Motivated by the use of EEG signals for analyzing the human's mental state in response to visual stimuli, the work done in this paper is focused on presenting a thorough analysis to find the relationship between EEG and video content perception. We present a rigorous study of human cognition utilizing the EEG responses of the participants while watching a video. The research presented in this paper is primarily focused on the following points:

- The human brain has a complex architecture where different parts of the brain involved differently in the processing of information such as visual information is processed at occipital, frontal and parietal lobe whereas motor-related tasks are processed at motor cortex area. Furthermore, the high-end EEG devices for capturing the brain signals from all scalp locations are costly and their prices vary according to the density of the electrodes. In this paper, our work is focused on analyzing the entire visual processing area of the brain and, to extract the most effective scalp location for tagging two categories of videos, so that the analyzed results can be used to further customize a low-cost and mobile EEG system for automatic analysis of videos.
- EEG signal can successfully capture the electrical activity of the brain and the decomposed frequency ranges of captured EEG signals, i.e., alpha, beta, gamma, theta and delta represents the specific cognitive state of human. In this paper, the analysis is performed to find the most effective frequency range for differentiating the two categories of visual stimuli.
- Furthermore, to establish the relationship between participant's EEG responses and video content perception, we presented a common affective response (CAR) through combining the participant's cognitive response at the most effective scalp location and frequency ranges.

3 Methodology

The objective behind analysis of affective video content is to automatically recognize the emotion behind the video according to its content. As a human's cognitive state is directly linked with subjective experiences like happiness, excitement, etc. [35, 36], thus, analyzing the congruency/relationship between a human's cognitive state and multimedia content can be highly beneficial for inducing subjective measures in existing multimedia analysis systems like multimedia tagging, video summarization, etc.

During the last decade, a lot of progress has been done in the field of neuroscience for examining the brain structure of humans using different experimental methods like an electroencephalogram (EEG), functional magnetic resonance imaging (fMRI), positron emission tomography (PET), etc. It is found that any brain action whether it is conscious or unconscious can be captured by generated electrical activities in the subject's body. Thus, electrophysiological signals-based techniques, like EEG, are very useful for better understanding of human's cognitive state while performing any activity.

A human brain can be roughly classified into four regions frontal lobe, parietal lobe, temporal lobe, and occipital lobe as shown in Fig. 1 [5]. Each brain region handles the information processing task in a different way, e.g., if a person is doing any visual task, decision-related activities usually occur at the frontal lobe, whereas action-related activities are handled by the parietal lobe. For object recognition-related tasks usually, the temporal lobe is active and the occipital lobe is active during attention-related activities.

EEG signals can be used for analyzing the human's cognitive state by extracting different frequency ranges from the original waveform, as it is evident that for certain cognitive tasks, signals in a certain frequency ranges are usually more prominent at specific brain locations only. Different frequency ranges which can be used for cognition analysis are high-frequency waves such as gamma (> 30 Hs), beta (12–30 Hz), and low-frequency waves such as alpha (8–12 Hz), theta (4–8 Hz), and delta (< 4 Hz). It is evident that low-frequency ranges are usually connected with the unconscious intellect, while during relaxation state of brain alpha waves are highly active at occipital and parietal brain regions. Also, the attentiveness of mind is usually linked with high-frequency ranges at frontal and other regions of the brain. Due to the importance of frequency ranges at certain scalp regions, in this paper, an attempt is made for extracting cognitive-affective features using specific frequency band and scalp location for analyzing the affective state of the human's mind for two different visual stimuli with the aim of automatic tagging of videos based on their content.

3.1 Dataset description

For examining the cognitive effects of participants in response to visual stimuli, a publicly available dataset "AMIGOS: A dataset for affect, personality and mood research on individuals and groups" [37] is used. AMIGOS is a multimodal dataset that contains the neurophysiological signals recordings like an electroencephalogram (EEG), electrocardiogram (ECG), and galvanic skin response (GSR) of forty participants (13 female) of mean age of 28.3, while viewing sixteen short duration (< 250 s) and four long duration (~ 20 min) video clips. The dataset also contains information about participant's mood, personality, and affective responses to video content while watching videos as an individual (alone) and as part of the audience (i.e., in a group). Frontal HD videos, RGB, and full-body depth videos of participants are also recorded in the dataset.

In a short video experiment, each visual stimuli is selected based on its position in valence and arousal quadrant. The intensity of valence and arousal [38] represents the positive–negative experience and exciting-calming experience of any viewer as shown in Fig. 2. In the dataset, to elicit specific emotions, 16 videos of less than 250 s duration have been selected as visual stimuli such that each quadrant of valence/arousal consists of 4 videos. Out of four videos in each quadrant, 3 videos have been selected from the DECAF dataset [39] and one is selected from the multimodal dataset for affect recognition [40]. Figure 2 presents four-quadrant, namely LVHA, HVHA, LVLA, and HVLA of valence-arousal where H, L, A and V, respectively, denoted as high, low, arousal and valence. During each trial, EEG signals have been recorded using a fourteen-channel Emotiv EPOC Neuroheadset with a 128 Hz sampling rate and 14-bit resolution. EEG electrodes for channels, AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4, were placed according to the 10–20 electrode system. After each trial, participants have been asked to rate the video on a scale of valence, arousal, dominance, familiarity, and liking.

In this paper for extracting, the effective cognitive features in response to visual stimuli, EEG recordings of 40 participants during a short video experiment for two

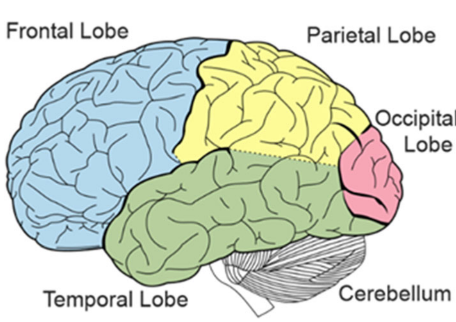


Fig. 1 Brain structure [5]

Fig. 2 Valence-Arousal quadrant

categories corresponding to 8 videos in the individual case are selected. We selected the HAHV and LALV video categories for the current research. Table 1 presents the information of videos taken during simulations along with their original video ID of the dataset [39, 40].

Trial structure In short video experiment, 8 videos correspond to two categories of videos (HAHV, LALV) have been presented to each participant in 8 different trials randomly. Before each trial, self-assessment of participants has been recorded to assess their basic emotions and level of valence and arousal. Then, sensory inputs are recorded for 5 s fixation period in which a fixation cross is presented to the participant, followed by the display of visual stimuli. After each trial, self-assessment of a participant is done to record their after-watch effects of stimuli as shown in Fig. 3.

Self-assessment/internal annotation Self-assessment is recorded before and at the end of each trial to analyze the affective state of the participant while performing the task. To do this, participants are asked to provide their feedback on the dimension of valence, arousal, dominance, liking, and familiarity. Then, they have been asked to provide at least one emotion from the category of basic emotions like neutral, happiness, sadness, surprise, fear, anger, and disgust.

3.2 Neurophysiological signals (EEG) analysis

To analyze the relationship between participant's cognitive state and affective video content, neurophysiological signals, i.e., EEG signals are used to record the participant's implicit response corresponding to affective video content. EEG signals are usually highly sensitive to noise; thus, extraction of meaningful information from them is a challenging task and it requires a lot of pre-processing effort. The computational procedure adopted for extracting the distinguishable cognitive features from EEG signals is depicted in Fig. 4. EEG data corresponding to 40 subjects for eight different videos of HAHV and LALV category as mentioned in Table1 are selected for analysis. EEG signals are then pre-processed followed by extraction of different frequency bands using the discrete wavelet transform (DWT) method. Finally, frequency domain features are extracted to find the effective bands and channel locations for analyzing the affective video content.

3.2.1 Raw EEG data

EEG data have been recorded continuously using 14 electrode positions at a sampling rate of 128 Hz. EEG data of all participants while watching eight short videos of categories HAHV and LALV in 8 trails are selected for further analysis. Sample EEG data corresponding to HAHV and LALV video categories at electrode position 'AF3' is shown in Fig. 5. Here, the time–amplitude domain representation of EEG waveform represents the differentiable cognitive response of participants for two categories of videos.

3.2.2 EEG data pre-processing

EEG signals usually have very low voltage variations. These signals are typically affected by various artifacts like blinks, muscle movement, and line noise, etc. Preliminary pre-processing of EEG signals to remove noise is done on EEGLAB software. The approach presented by authors in [37] is used to remove noise from EEG signals. To remove eye artifacts from all electrode positions, the blind source separation technique is adopted, where to eliminate the artifactual activity from EEG signals, an independent component analysis (ICA)-based approach is used. The average re-referencing is done using "Compute Average Reference" option in EEGLAB Toolbox. Furthermore, EEG data are filtered from 4 to 45 Hz using (FIR) finite impulse filter.

3.2.3 Data selection and frequency band isolation

Various brain regions are involved in the processing of visual information differently. Thus, to focus on EEG dynamic induced by visual stimuli, data on all 14 channels positions as shown in Fig. 6 are used for analysis.

Table 1 Details of videos used

Video source	Video id [39, 40]	Video_No. AMIGOS [37])	Category
Airplane	4	12	HAHV
When Harry Met Sally	5	13	
Hot Shots	9	16	
Love Actually	80	15	
Exorcist	19	5	LALV
My girl	20	6	
My bodyguard	23	7	
The Thin Red line	138	3	

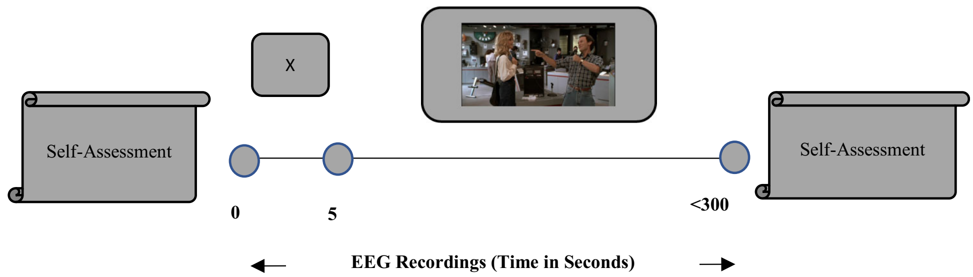


Fig. 3 Data acquisition procedure for each stimulus

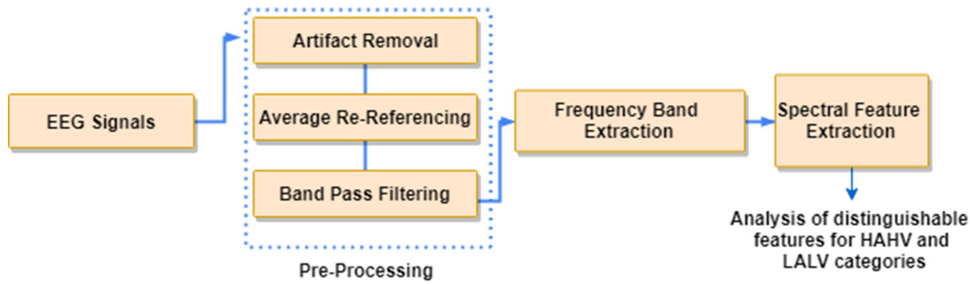


Fig. 4 Computational procedure for EEG data analysis

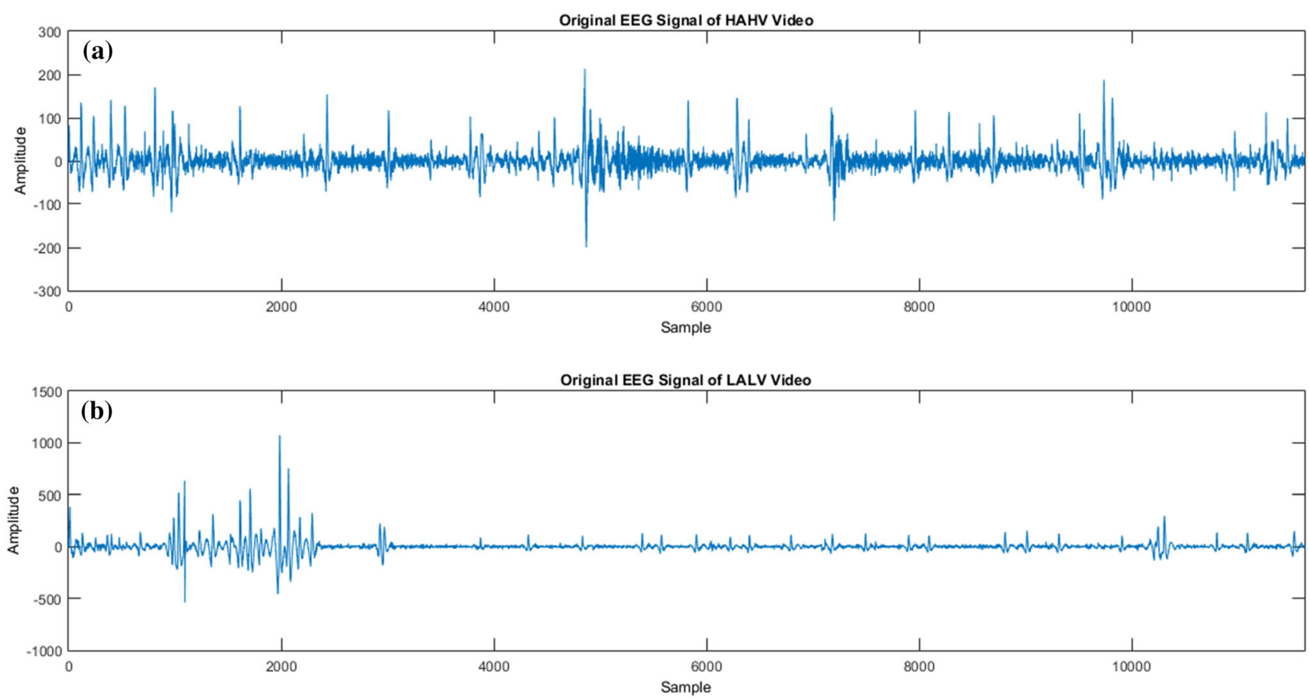


Fig. 5 Raw EEG data of **a** HAHV **b** LALV type video

Furthermore, for analyzing human brain responses and emotional activity, five main frequency bands have been extracted from EEG signals using the DWT-based approach.

EEG signals are usually non-stationary in nature. Thus, to find out an appropriate spectral component from them, conversion of signals from the time-domain to frequency domain is needed. Wavelet transform (WT) is a well-known method of research in terms of analyzing non-stationary signals in the frequency domain. It provides the multi-resolution description of a non-stationary signal [41]. Multi-resolution analysis of a signal deals with analyzing the signal at different frequencies with different resolutions. In wavelet transform, this can be achieved by convolving the signal with a small oscillatory function called the mother wavelet with its different translation and scaled versions. Here, translation represents the location of the window which provides the time information in the transformed domain. The term scale is used to represent the global and detailed information about the signal.

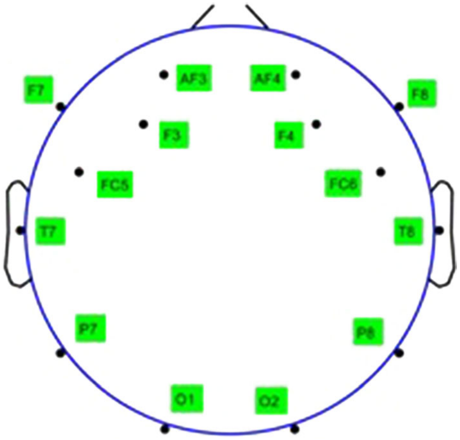


Fig. 6 Layout of used EEG channels

DWT performs a multi-resolution analysis of a signal by decomposing it into a number of scales. In terms of frequency, low frequencies, i.e., high scales represent the global information about the signal whereas high frequencies, i.e., narrow scales correspond to the detailed view of the signal. In DWT, for analyzing the signal at a different resolution, filters of different cutoff frequencies are usually used [42]. Thus, in DWT-based decomposition of the signal, consecutive convolution of signals with a high pass (HP) and low pass (LP) filters is performed to decompose it into various high and low frequencies components, respectively [43, 44]. Furthermore, scaling of a signal can be performed using a down-sampling operation, which is a process of reducing the sampling rate of the signal.

In current research for performing the wavelet transformation on signal, Daubechies-four (db4) mother wavelet is used. Daubechies wavelets are a popular wavelet family for signal transformation, which is used to perform orthogonal multi-resolution analysis on the signal. After testing different wavelets from the Daubechies wavelet family, the Db4 wavelet is chosen due to its high performance and less computational complexity than other wavelets. Db4 mother wavelet and its scaling function are shown in Fig. 7.

Furthermore, for DWT-based decomposition, the number of decomposition levels is determined using the dominant frequency component of the signal. As EEG data used in this analysis are recorded at a 128 Hz sampling rate, thus decomposition level of four is chosen to decompose the signal in required frequency ranges, i.e., gamma, beta, alpha, theta, and delta.

Scaling [$\phi_{j,k}(n)$] and wavelet [$\psi_{j,k}(n)$] functions [42] are dependent on low and high pass filters, respectively, as given in Eqs. (1) and (2), respectively.

$$\varnothing_{j,k}(n) = 2^{-j/2} h\left(2^{-j} n - k\right) \tag{1}$$

$$\psi_{j,k}(n) = 2^{-j/2} g\left(2^{-j} n - k\right) \tag{2}$$

where n = 0, 1, 2….M − 1, j = 0, 1, 2… J − 1, k = 0, 1, 2… 2^j − 1. M represents the length of the signal, and J represents no. of levels, i.e., 4. At each level of decomposition, DWT generates approximate (A_i) and detailed coefficients (D_i) by applying successive high pass and low pass filtering with a down-sampling rate of 2 [43]. Approximate coefficient (A_i) and detailed coefficient (D_i) at the ith level are determined using Eqs. (3) and (4), respectively.

$$A_i = \frac{1}{\sqrt{M}} \sum_n x(n).\varnothing_{j,k}(n) \tag{3}$$

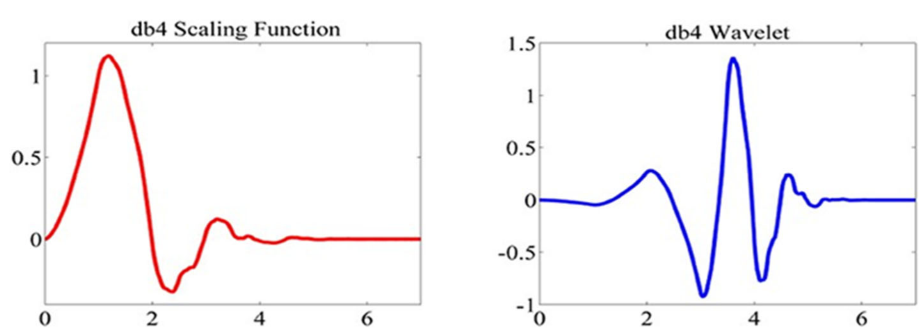
$$D_i = \frac{1}{\sqrt{M}} \sum_n x(n).\psi_{j,k}(n) \tag{4}$$

The approximation coefficients are then further decomposed to extract localized information from the sub-band of detailed coefficients as shown in Fig. 8.

Among extracted sub-bands as given in Table 2, alpha, beta, gamma, and theta bands have been used for analyzing human responses to the affective content of video.

Figure 9 shows the frequency bands (as depicted in Table 2 in column 3) for two videos from HAHV and LALV categories.

Fig. 7 Mother wavelet and scaling function [34]



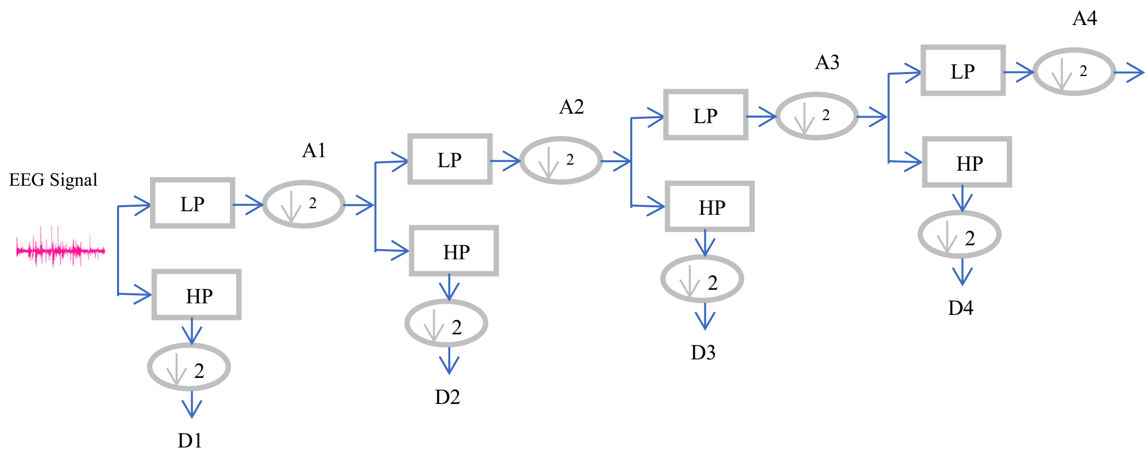


Fig. 8 DWT-based signal decomposition into detailed and approximate coefficients

Table 2 Extracted EEG frequency bands

DWT coefficients	Frequency range	Frequency bands
D_1	32–64	Gamma
D_2	16–32	Beta
D_3	8–16	Alpha
D_4	4–8	Theta
A_4	0–4	Delta

3.2.4 Cognitive feature extraction

The most common method for analyzing the dominant frequency band in response to an affective human's cognitive state is to investigate the absolute or relative power of signals. In the current research, power spectral density (PSD) using Welch [45] method is extracted for fixation and stimuli period separately. For EEG data acquisition, each trial starts with 5 s baseline period (as discussed in Sect. 3.1) followed by a video stimuli period. The length of each trial varies according to the duration of a video. As described in Table 1, four different video trials for both categories (HAHV and LALV) of videos have been presented to the participants. Videos in both quadrants of HAHV and LALV are denoted as $HAHV_i$ and $LALV_i$ (i = 1.0.4 for each category per participant), respectively. Furthermore, for each participant, $HAHV_1^{fixation}$ and $HAHV_1^{stimuli}$ represent the first video in the HAHV category for fixation and stimuli period, respectively, for the

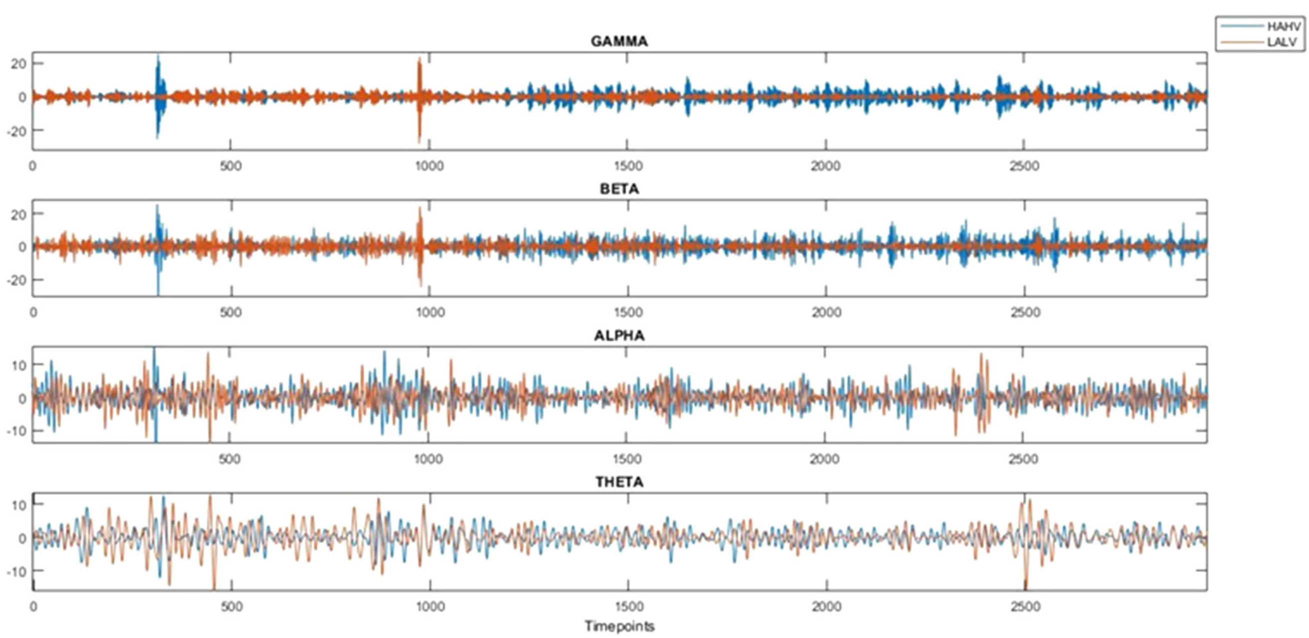


Fig. 9 Various extracted bands for HAHV and LALV category of video

EEG signals. For analyzing the effect of video on a human's mental state, power spectral features are extracted on fixation and visual stimuli period separately. In each video trial, the spectral power is calculated in all frequency bands mentioned in Table 2. Then, average EEG power spectra for all 14 channel locations (Fig. 6) in frequency bands alpha, beta, gamma, and theta are calculated in fixation and stimuli period separately for both HAHV and LALV categories of videos. To reduce the inter-subject variance of EEG signals, power values in the stimuli period of each video and subject have been normalized w.r.t fixation (baseline) period using Eq. (5). The normalization of power will remove all those activities from EEG signals which were constant over that period and are not related to visual stimuli. The normalized power is calculated for each frequency band for all 8 trials for 40 participants on 14 channel locations.

$$P^{N}_{c,b} = \left(P^{\text{stimuli}} - P^{\text{fixation}}\right)/P^{\text{fixation}} \quad (5)$$

where c represents channels [1–14], b represents bands [1–4] and P^N represents normalized power, P^{stimuli} and P^{fixation} represent power values in stimuli and fixation period for channel c and b. After normalization, EEG signals of one video trial in both HAHV and LALV categories are characterized by feature matrix (F) of dimension 14 (channels)* 4 (frequency bands). Furthermore, the power values of each video's category have been averaged over participants using Eq. (6). The outcome of Eq. (6) is used to analyze the effective frequency bands and channels. It helps to investigate the distinguishable power features of LALV and HAHV categories of videos.

$$P_{\text{avg}} = \sum_{i=1}^{40} P_i^N \quad (6)$$

Response of HAHV and LALV videos on human's cognitive state in terms of normalized average power in alpha, beta, gamma, and theta bands on different scalp locations is presented in Sect. 4. Furthermore, its significance w.r.t frequency band and scalp positions are also analyzed and discussed.

4 Results and discussions

In the present study, an attempt is made for extracting the cognitive feature based on human's EEG responses on content of video. It can be used to automatically tag a video as affective or otherwise. As EEG signals are highly complex and temporal in nature, hence extraction of meaningful information from EEG signals is difficult. Time-domain representation of a participant for both HAHV and LALV categories of videos at FC5 location is presented in Fig. 10, which represents the complex behavior of EEG response in the time domain.

Furthermore, the absolute power of EEG response is plotted in Fig. 11. It shows that in a time-domain representation of EEG response, clear discrimination is not visible. Most distinguished information of the signal is usually contained in its frequency content. In addition, human cognitive behavior is usually examined by considering certain frequency ranges of brain signals. Hence, conversion of the time-domain signal into the frequency domain is done.

In the frequency domain, a signal is characterized in terms of its component frequencies. The strength of the signal in a particular frequency bin can be used for different interpretations like how much participant is active or not during the recording period. Power spectra of EEG response of a participant for HAHV and LALV categories of videos at all channels are presented in Fig. 12. Figure 12 shows a clear distinguishable behavior of signal and it is visible in the frequency domain. Following this analysis, we extracted various frequency bands using DWT for further analysis.

It has been found in the literature that based on the cognitive activity of the brain, EEG responses captured at different brain regions are noticeable in specific frequency ranges only. Therefore, after the extraction of DWT-based frequency bands, responses of these different frequency ranges, i.e., gamma, beta, alpha, and theta were analyzed to establish their relationship with affectiveness of video content. Furthermore, to find the most active brain regions to differentiate two types of videos, all channel locations were also analyzed to extract the most prominent brain locations for measuring the affectiveness of video. For this, we have performed the analysis of brain responses with respect to each frequency band and channel location. Average normalized power, i.e., P_{avg} at all scalp locations and bands in both HAHV and LALV categories of videos has been used for finding the effective bands and channel locations. The analysis of effective frequency bands and channels using average normalized power features is presented in subsequent sections.

4.1 Frequency bands analysis

Previously, the researchers tried to link various emotional and working states of the human mind with different frequency ranges. It is noticed that higher frequency ranges, i.e., beta and gamma waves are related to high engagement of brain activity. Alpha waves are known as idling or relaxation rhythms, and they are generally high when eyes are closed. On the other hand, theta waves are deal with the emotional process, daydreaming, stress and frustration. Average power content in different frequency ranges of

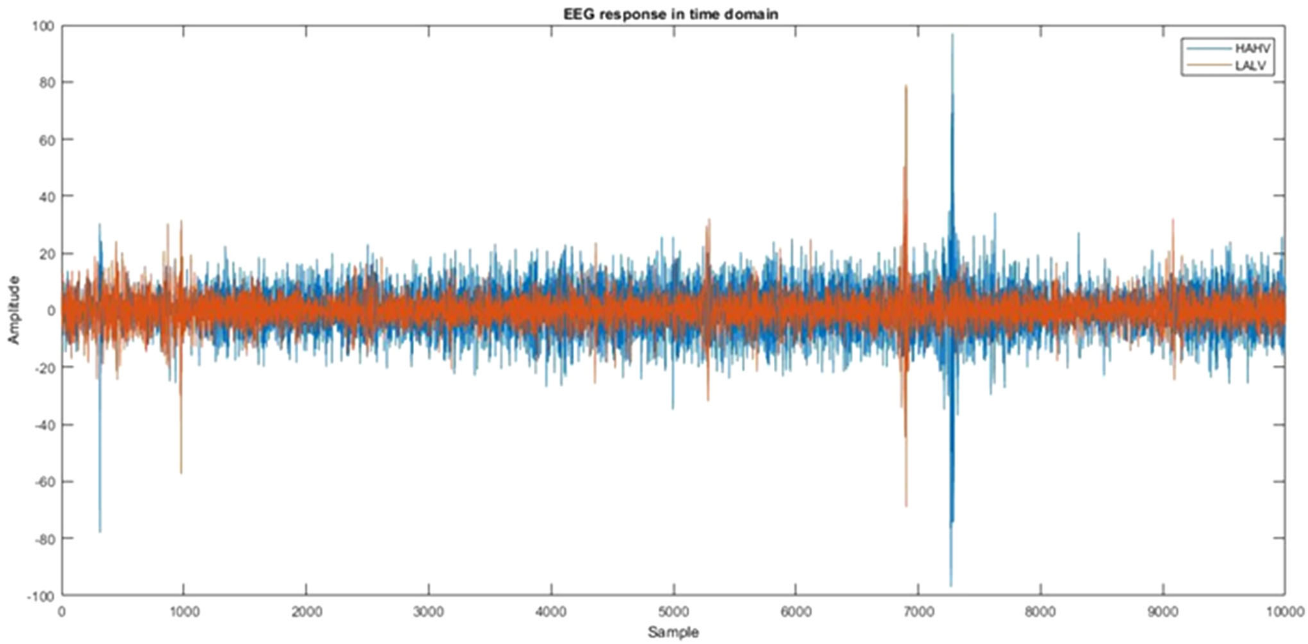


Fig. 10 EEG response of participant for HAHV and LALV at FC5

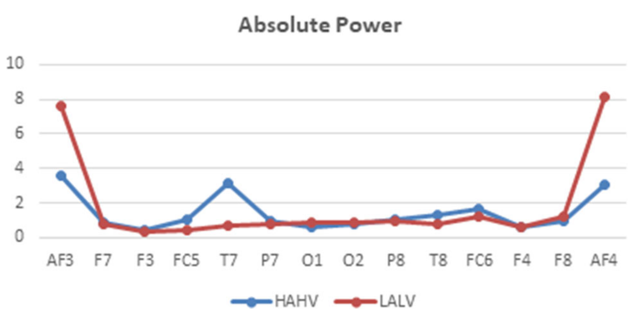


Fig. 11 Absolute Power for HAHV and LALV

EEG responses at all channel locations for HAHV and LALV category of videos is given in Table 3. The cognitive aspect of these power values at a particular band and channel with a human's affective state is discussed in the next subsection along with the graphical representations.

Cognitive aspect The distribution of normalized power across all channels in the theta band is presented in Fig. 13. Theta bands represent the sub-conscious state of humans and it is usually found in daydreaming and sleep. In healthy participants, these waves are not visible in excess during waking hours. Thus, a lower amount of power distribution across all channels is visible for both HAHV and LALV categories of video. A clear difference is also visible in the result on various visual processing scalp locations, like frontal and occipital channels which represent the visual information processing related tasks. The result indicates that HAHV videos are more affective than LALV mainly, because various visual processing areas are involved in processing visual information which results in the reduction of theta band power.

The alpha band represents a human's relaxation state and is usually found in large amounts during the eye's close state. A reduced amount of alpha power represents the processing of information by the brain; thus, the brain will be more active in this situation. They are usually prominent at the occipital and parietal lobe which can be seen as responses at electrodes O1, O2, P7 and P8 as depicted in Fig. 14. Results indicate that low power distribution in alpha waves during HAHV videos is directly correlated with human's high affective state in response to a particular video.

A high-frequency ranges such as gamma and beta frequency waves deals with the effective engagement of the brain. Beta waves represent the human attentional state and they appear in large amounts when there is focused concentration. Similarly, large gamma waves can be seen during high cognitive functioning state. We noticed more power values in beta and gamma waves (Figs. 15, 16) for the HAHV category as compared to LALV. It indicates more human response to HAHV video category than the LALV.

4.2 Effective channel analysis

To find the most effective brain regions for representing a human's cognitive state during the video stimuli period, the difference of normalized power at all channels and bands is calculated. Table 4 represents HAHV and LALV PSD difference across all bands and channels.

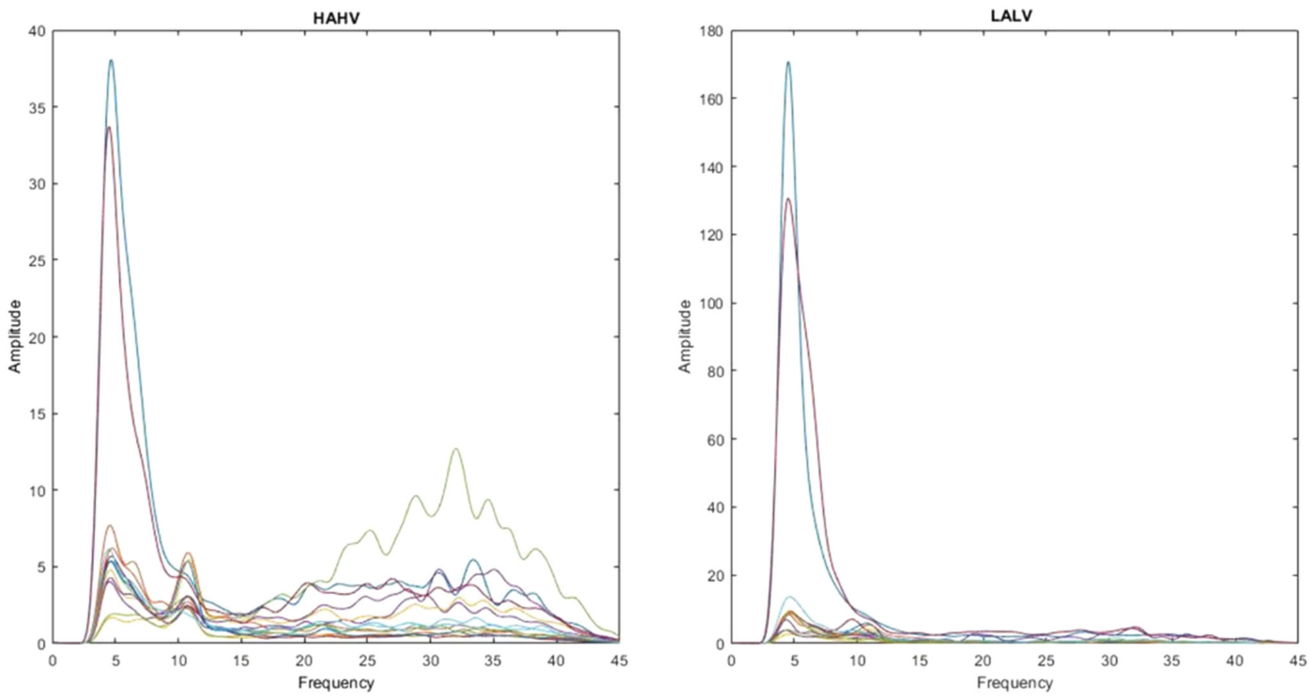


Fig. 12 Power spectra of one participant for HAHV and LALV

Table 3 Average normalized power values across all channels and bands

Band/Channel	Theta		Alpha		Beta		Gamma	
	HAHV	LALV	HAHV	LALV	HAHV	LALV	HAHV	LALV
AF3	0.16	0.32	0.30	0.40	0.68	0.76	0.81	0.81
F7	0.29	0.30	0.47	0.41	0.66	0.52	0.72	0.45
F3	0.43	0.53	0.80	0.78	0.89	0.91	0.79	0.76
FC5	0.26	0.32	0.48	0.48	0.70	0.60	0.80	0.56
T7	0.27	0.41	0.53	0.65	1.00	0.61	1.33	0.45
P7	0.34	0.48	0.68	0.78	0.80	0.63	0.88	0.52
O1	0.30	0.42	0.81	0.86	0.82	0.75	0.85	0.50
O2	0.34	0.47	0.67	0.79	0.69	0.81	0.74	0.55
P8	0.31	0.40	0.65	0.67	0.71	0.65	0.75	0.53
T8	0.26	0.36	0.57	0.59	0.88	0.63	0.98	0.55
FC6	0.29	0.44	0.54	0.56	0.86	0.66	1.04	0.51
F4	0.49	0.85	0.63	0.77	0.78	0.73	0.85	0.79
F8	0.34	0.42	0.42	0.48	0.67	0.52	0.89	0.50
AF4	0.14	0.31	0.27	0.45	0.73	0.81	0.75	0.80

The difference of PSD values for HAHV and LALV video categories is calculated and plotted based on the frequency bands analysis presented in Sect. 4.2. The difference is taken and plotted as LALV-HAHV for theta and alpha bands, whereas for beta and gamma bands the difference is plotted as HAHV-LALV.

Comparison of difference of power values for HAHV and LALV video categories at all bands and channel locations are presented in Fig. 17. It can be seen that theta and alpha bands are having very little difference of $< 20\%$ at all channel locations except frontal location F4. As theta band is usually visible in large amounts during sleep state and alpha band during relaxation state, a little difference in human's cognitive state represents consistent relation with the working nature of alpha and theta bands.

Furthermore, as beta and gamma waves represent an active engagement of the brain, the result shows a large amount of difference at temporal locations T7 and T8 for beta waves, whereas the gamma band is visible as most effective at the major brain portion, as the difference can be

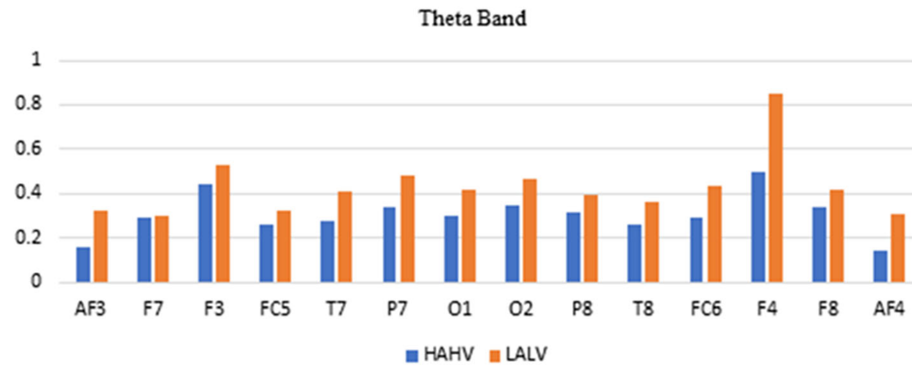


Fig. 13 Distribution of theta power values across all channels

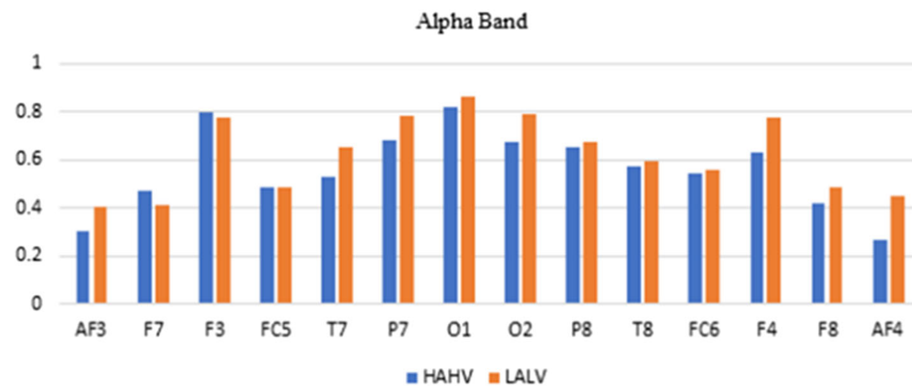


Fig. 14 Distribution of alpha power values across all channels

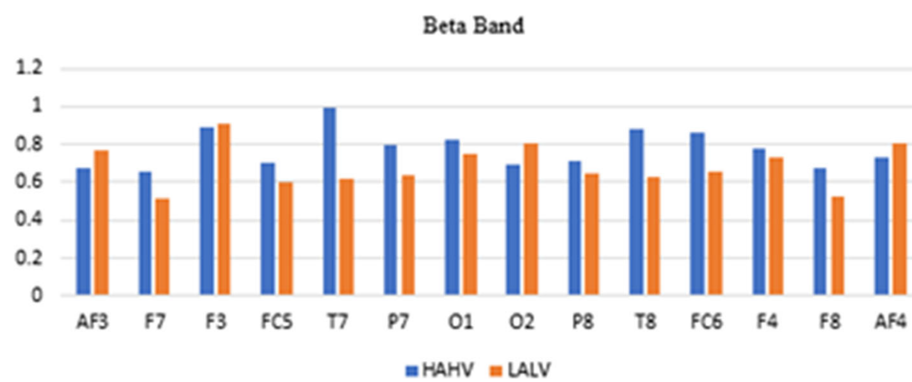


Fig. 15 Distribution of beta power values across all channels

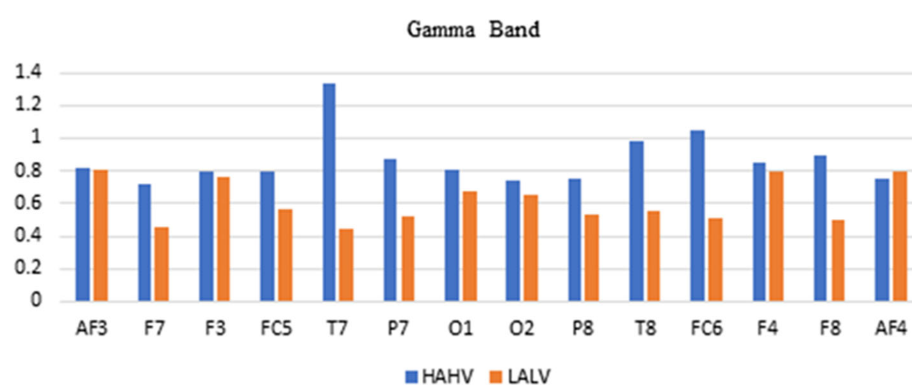


Fig. 16 Distribution of gamma power values across all channels

Table 4 HAHV and LALV PSD difference across all bands and channels

PSD_Diff	Theta	Alpha	Beta	Gamma
AF3	0.16	0.1	− 0.08	0.00
F7	0.01	− 0.06	0.14	0.27
F3	0.1	− 0.02	− 0.02	0.03
FC5	0.06	0.00	0.10	0.24
T7	0.14	0.12	0.39	0.88
P7	0.14	0.10	0.17	0.36
O1	0.12	0.05	0.07	0.13
O2	0.13	0.12	− 0.12	0.09
P8	0.09	0.02	0.06	0.22
T8	0.1	0.02	0.25	0.43
FC6	0.15	0.02	0.2	0.53
F4	0.36	0.14	0.05	0.06
F8	0.08	0.06	0.15	0.39
AF4	0.17	0.18	− 0.08	− 0.05

seen at frontal (F7, FC5, FC6, F8), temporal (T7 and T8), parietal (P7 and P8), and occipital (O1, O2) location. This clear difference of EEG responses in both categories can be used for modeling a human's cognitive state during visual stimuli.

5 Key observations and findings

This study is planned to ascertain a successful relationship between induced emotions in participants and the video content so that a low-cost neural device can be generated for automatic tagging of videos based on their content. In this section, we highlighted the key observations of our work based on the experimental outcomes and analysis presented in the previous section.

- First, we showed that as slow frequency waves alpha and theta deals with the relaxation and sleep behavior of the subject, thus, a little difference between the two categories of videos is visible at alpha and theta bands.
- Second, we noticed that as high-frequency ranges such as gamma and beta shows an active cognitive situation of human, thus, a higher difference for two categories of videos across all brain regions is noticed at gamma and beta frequency waves.
- Third, we found that a considerable difference in signal values is located in the frontal lobe around electrodes F7, FC5, FC6, F8, temporal lobe around T7, T8, parietal lobe around P7, P8, O1, and O2 of the occipital lobe.
- Fourth, to present the differentiable cognitive feature in response to video stimuli, common effective response (CAR) is determined by combining the power values at the most effective channel locations and frequency ranges.
- Finally, we presented the extracted cognitive feature in terms of the combination of most effective channel locations and frequency ranges as presented in Table 5. The obtained results also justify the cognitive aspect of human behavior. The information processing related to attention is processed at the occipital lobe, whereas the frontal lobe represents working memory. Temporal and parietal lobe deals with spatial movement and object perception related tasks. As these four regions are highly involved in the processing of visual information, the same effect is shown in Table 5 at high-frequency range, i.e., gamma, which shows the active cognitive state of participant.

It can be seen in the literature that a lot of work has been done on participant's emotional state classification using EEG signals while watching a video [46, 47]. For the classification of a particular type of emotion, the studies classified the collected EEG signals according to the valence-arousal emotion space. With this work, our goal is not to classify the emotional state of participants using

Fig. 17 Visualization of PSD difference for HAHV and LALV

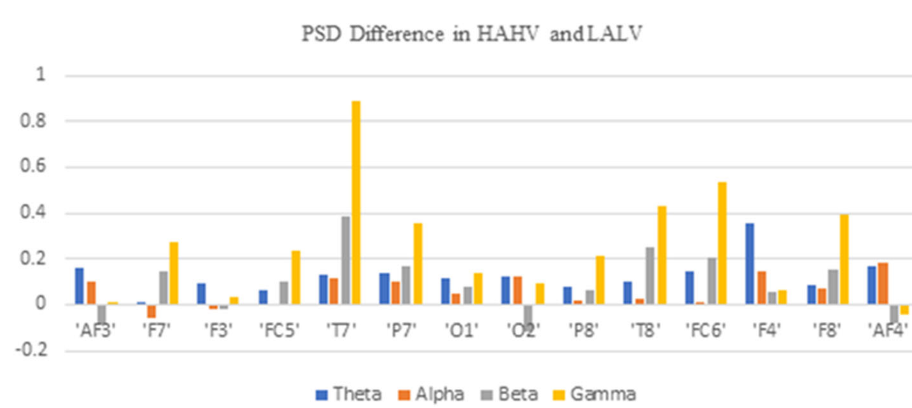


Table 5 Comparative analysis of gamma power values at different brain locations

Video Categories	Frontal (F7, FC5, FC6, F8)	Temporal (T7, T8)	Parietal (P7, P8)	Occipital (O1, O2)
HAHV	86.25	57.75	40.75	79.5
LALV	50.5	25	26.25	52.5

EEG signals. Instead, we are here establishing a relationship between a particular video category according to its content and participant's cognitive state, so that this relationship can be further used to develop a more effective computational model and BCI-based devices for automatic tagging of videos. Here, the interesting use of EEG can

Table 6 Comparative analysis of proposed work

Paper	Task	Data Used	Analysis
Soleymani et al. [28]	Here authors used EEG signals to analyze the participant reaction for matched and unmatched tags for each video to perform tag validation	Self-collected data of 30 participants using 32 active electrodes on 10–20 international systems using a Biosemi Active II system	To analyze participant's responses to matched and unmatched tags, authors extracted PSD-based features from five frequency bands to train the SVM classifier
Tauscher et al. [29]	The authors compared three different modalities, i.e., user ratings, eye tracking, and EEG data to analyze the participant's perception against artifacts in videos	Self-collected data of 10 participants using a gamma cap2 with 16 active electrodes.	Here authors have used EEG as one of the modalities to analyze participant's perception for video quality. The ERP (Event-related potential) based analysis is done on EEG signals for establishing the relation between EEG signals and video quality
Mutasim et al. [30]	The authors presented a wireless EEG-based video classification model to automatically classify three categories of videos	Self-collected data from 23 participants using Muse headband with 5 Channels	Authors tried various combinations of feature extraction methods like DWT, FFT, STFT, etc., and different classifiers to test the model on each channel location. Authors claimed accuracy of 80% on one channel location AF8, whereas the thorough analysis behind the selection of particular features, classifier, and channel location is missing
Salehin et al. [34]	Here authors presented a model to perform video summarization using EEG signals	Self-collected data	Empirical mode decomposition (EMD) based analysis is done on EEG signals to analyze different frequency components called IMF (Intrinsic mode function). Here authors generated a neuronal attentional curve using EEG features to summarize video
Sai Sukruth Bezugam et al. [48]	Here a video summarization approach is presented using EEG and Eye-tracking Signals	Self-collected EEG data of 15 participants using Brain Vision acti Champ with 64 channels	A CWT (continuous wavelet transform) based analysis is done on EEG signals to generate the attention curve of human perception. Video summarization is then performed by extracting the important events using extracted EEG signals-based attention curve
Proposed work	An affective video content analysis approach is presented for automatic recognition of elicited emotions by videos	AMIGOS Dataset of 40 participants at 14 channel locations with video clips of benchmark dataset for affect recognition [39, 40]	A DWT-based analysis is done on all frequency components to extract a combination of the most effective frontal, temporal and parietal channel locations of the brain for affective tagging of videos. The analysis presented here matches with the human's cognitive behavior and can be used to generate low-cost EEG-based automatic tagging of videos based on their content

provide the possibility of implicit indexing, i.e., the automatic video indexing or tagging can be done by simultaneously analyzing the EEG data without conscious effort.

In the literature, researchers have also used EEG signals for multimedia content analysis [20–34]. Some of the work related to EEG signals based on video analysis is summarized in Table 6, to differentiate our work from the existing work. The most relevant research related to our work is presented by Mutasim et al. [30], as authors tried to model EEG signals for automatic classification of three video categories. The authors tried a combination of different features and classification models to get good classification accuracy on each five-channel data, whereas the proper reason for selecting a particular channel location, features, and classifier is not mentioned properly. Our work is also motivated by work presented by authors in [34] and [48], where authors performed a rigorous analysis on extracted frequency bands to extract the attentional curve for extracting highlights from the video.

6 Conclusion

In this paper, a method is proposed to establish the relationship between the EEG response and video content perception. Here, extensive computer simulations have been conducted on a standard publicly available dataset "AMIGOS" [37] to find the distinguishable features of two categories (HAHV and LALV) videos. We demonstrated that the proposed system is capable of establishing the relationship between the EEG response and video content perception. Here, the proposed system is robust and works successfully based on the EEG response. We presented a thorough analysis in Sect. 4 by accounting cognitive aspects of all high and low-frequency ranges at most effective brain regions. We summarized the experimental outcomes in Sect. 5 by presenting key observations. We strongly believe that the results and analysis presented in this paper suggest a robust neural correlate to video content and it will motivate the researchers to develop the computational model for affective video content-based applications using EEG signals in the near future.

Acknowledgements The study presented in this paper is based on the dataset "AMIGOS" [37] collected by Juan Abdon et al. In this dataset, a wide range of physiological parameter collection of participants is done using EEG, ECG, and GSR for videos taken from all valence/arousal quadrants. We would like to thank the developers for providing us access to this dataset.

Authors' contributions SS contributed to conceptualization, methodology, software, data curation, validation, and writing—original draft preparation. AKD contributed to conceptualization, methodology, supervision, reviewing and editing. PR contributed to conceptualization, supervision, reviewing and editing. A contributed to reviewing and editing.

Funding Not applicable.

Availability of data and material AMIGOS.

Code availability Custom code.

Declarations

Consent to participate I consent to participate.

Consent for publication I consent for publication.

Informed consent The article contains methodologies performed on publicly available data AMIGOS. As per the dataset, description participants have provided the written consent before participation to the developers.

Ethics approval Not applicable.

Human and animal rights This article does not contain any studies with animals performed by any of the authors.

Conflict of interest There is no conflict of interest.

References

1. Caviedes JE (2012) The evolution of video processing technology and its main drivers. Proc IEEE 100(4):872–877. https://doi.org/10.1109/JPROC.2011.2182072
2. Dimitrova N, Zhang HJ, Shahraray B, Sezan I, Huang T, Zakhor A (2002) Applications of video-content analysis and retrieval. IEEE Multimed 9(3):42–55. https://doi.org/10.1109/MMUL.2002.1022858
3. Smith MA, Chen T (2005) 9.1: image and video indexing and retrieval. In: Bovik AL (ed) In communications, networking and multimedia, handbook of image and video processing, 2nd edn. Academic Press, New York. https://doi.org/10.1016/B978-012119792-6/50121-2
4. Isola P, Xiao J, Parikh D, Torralba A, Oliva A (2013) What makes a photograph memorable? IEEE Trans Pattern Anal Mach Intell 36(7):1469–1482. https://doi.org/10.1109/TPAMI.2013.200
5. Müller V (2008) Margaret A. Boden, Mind as machine: a history of cognitive science, 2 vols. Mind Mach 18:121–125. https://doi.org/10.1007/s11023-008-9091-9
6. Hassanien AE, Azar A (2014) Brain computer interfaces: current trends and applications, intelligent systems reference library, vol 74. Springer, Cham
7. Ghaemmaghami P (2017) Information retrieval from neurophysiological signals. Ph.D. thesis, University of Trento
8. Yang Y, Bloch I, Chevallier S, Wiart J (2015) Subject-specific channel selection using time information for motor imagery brain-computer interfaces. Cogn Comput 8:505–518. https://doi.org/10.1007/s12559-015-9379-z
9. Duan L, Bao M, Cui S, Qiao Y, Miao J (2017) Motor imagery EEG classification based on kernel hierarchical extreme learning machine. Cogn Comput 9:758–765. https://doi.org/10.1007/s12559-017-9494-0
10. Padfield N, Zabalza J, Zhao H, Vargas VM, Ren J (2019) EEG-based brain-computer interfaces using motor-imagery: techniques and challenges. Sensors. https://doi.org/10.3390/s19061423

11. Kumar S, Riddoch MJ, Humphreys G (2013) Mu rhythm desynchronization reveals motoric influences of hand action on object recognition. Front Hum Neurosci 7:66. https://doi.org/10.3389/fnhum.2013.00066
12. Hiyoshi-Taniguchi K, Kawasaki M, Yokota T, Bakardjian H, Fukuyama H, Cichocki A, Vialatte FB (2015) EEG correlates of voice and face emotional judgments in the human brain. Cogn Comput 7:11–19. https://doi.org/10.1007/s12559-013-9225-0
13. Li J, Zhang Z, He H (2018) Hierarchical convolutional neural networks for EEG-based emotion recognition. Cogn Comput 10:368–380. https://doi.org/10.1007/s12559-017-9533-x
14. Gawali BW, Rao S, Abhang P, Rokade P, Mehrotra SC (2012) Classification of EEG signals for different emotional states. In: Fourth international conference on advances in recent technologies in communication and computing (ARTCom2012), pp 177–181. https://doi.org/10.1049/cp.2012.2521
15. Frydenlund A, Rudzicz F (2015) Emotional affect estimation using video and EEG data in deep neural networks. In: Barbosa D, Milios E (eds) Advances in artificial intelligence. Canadian AI 2015. Lecture notes in computer science, vol 9091. Springer, Cham. https://doi.org/10.1007/978-3-319-18356-5_24
16. Alarcao SM, Fonseca MJ (2018) Emotions recognition using EEG signals: a survey. IEEE Trans Affect Comput. https://doi.org/10.1109/TAFFC.2017.2714671
17. Roy Y, Banville H, Albuquerque I, Gramfort A, Falk TH, Faubert J (2019) Deep learning-based electroencephalography analysis: a systematic review. J Neural Eng 16(5):051001. https://doi.org/10.1088/1741-2552/ab260c (**PMID: 31151119**)
18. Vecchiato G, Cherubino P, Maglione AG, Ezquierro MT, Marinozzi F, Bini F, Trettel A, Babiloni F (2014) How to measure cerebral correlates of emotions in marketing relevant tasks. Cogn Comput 6:856–871. https://doi.org/10.1007/s12559-014-9304-x
19. Gupta A, Shreyam R, Garg R, Sayed T (2017) Correlation of neuromarketing to neurology. IOP Conf Ser Mater Sci Eng 225:012129. https://doi.org/10.1088/1757-899X/225/1/012129
20. Bigdely-Shamlo N, Vankov A, Ramirez RR, Makeig S (2008) Brain activity-based image classification from rapid serial visual presentation. IEEE Trans Neural Syst Rehabil Eng 16(5):432–441. https://doi.org/10.1109/TNSRE.2008.2003381
21. Wang J, Pohlmeyer E, Hanna B, Jiang YG, Sajda,P, Chang SF (2009) Brain state decoding for rapid image retrieval. In: Proceedings of the 17th ACM international conference on multimedia, pp 945–954. ACM, New York. https://doi.org/10.1145/1631272.1631463
22. Huang Y, Erdogmus D, Pavel M, Mathan S, Hild KE (2011) A framework for rapid visual image search using single-trial brain evoked responses. Neurocomputing 74(12):2041–2051. https://doi.org/10.1016/j.neucom.2010.12.025
23. Lees S, Dayan N, Cecotti H, McCullagh P, Maguire L, Lotte F, Coyle D (2018) A review of rapid serial visual presentation-based brain- computer interfaces. J Neural Eng 15(2):021001. https://doi.org/10.1088/1741-2552/aa9817
24. Kapoor A, Shenoy P (2008) Combining brain computer interfaces with vision for object categorization. In: 2008 IEEE conference on computer vision and pattern recognition, pp 1–8. https://doi.org/10.1109/CVPR.2008.4587618
25. Mohedano E, Healy G, McGuinness K, Giró-i-Nieto X, O'Connor NE, Smeaton AF (2014) Object segmentation in images using EEG signals. In: Proceedings of the 22Nd ACM international conference on multimedia, pp 417–426. ACM, New York. https://doi.org/10.1145/2647868.2654896
26. Mohedano E, McGuinness K, Healy G, O'Connor NE, Smeaton AF, Salvador A, Porta S, Nieto XG (2015) Exploring EEG for object detection and retrieval. In: Proceedings of the 5th ACM on international conference on multimedia retrieval, pp 591–594. ACM, New York. https://doi.org/10.1145/2671188.2749368
27. Healy G, Smeaton AF (2011) Optimising the number of channels in EEG-augmented image search. In: Proceedings of the 25th BCS conference on human–computer interaction, pp 157–162. British Computer Society, Swinton
28. Soleymani M, Pantic M (2013) Multimedia implicit tagging using EEG signals. In: 2013 IEEE international conference on multimedia and expo (ICME), San Jose, CA, USA, 2013, pp 1–6. https://doi.org/10.1109/ICME.2013.6607623
29. Tauscher JP, Mustafa M, Magnor M (2017) Comparative analysis of three different modalities for perception of artifacts in videos. ACM Trans Appl Percept. https://doi.org/10.1145/3129289
30. Mutasim AK, Tipu RS, Bashar MR, Amin MA (2017) Video category classification using wireless EEG. In: Zeng Y, He Y, Kotaleski JH, Martone M, Xu B, Peng H, Luo Q (eds) Brain informatics. Lecture notes in computer science, vol 10654. Springer, Cham, pp 39–48. https://doi.org/10.1007/978-3-319-70772-3_4
31. Nussbaum PA, Herrera A, Joshi R, Hargraves R (2012) Analysis of viewer EEG data to determine categorization of short video clip. Procedia Comput Sci 12:158–163. https://doi.org/10.1016/j.procs.2012.09.047
32. Wehbe RR, Kappen DL, Rojas D, Klauser M, Kapralos B, Nacke LE (2013) EEG-based assessment of video and in-game learning. CHI Ext Abstr. https://doi.org/10.1145/2468356.2468474
33. Moon J, Kim Y, Lee H, Bae C, Yoon WC (2013) Extraction of user preference for video stimuli using EEG-based user responses. ETRI J 35(6):1105–1114. https://doi.org/10.4218/etrij.13.0113.0194
34. Salehin MM, Paul M (2017) Affective video events summarization using EMD decomposed EEG signals (EDES). In: 2017 international conference on digital image computing: techniques and applications (DICTA), pp 1–6. https://doi.org/10.1109/DICTA.2017.8227402
35. Baveye Y, Chamaret C, Dellandréa E, Chen L (2018) Affective video content analysis: a multidisciplinary insight. IEEE Trans Affect Comput 9(4):396–409. https://doi.org/10.1109/TAFFC.2017.2661284
36. Hanjalic A, Xu L (2005) Affective video content representation and modeling. IEEE Trans Multimed 7(1):143–154. https://doi.org/10.1109/TMM.2004.840618
37. Correa JAM, Abadi MK, Sebe N, Patras I (2018) AMIGOS: a dataset for affect, personality and mood research on individuals and groups. IEEE Trans Affect Comput 12(2):479–493. https://doi.org/10.1109/TAFFC.2018.2884461
38. Kossaifi J, Tzimiropoulos G, Todorovic S, Pantic M (2017) AFEW-VA database for valence and arousal estimation in-the-wild. Image Vis Comput 65(C):23–36. https://doi.org/10.1016/j.imavis.2017.02.001
39. Abadi MK, Subramanian R, Kia SM, Avesani P, Patras I, Sebe N (2015) DECAF: MEG-based multimodal database for decoding affective physiological responses. IEEE Trans Affect Comput 6(3):209–222. https://doi.org/10.1109/TAFFC.2015.2392932
40. Soleymani M, Lichtenauer J, Pun T, Pantic M (2012) A multimodal database for affect recognition and implicit tagging. IEEE Trans Affect Comput 3(1):42–55. https://doi.org/10.1109/T-AFFC.2011.25
41. Akansu AN, Haddad RA (2001) Chapter 6: wavelet transform. In: Akansu AN, Haddad RA (eds) Multiresolution signal decomposition, 2nd edn. Academic Press, London, pp 391–442. https://doi.org/10.1016/B978-012047141-6/50006-9
42. Kehtarnavaz N (2008) Chapter 7: frequency domain processing. In: Kehtarnavaz N (ed) Digital signal processing system design, 2nd edn. Academic Press, London, pp 175–196. https://doi.org/10.1016/B978-0-12-374490-6.00007-6
43. Vivas EL, García-González A, Figueroa I, Fuentes RQ (2013) Discrete wavelet transform and ANFIS classifier for brain-

machine interface based on EEG. In: 2013 6th international conference on human system interactions (HSI), pp 137–144. https://doi.org/10.1109/HSI.2013.6577814
44. Subasi A (2007) EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst Appl 32:1084–1093. https://doi.org/10.1016/j.eswa.2006.02.005
45. Welch P (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans Audio Electroacoust 15(2):70–73. https://doi.org/10.1109/TAU.1967.1161901
46. Doma V, Pirouz M (2020) A comparative analysis of machine learning methods for emotion recognition using EEG and peripheral physiological signals. J Big Data 7:18. https://doi.org/10.1186/s40537-020-00289-7
47. Hu X, Chen J, Wang F, Zhang D (2019) Ten challenges for EEG-based affective computing. Brain Sci Adv 5(1):1–20. https://doi.org/10.1177/2096595819896200
48. Bezugam S, Majumdar S, Ralekar C, Gandhi T (2021) Efficient video summarization framework using EEG and eye-tracking signals. ArXiv: arXiv:2101.11249

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.