ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 9, Issue 12, June 2020

Exploration on programmed feature scene extraction frameworks by using EEG

Tadanbu Misawa, Masashi Miyazaki, Yasuhiro Inazumi, Shigeki Hirobayashi
Graduate School of Science and Engineering for Research, University of Toyama

***Abstract*— In the administration field, the frameworks created depend on mind capacities. Inclination happens on account of mental variables in different circumstances. For instance, there is inclination and enthusiasm for specific recordings. Assessment of recordings has been examined by using EEG, yet these frameworks could just perform assessments. Consequently, we attempted to build up a programmed extraction framework for feature scenes dependent on inclinations in the media field. To build up this framework, we led three confirmation tests. These investigations demonstrated the connection between scenes extricated utilizing mind waves and client chose feature scenes. Our outcomes recommend that, specifically, the base adequacy estimation of fleeting information of alpha waves is a powerful component for feature scene extraction.**

***Index Terms*—Scene, EEG.**

I. INTRODUCTION

The use of EEG is predicted to spread to ordinary homes in the future, facilitated even further by reduced costs. For this reason, systems that use brain activity have been recently attracting the attention of the public. Systems using EEG have been developed and investigated for their use in brain-computer interface (BCI) systems in the welfare field. For example, there are text entry systems (called “spellers”) which use EEG [1]. Research on EEG-based systems is not restricted to the welfare field but has also spread to the business management field. Present research in the business management field is aimed at trying to elucidate the brain functions associated with purchase decision making. “Preference” is one such brain function; it is considered a psychological factor that occurs in various daily situations. For example, there is preference for particular videos. On the other hand, EEG has been used in studies that involve the evaluation of image quality and the estimation of emotions when watching videos [3]-[8].

Currently, the highlight scenes of videos are chosen by an editor. There is also an objective method for the same purpose that is determined by sound and moving image processing. However, these methods do not include the psychological factors (preferences) of video viewers. It may be possible to extract scenes according to the psychological factors of video viewers using EEG, which would make it possible to detect highlight scenes with higher accuracy than through conventional methods.

Manuscript received: 20 May 2020
Manuscript received in revised form: 13 June 2020
Manuscript accepted: 02 July 2020
Manuscript Available online: 15 July 2020

Therefore, in this study, we investigated whether highlight scenes can be extracted from EEG signals to develop a highlight scenes extraction system using EEG. In order to develop the system, three experiments were conducted in this research. In Experiment 1, we investigated whether a highlight scene can be detected in the brain waves. For this purpose, the brain activity of 20 subjects during video watching was measured by EEG. In Experiment 2, an evaluation experiment was carried out using the data obtained in Experiment 1. The scenes extracted for each feature were given a score. In Experiment 3, an evaluation experiment was conducted using a system that automatically extracts scenes based on brain wave features. In this system, EEG electrodes were attached to the subject, a video was shown, and then the scenes were extracted based on the features of the brain waves immediately after video watching. In addition, a score was given to the extracted scenes. From these three experiments, we investigate the possibility of extracting highlight scenes using EEG.

A. Related Work

This section introduces previous research related to extracting highlight scenes from EEG. In subsection (1), we introduce previous research about emotional evaluation using EEG. In subsection (2), we introduce previous research related to video summarization.

(1) Emotional Evaluation

Elucidation of the emotional mechanisms of the brain has progressed by using EEG measuring equipment, and studies are being made to try to apply it to emotion estimation and user interfaces. Two important factors in emotion are the emotional value (valence) indicating the direction of the evoked emotion as positive or negative, and the strength of the evoked emotion, which is the degree of arousal. Valence is a concept of psychology expressing whether the direction of emotion is positive or negative, and is used as an indicator of basic emotion discrimination. Therefore, brain activity when researching positively or negatively evaluated music and images when presented to subjects is currently being investigated. It has been reported that the response in the α-band (8-13 Hz) increases in the left side of the frontal lobe during positive tasks and in the right side of the frontal lobe during negative tasks [7], [10].

(2) Video Summarization Method

Owing to the excessive video content available in present times, studies are under way to develop a system that enables users to understand summaries of video contents and search

DOI:10.17605/OSF.IO/VXWYT

or manage video content. By characterizing through image processing, speech recognition, and character recognition as an objective method of summarizing videos, it is possible to apply it to a state transition pattern of a flow in a hidden Markov model to determine highlight scenes. Other studies to detect highlight scenes have also been reported [11]. In such studies, highlight scenes are determined by fitting new data using image, sound, and character features from highlight scenes that were defined as such in advance.

Thus, studies aimed at improving the precision of summarization techniques by analyzing the emotions of viewers along their biological signals have been reported.

II. MATERIALS AND METHODOLOGY

A. *Equipment and Setup*

EEG signals were recorded with four-channel ProComp Infiniti equipment (Thought Technology Inc). The sampling rate of EEG signals was 256Hz. The electrode positions are shown in Fig. 1, according to the 10-10 system. The positions marked in red (Fz, Cz, O1, O2) were used in Experiment 1 and the positions marked in blue (F3, Fz, Oz, O2) were used in Experiment 3. For analysis, only brain wave data from electrode Fz, which is the area of the brain related to emotions, was used. In a previous study, this electrode was placed at a position that was related to pleasant emotions during video watching [17]. Subjects performed experimental tasks in a dark room. When the audio of the video was used, subjects wore earphones. The impedance of each electrode was required to be less than 5 kΩ.

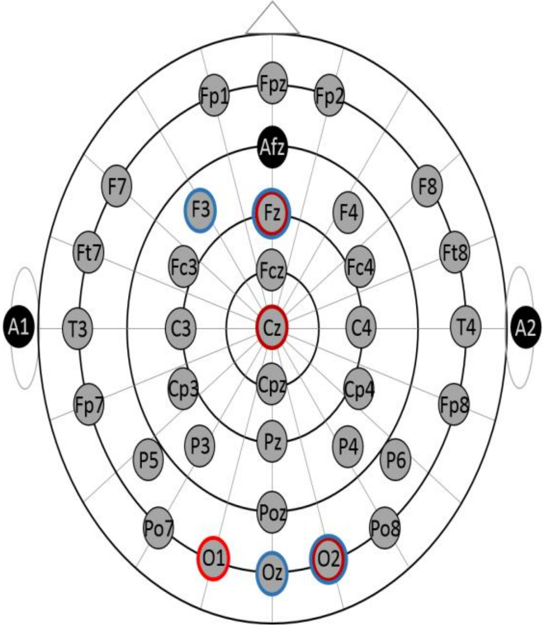

Fig. 1: Electrode positions

B. *Analysis Methods*

In this research, we conducted three experiments. Since the analysis methods differ for each experiment, we will explain each one separately. The analysis method of each experiment is shown in Fig. 2. The window length of the short-time Fourier transform (STFT) was 256 points, and shift length, the amount of change per frame, was 8 points.

- ***Experiment 1***

In Experiment 1, we investigated whether the scene extracted from the EEG signals of the group was related to the highlight scene. Therefore, the data processed by the STFT was averaged. After that, in order to remove noise, the temporal data was processed with a low pass filter (cutoff frequency = 5 Hz).

- ***Experiment 2***

In Experiment 2, the brain wave data obtained in Experiment 1 was used. The data after STFT processing was normalized by Z-score. This was done because the amplitude value differs for each individual. Next, the same averaging and low pass filtering process of Experiment 1 was performed.

- ***Experiment 3***

Experiment 3 was carried out using the system described in section *IV*. This system extracts brain wave features for each individual. The data processed by the STFT was low pass filtered.

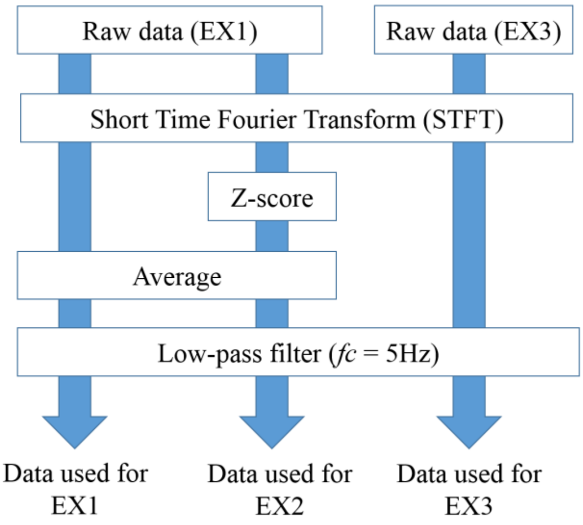

Fig. 2: Analysis method for each experiment

C. *Videos*

In this research, commercial messages (CM) were used as our experimental videos. CMs were 30 seconds long and had 30 fps. The CMs contained messages that the editor wanted to convey during a short time. We tried to extract those messages in a short time using brain features. We evaluated 126 CMs as "interesting" or "not interesting". We selected 50 CMs as experimental videos (30 CMs evaluated as interesting and 20 CMs evaluated as not interesting).

III. PROPOSED SYSTEM

A. *System Overview*

In this section, the proposed system will be explained. The system overview is shown in Fig. 3. This system was also used in EX3. The experimental procedure was as follows.

(1) The user was made to watch the moving images and brain activity was measured by EEG during that time.

(2) In order to obtain the features, frequency domain analysis and statistical processing are performed on the measured brain wave data.

(3) The highlight scenes were determined based on the features obtained from the electrode on the forehead, which is related to emotion, and are displayed on the display.

B. Features

EEG data was analyzed using the STFT. The bandwidth used in the STFT were the α-band (8-13Hz), the β-band (13-30Hz), and the γ-band (30-50Hz). The γ-band is generally 30-70 Hz, but since the power supply noise was at 60 Hz, it was defined as 30-50Hz.

The features used in each experiment are given in Table 1. The meaning of Max in Table 1 corresponds to the maximum amplitude value in the temporal data of the brain waves during video watching. Similarly, the meaning of Min in Table 1 corresponds to the minimum amplitude value in the temporal data of the brain waves during video watching. In addition, and since audio features were used in EX2 and EX3, the value with the maximum volume was used. Volume Max is one of the features in the video summarization method. Volume is an important feature for extracting exciting scenes without depending on the content of the video. Therefore, it was used as a feature for comparison. In Table 1, Random means that scenes were randomly extracted. Random scenes were used to investigate whether they were evaluated higher than scenes extracted appropriately. A method for extracting a scene using these features will be described in subsection *C*.

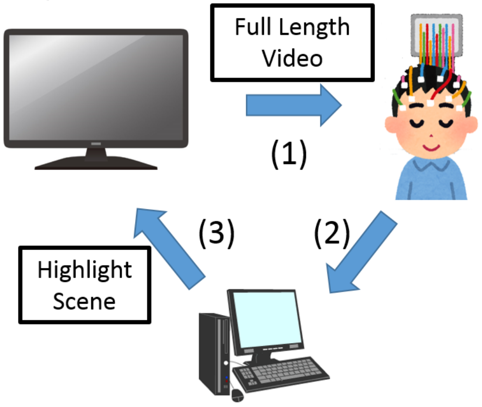

Fig. 3: Outline of the proposed system

Table I: Features used for each experiment

Feature	EX1	EX2	EX3
EEG	α-Max	α-Max, Min β-Max γ-Max	α-Max, Min β-Max γ-Max
Audio		Volume Max	Volume Max
Random			Random

C. Extraction Method

In this subsection, we will describe the highlight scene extraction method based on the aforementioned features. In order to extract the most characteristic scene from each video, it was extracted in the form of shots. In order to divide the video into shots, Adobe Speed Grade was used.

Fig. 4 is an explanatory diagram of the extraction method. Fig. 4 shows the α-wave time series data. The following example serves to explain the case in which the feature to be extracted is α-Max. In Fig. 4, the red circle indicates the maximum amplitude value (which is α-Max). In this case, the system extracts Shot 8, which corresponds to α-Max, and the extracted shot is displayed.

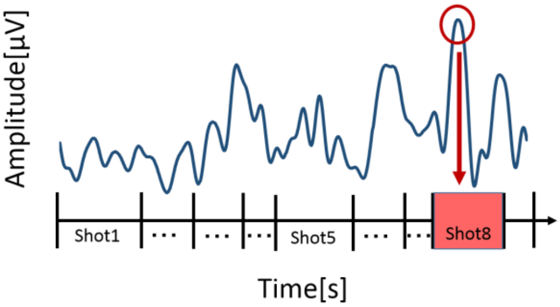

Fig. 4: Shot extraction method

IV. EXPERIMENT 1

Twenty healthy Japanese participants took part in our research after providing written informed consent. All participants were university students aged between 21 and 24 years.

A. Task

Fig. 5 shows the outline of the experiment. This experimental task involved a sequential process that consisted of a 20 s rest period, 30 s of video watching, and an evaluation of the watched video with either of two options (interesting, not interesting) using the mouse. Each subject underwent a total of 50 trials (about 1 hour), which were conducted in two blocks of 25 trials each with a ten-minute break in between. Subjects were sitting in a dark room throughout the experiment.

B. Evaluation Method

To evaluate the estimated highlight scenes obtained from the acquired EEG data, we investigated whether these highlight scenes were appropriate by subjective evaluation. Fig. 6 presents a schematic diagram of the subjective estimation method. Ten subjects watched ten randomly chosen videos. Subjects selected six highlight scenes from each video and ranked them as “TOP1,” “TOP2,” and “TOP3”. Since this study’s goal is to demonstrate that scenes extracted by brain data pertain to highlight scenes, we compared the highlight scenes determined from brain data with those marked as “TOP1”, “TOP1&TOP2”, and “TOP1&TOP2&TOP3”. The duration of each highlight scene is 3 s.

C. Results and Discussion

Fig. 8 shows the estimation accuracy for ten videos. When “TOP1&TOP2&TOP3” highlight scenes were compared with those determined from the acquired brain data, the estimation accuracy for 4 videos (3, 7, 8, and 10) was more than 80 %. Fig. 7 shows snapshots from video 7. Scene 6 in Fig 7, where the girl is breaking tiles, was estimated from the EEG data to be a highlight scene. This result is indicative of the possibility that can extract highlights from videos using EEG. Furthermore, the estimation accuracy for 2 videos (1 and 5) was more than 50 %. However, the estimation accuracy for other videos was less than 50 %. This is due to variations in highlight scenes selected by the subjects.

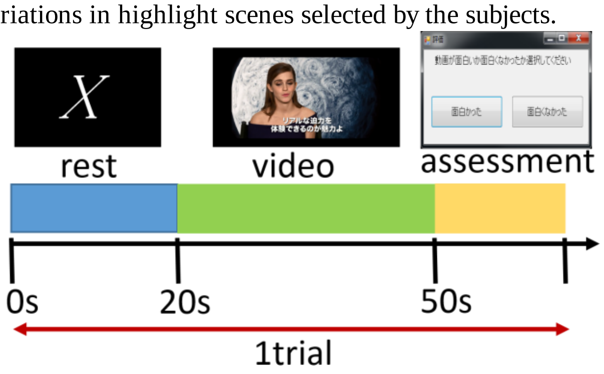

Fig.5: EX1 task flow

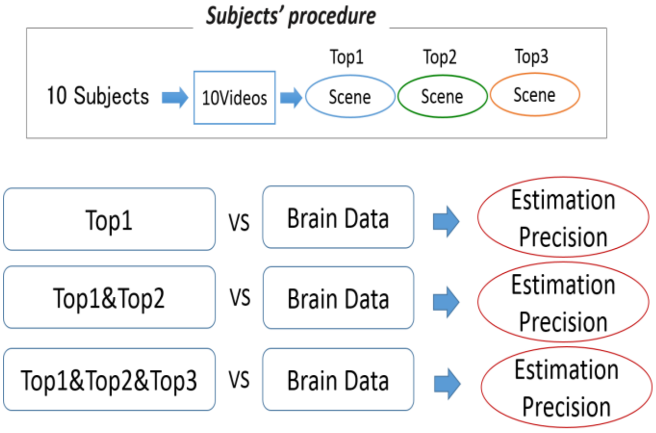

Fig. 6: Subjects' procedure

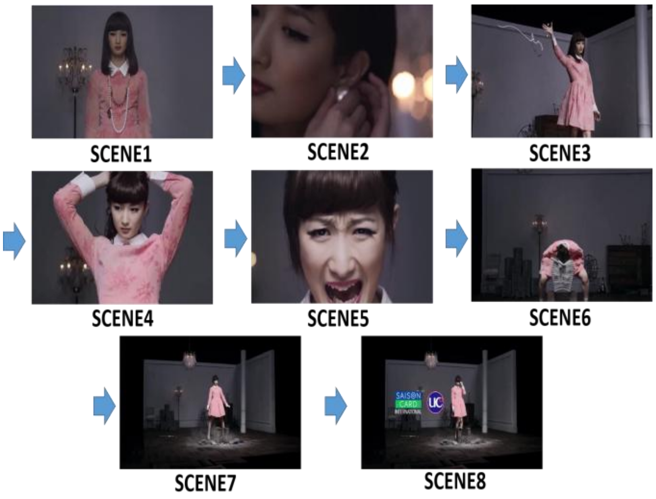

Fig. 7: Scene captures

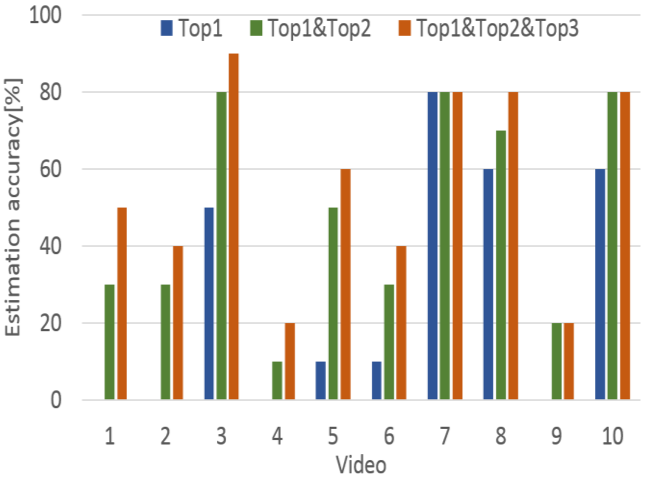

Fig. 8: Estimation accuracy

V. EXPERIMENT 2

Ten healthy Japanese participants took part in our research after providing written informed consent. All participants were university students aged between 21 and 24 years.

In EX2, the EEG data obtained in Experiment 1 was used. Therefore, we did not measure the EEG signals of our subjects in this experiment. Subjects evaluated the shots of all videos extracted from the EEG data of Experiment 1.

In EX2, shots extracted using the features indicated in Table 1 was given a score from one to 5. The 5-point scale ranges from "appropriate highlight" to "not an appropriate highlight". Shots extracted using the Volume Max value and shots extracted using EEG data are compared. We investigated the effectiveness of these features as a way to extract highlight shots.

A. *Task*

Fig. 9 shows the outline of the experiment. We define the steps marked from (1) to (3) as a single trial.

(1) Twenty seconds of rest.
(2) Thirty seconds of video-watching.
(3) Evaluation.

In the screen pertaining to step (3), several buttons are arranged. Each subject presses the desired button, and the shot extracted using the corresponding feature indicated in Table 1 is displayed. Subjects watch the extracted shots and score points. From each video, 1 shot is extracted per feature. Up to six buttons are displayed (since the number of features used is six). When the same shot is extracted for more than one feature, the number of buttons decreases. Subjects performed 50 trials. Subjects rested for ten minutes after 25 trials.

B. *Results and Discussion*

We examined the extracted shots that were given a score of either 1 or 2 and 4 or 5. We defined that shots scoring 1 and 2 are not appropriate highlights, and those that scored 4 or 5 are appropriate highlights. We counted the number of shots

DOI:10.17605/OSF.IO/VXWYT

that scored 1 or 2 and those that scored 4 or 5 for each feature. Then, we took the results from each subject and averaged them. Fig. 10 shows the results for shots that scored either 4 or 5, and Fig. 11 shows the results for shots that scored either 1 or 2.

Fig. 10 shows that the feature with the highest average value is Volume Max. The next highest valued feature is α-Min. There is no significant difference between Volume Max and α-Min. This result is indicative of the possibility that α-Min is a feature that can be used to extract highlights from videos.

Fig. 11 shows that the feature with the lowest average value is Volume Max. The next lowest valued feature is α-Min. There is no significant difference between Volume Max and α-Min. This result is indicative of the possibility that α-Min does not extract shots that have no relationship at all with the highlight scenes.

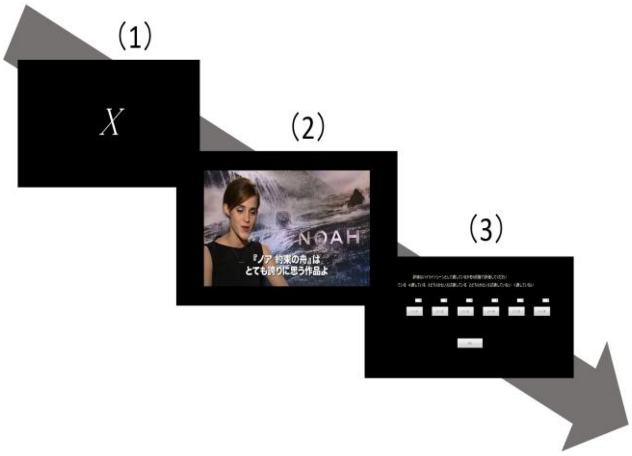

Fig. 9: EX2 and EX3 task flow

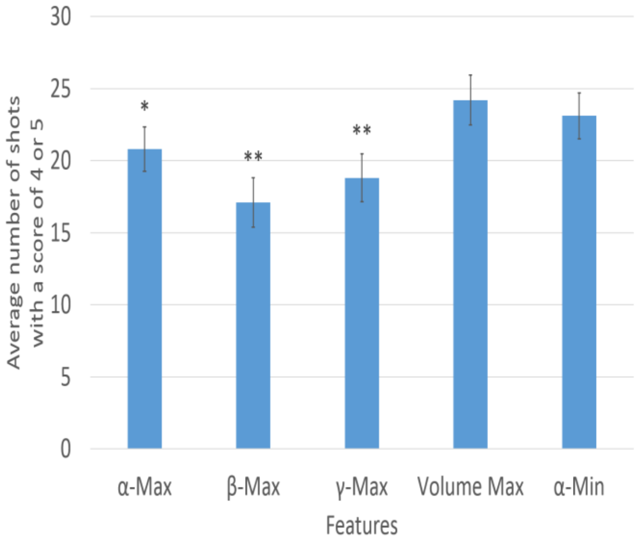

Fig. 10: Average number of shots that scored 4 or 5

In EX3, we used the system proposed in Section *IV* to investigate whether the same results are obtained for individual EEG signals. In addition, as Volume Max and α-Min are similar, we investigated whether the same shot is extracted for each.

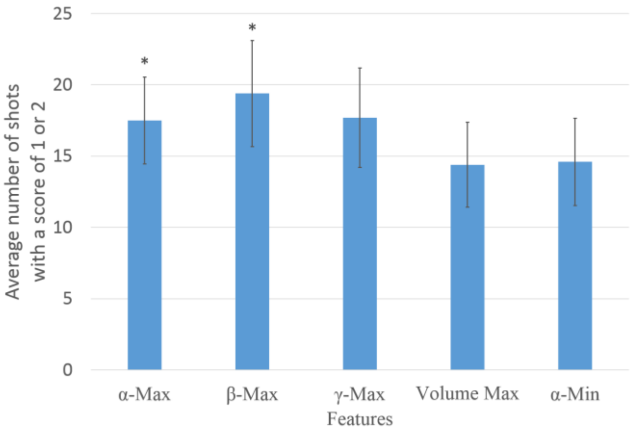

Fig. 11: Average number of shots that scored 1 or 2

VI. EXPERIMENT 3

Ten healthy Japanese participants took part in our research after providing written informed consent. All participants were university students aged between 21 and 24 years.

The system described in Section III was developed and used in EX3. We measured the subjects' brain waves during video watching. Shots were automatically extracted from the measured EEG data and subjects were asked to evaluate them on the spot.

In EX3, we also investigated whether the EEG signals were affected by volume. To investigate this, we used both videos with audio and videos without audio in our experiment. The contents of both videos were the same. Subjects were divided into 2 groups of 5 people (group A and group B). Group A performed experiments using the first half of the total number of videos with audio and group B performed experiments using the same videos with no audio. One month later, group A performed experiments using the second half of the total number of videos with no audio, and group B performed experiments using the same videos with audio.

A. Task

The task for EX3 is the same as for EX2. We excluded 2 videos with extremely few shots. Therefore, subjects performed 48 trials.

B. Results and Discussion

As in EX 2, the results of EX3 also examined only shots that scored 1 or 2, and 4 or 5. Fig. 12 shows the average number of shots that scored either 4 or 5 and correspond to the results obtained for the first half of the videos watched by group A and the second half of the videos watched by group B (i.e., videos with audio). Fig. 13 shows the average number of shots that scored either 4 or 5 and correspond to the results obtained for the second half of the videos watched by group A and the first half of the videos watched by group B (i.e., videos with no audio). '*' indicates the t-test results between Volume Max and other features. Fig. 12 shows that the feature with the highest average value is Volume Max. The next highest valued feature is α-Min. There is no significant difference between Volume Max and α-Min. Fig. 13 shows that the feature with the highest average value is α-Min. The next highest valued feature is Volume Max.

Fig. 14 shows the average number of shots that scored either 1 or 2 and correspond to the results obtained for the first half of the videos watched by group A and the second half of the videos watched by group B (i.e., videos with audio). Fig. 14 shows the average number of shots that scored either 5 or 2 and correspond to the results obtained for the second half of the videos watched by group A and the first half of videos watched by group A (i.e., videos with no audio). ‘*’ indicates the t-test results between Random and other features. Fig. 14 shows that the feature with the lowest average value is Volume Max. The next lowest valued feature is α-Min. There is no significant difference between Random and γ-Max, Volume Max, or α-Min. Fig. 15 shows that the feature with the lowest average value is α-Min. There was a significant difference between Random and other features, except Volume Max. This result strengthens the possibility that α-Min does not extract shots that have no relationship at all with the highlight scenes.

Fig. 16 shows the result of examining whether the shot extracted using Volume Max and the shot extracted using α-Min were the same. Volume Max and α-Min, as shown in Fig. 15, extracted the same number of shots. In Fig. 16, the label Volume Max on the horizontal axis indicates the number of extracted shots that were not extracted by α-Min, and the label α-Min is the number of extracted shots that were not extracted by Volume Max. The results show that Volume Max and α-Min are extracting different shots. These results are indicative of a possibility that brain waves could be a novel feature based on emotions that can be used to extract highlight scenes. Previous studies have reported that α-waves weaken when subjects watch their favorite CMs [18]. The results of this study are similar to those of previous studies.

In this research, because the video durations were short, extraction was done on a per shot basis. In our research, we were able to extract shot units, so the extraction method might be applied to scenes composed of shots.

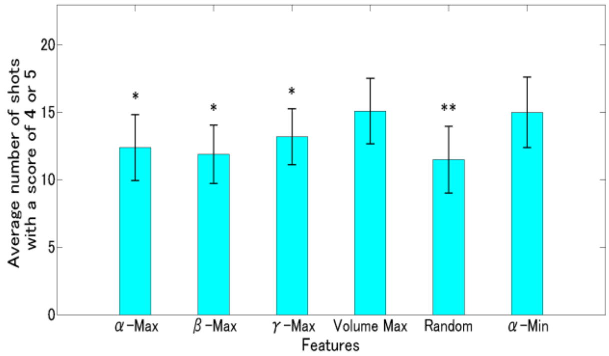

Fig. 12: Average number of 4 and 5 score for video with audio

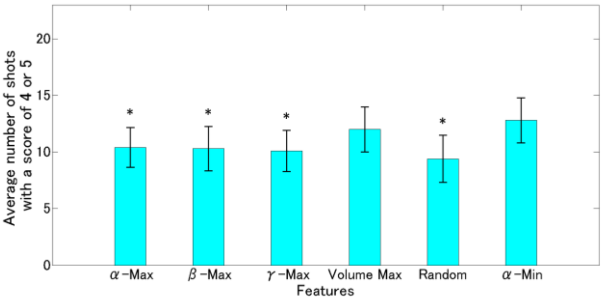

Fig. 13: Average number of 4 and 5 score for video with no audio

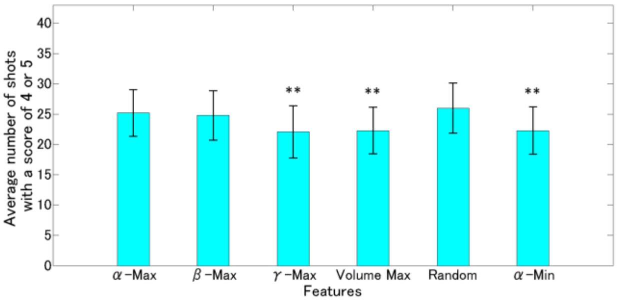

Fig. 14: Average number of 1 and 2 score for video with audio

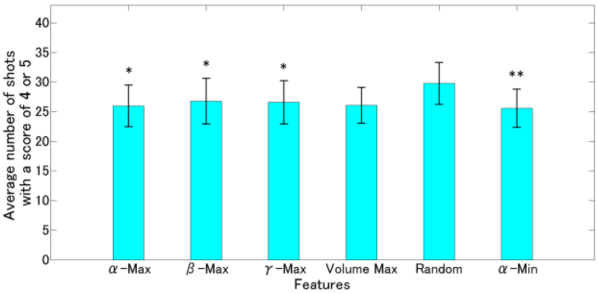

Fig. 15: Average number of 1 and 2 score for video with no audio

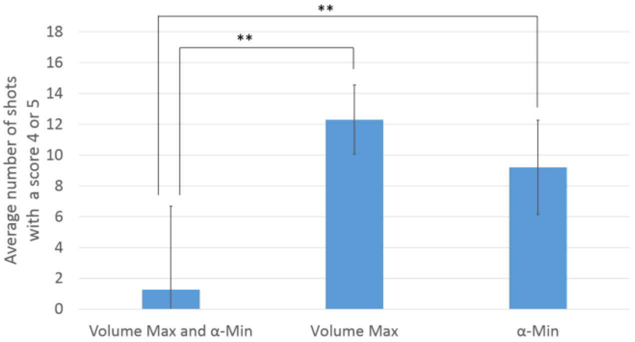

Fig. 16: Average number of same shots and independent shots

In the future, it is necessary to experiment with a video labeled with emotions and verify the effectiveness of brain wave features. In the study of emotion estimation, the EEG features vary depending on the subject’s emotions [19], [20]. The accuracy of the method for extracting highlight scenes could be improved by using brain wave features matching the content (emotions) of the video.

VII. CONCLUSION

In EX1, EEG signals from twenty subjects were measured while they were watching videos; the scenes extracted using the obtained EEG data were compared with the scenes selected by the subjects, and the estimation accuracy was calculated. When using scenes classified as “TOP1 & TOP2 & TOP3”, the estimation accuracy for four videos exceeded 80%. This suggests that it may be possible to extract highlight scenes using EEG. In EX2, evaluation experiments were performed by assigning scores to shots extracted using different features. In EX3, we used a system that extracted shots in real time using EEG. This system was developed by us. From the results of EX2 and EX3, we showed that α-Min is an effective feature to extract highlight scenes. In future, we will investigate brain features and their relationship to emotion by using videos labeled with emotions.

REFERENCES

[1] Salvaris, Mathew and Sepulveda, Francisco, “Visual modifications on the P300 speller BCI paradigm”, Journal of neural engineering, vol. 6, 2009.

[2] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu, “Emotional state classification from EEG data using machine learning approach”, Elsevier Journal Neurocomputing, vol. 129, 2014, pp. 94-106.

[3] Murugappan Murugappan, Nagarajan Ramachandran, Yaacob Sazali, et al., “Classification of human emotion from EEG using discrete wavelet transform”, Journal of Biomedical Science and Engineering, vol. 3, 2010 p. 390-394.

[4] Keng-Sheng Lin, Ann Lee, Yi-Hsuan Yang, Cheng-Te Lee, and Homer H Chen, “Automatic highlights extraction for drama video using music emotion and human face features”, Elsevier Journal Neurocomputing, vol. 119, 2103, pp. 111-117.

[5] Maryam Mustafa, Stefan Guthe, and Marcus Magnor, “Single-trial EEG classification of artifacts in videos”, ACM Trans on Applied Perception (TAP), vol. 9, 2012.

[6] Mohammad Soleymani, Maja Pantic, and Thierry Pun, “Multimodal emotion recognition in response to videos”, IEEE transaction on Affective Computing, vol. 3, 2012, pp. 211-223.

[7] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yaz-dani, Touradj Ebrahimi et al., “Deap: Adatabase for emotion analysis; using physiological signals”, IEEE transaction. on Affective Computing, vol. 3, 2012,pp. 18-31.

[8] Eleni Kroupi, Philippe Hanhart, Jong-Seok Lee, Martin Rerabek, and Touradj Ebrahimi, “Eeg correlates during video quality perception”, In Signal Processing Conference (EU-SIPCO), Proceedings of the 22nd European,2014, pp. 2135-2139.

[9] Stefano Valenzi, Tanvir Islam, Peter Jurica, and Andrezj Cichocki, “Individual classification of emotions using EEG”, Journal of Biomedical Science and Engineering, vol. 2014, 2014.

[10] Shangfei Wang, Yachen Zhu, Guobing Wu, and Qiang Ji, “Hybrid video emotional tagging using users EEG and video content”, Multimedia tools and applications, vol. 72, 2014, pp. 1257-1283.

[11] Taeyang Yang, Do-Young Lee, Youngshin Kwak, Jinsook Choi, Chajoong Kim, and Sung-Phil Kim, “Evaluation of tv commercials using neurophysiological responses”, Journal of physiological anthropology, vol. 34, 2015, p. 19.

[12] Peng Chang, Mei Han, and Yihong Gong, “Extract highlights from baseball game video with hidden markov models”, International Conference on. In Image Processing. Proceedings, vol. 1, 2002.

[13] Mei-Chen Yeh, Yen-Wei Tsai, and Hao-Chen Hsu, “A content-based approach for detecting highlights in action movies”, Multimedia Systems, vol. 22, 2016, pp. 287-295.

[14] Yasuo Ariki, Masahito Kumano, and Kiyoshi Tsukada, “Highlight scene extraction in real time from baseball live video”, In Proceedings of the 5th ACM SIGMM international, workshop on Multimedia information retrieval, 2003,pp. 209-214.

[15] Arthur G Money and Harry Agius, “Analysing user physiological responses for affective video summarization”, Displays, vol. 30, 2009, pp. 59-70.

[16] Christophe Chenes, Guillaume Chanel, Mohammad Soleymani, and Thierry Pun, “Highlight detection in movie scenes through inter-users”, physiological linkage. In Social Media Retrieval, Springer, 2013, pp. 217-237.

[17] Michael E Smith and Alan Gevins, “Attention and brain activity while watching television: Components of viewer engagement”, Media Psychology, vol. 6, 2004, pp. 285-305.

[18] Giovanni Vecchiato, Jlenia Toppi, Laura Astol Fabrizio De Vico Fallani, Febo Cincotti, Donatella Mattia, Francesco Bez, and Fabio Babiloni, “Spectral eeg frontal asymmetries correlate with the experienced pleasantness of tv commercial advertisements”, Medical & biological engineering & computing, vol. 49, 2011,pp. 579-583.