http://dx.doi.org/10.14236/ewic/HCI2022.23

A Model for Estimating Subjective Evaluation Values of Video Degradation from Viewers' Physiological Signals

Masaki Omata
University of Yamanashi
Kofu, Yamanashi, Japan
omata@hci.media.yamanashi.ac.jp

Naho Kiriyama
University of Yamanashi
Kofu, Yamanashi, Japan
kiri21@hci.media.yamanashi.ac.jp

This paper describes that it is possible to estimate a viewer's five-level subjective evaluation of video degradation with an estimation accuracy of 0.90 or better by using physiological data such as blood volume pulses of viewers. To this end, we conducted an experiment to record participants' EEG, BVP, gaze, pupil diameter, and subjective evaluation values of video degradation while they watched videos. We then created five different datasets from the data and built estimation models using machine learning based on random forest or neural network. As a result, the coefficient of determination for the physiological data with top importance trained by random forest was 0.997. The results contribute to an objective, continuous, unconscious, and quantitative method for estimating Quality of Experience (QoE) during video viewing.

Physiological signals. Quality of experience. Video degradation.

1. INTRODUCTION

This paper experimentally validates whether it is possible to estimate a viewer's response to degradation of video resolution or frame rate from the viewer's physiological signals during video viewing. As an experimental task, participants watched videos that we intentionally degraded partially, while continuously inputting their own subjective evaluation values regarding the degradation. During this time, EEG (ElectroEncephaloGraphy), Blood Volume Pulse (BVP), on-screen eye position, and pupil diameter were recorded as the physiological signals. We used 80% of the physiological data, the genre of the video, and the degradation information as training data for machine learning to construct models for estimating subjective evaluation values of degraded videos, and used the remaining 20% as test data to evaluate the models.

The contributions of this paper are as follows:

(i) In the construction of a model for estimating the subjective evaluation of video degradation, the model with physiological data is far more accurate than the model without physiological data.
(ii) The estimation accuracy of the model, which learns physiological data of high importance using the random forest method to estimate five levels of subjective ratings, was 0.997.
(iii) BVP is strongly correlated with the subjective evaluation value among EEG, BVP, eye gaze, and pupil diameter.

The background of this study is a problem that conventional network control based on Quality of Experience (QoE) does not directly reflect QoE of users (Skorin-Kapov et al., 2018). QoE has been defined that the degree of delight or annoyance of the user of an application or service (ITU, 2019). The current network control philosophy has shifted from device-centric to application-centric (Cao et al., 2015; Huang, 2015; Seddiki et al., 2015). Methods to control the network based on the user's network usage have also been proposed (Kolb et al., 2015; Bentaleb et al., 2016). However, the methods do not solve the problem that they do not directly reflect the quality of the user's own experience. Most conventional methods estimate QoE from network quality measures (e.g., delay) and application quality measures (e.g., sound quality, image quality, response time) and reflect them in network control (Liotou et al., 2015; Hayashi, 2015).

As a solution to the problem, we have proposed "affective network control system," in which physiological signals that can be measured objectively, continuously, and unconsciously are

Proceedings of the 35th British HCI and Doctoral Consortium 2022, UK

used to estimate physiological psychological experience quality, and networks are controlled in synchronization with the estimated values. As an experiment to obtain basic findings for the purpose, we attempted to build a machine learning model to estimate subjective evaluation values for video degradation based on viewers' physiological responses to video degradation during video viewing. In this paper, Section 2 introduces related research and the differences between them and our proposal, Section 3 describes the experiment of recording physiological signals, Section 4 describes the construction of the estimation model using machine learning, and Section 5 concludes.

2. RELATED WORK

Various methods have been studied to obtain or estimate QoE. These can be broadly classified into three methods: estimating QoE from quality of service (QoS) such as a state of network traffic and computer resources; obtaining QoE from user questionnaires such as oral examinations and questions; and estimating QoE from user's physiological signals such as EEG, gazing, and facial expressions. Each method has its advantages and disadvantages. Estimating from QoS has the advantage of being easy to estimate continuously and quantitatively, but has the disadvantage of not directly reflecting the user's condition. Obtaining data from questionnaires has the advantage of directly obtaining the user's condition and mental state, but has the disadvantages of being difficult to obtain continuously, interfering with the user's work, and causing bias due to subjectivity. Estimating from physiological responses has the advantage of continuously and objectively obtaining the user's responses, but has the disadvantages of a large apparatus, instability of estimation accuracy, and individual differences.

In studies using QoS, Hayashi proposed using network quality measures such as packet loss and application quality measures such as audio and picture quality to quantify QoE (Hayashi, 2015). Liotou et al. listed two major QoE influence factors: service-independent factors such as network layer and physical layer, and service-dependent factors such as video specific and voice, and introduced a QoE estimation formulas based on the factors (Liotou et al., 2016).

In a study using a questionnaire survey, Robitza et al. asked participants to watch videos of different quality and then to answer questions about their impressions of the video viewing experience and their behavior when the quality was poor (Robitza et al., 2016). The International Telecommunication Union (ITU) has developed recommendations on methodologies for subjective quality evaluation of video images, specifying viewing environments and viewing conditions with the aim of obtaining reproducible quality evaluation results. In particular, BT.500 recommends a method for subjective evaluation of TV video quality (ITU, 2019).

Regarding use of physiological signals, Arndt et al. reviewed previous studies that have primarily used EEG. P300 and alpha and theta waves of video and audiovisual viewers have been used for the EEG analyses (Arndt et al., 2016). Porcu et al. surveyed previous studies using EEG, facial expressions, and eye gaze, noting that it is difficult to investigate visual local interest and load when EEG is the primary method, so they used facial expressions and eye gaze (Porcu et al., 2020).

The novelty of our approach, which differs from the previous studies, is to combine video quality parameters which are QoS, with physiological data which are direct reactions of viewers, in order to estimate subjective evaluation values regarding video degradation, which is one of QoE. The utility of our approach is to be able to quantify QoE directly, objectively, continuously, and unconsciously, and to be able to flexibly control network and/or computer resources in synchronization with the QoE.

3. EXPERIMENT

The purpose of this experiment was to obtain the viewer's physiological signals and subjective evaluation values during video degradation in order to construct estimation models. Experimental participants were instructed to use a slider to input their subjective evaluation of the video quality while watching a video. At the same time, participants' EEG, fingertip BVP, and gaze were recorded. Four videos of different genres with playback durations ranging from 5 to 7 minutes were prepared as experimental stimulus videos. Each video contained scenes that were intentionally degraded by the experimenter in advance.

3.1 Experiment environment

The experiment was conducted in a private room with no ambient noise and no outsiders. The brightness of the room was adjusted to a constant level by closing the curtains and turning on a light, and the room temperature was adjusted to 22°C with an air conditioner in order to ensure that the experimental environment was the same for all participants.

3.2 Experimental apparatus and physiological signals

Figure 1 shows the experimental apparatus and a participant performing the experimental task. Each of the equipments is described below.

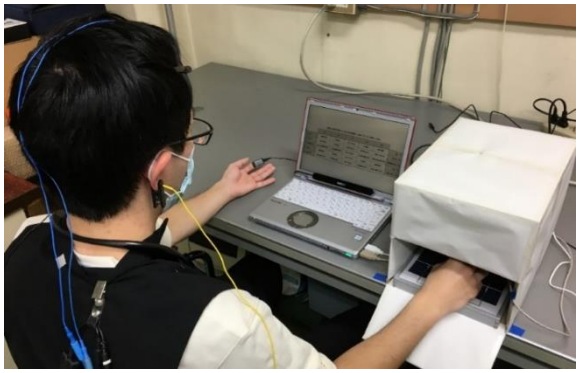


Figure 1: *The experimental equipment and a participant performing the task*

3.2.1. Video playback equipment
A laptop computer (Panasonic CF-SZ6) with a screen size of 12.1 inches was used to play the stimulus videos. Each video was fixed to 1280 x 720 px and displayed in the center of the laptop. A neck-mounted loudspeaker (JVCKENWOOD SP-A7WT-B) was used to play the sound of the videos.

3.2.2. Subjective evaluation value input device
To obtain subjective evaluation values regarding video quality, the Single Stimulus Continuous Quality Evaluation (SSCQE) method (ITU, 2019) was implemented, in which a participant inputs the evaluation values by moving a slider (Phidgets 1112-Slider 60, see Figure 2) while watching a video. The maximum movement range of the slider is 60 mm, and the sampling frequency is 10 Hz. The movement range of the slider was divided into five segments (12 mm each), and one of the five evaluation values shown in Table 1 was continuously input as a time-continuous value. The five evaluation terms are based on ITU-R BT.500 (ITU, 2019).

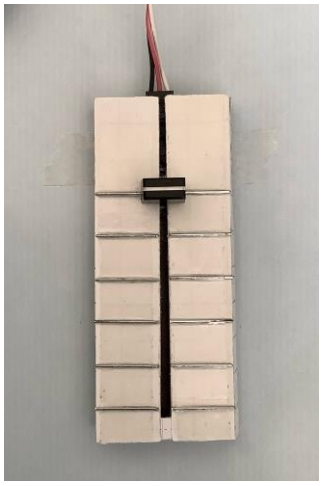


Figure 2: *The slider for inputting subjective evaluation value*

Table 1: *Subjective evaluation value for video degradation*

Evaluation value	Evaluative phrase
5	*Deterioration is imperceptible*
4	*Deterioration is perceptible, but not annoying*
3	*Deterioration is slightly annoying*
2	*Deterioration is annoying*
1	*Deterioration is very annoying*

Because the ITU-R BT.500 describes a “fixed or desk-mounted position,” and to prevent participants from looking at their hands while concentrating on watching a video, the slider was hidden by a box and fixed to the experimental table, as the right hand of the participant shown in Figure 1.

3.2.3. Physiological Sensors
In this experiment, the viewer’s electroencephalograph (EEG), Blood Volume Pulse (BVP), on-screen eye position, and pupil diameter were measured and recorded. we checked the wear of the physiological sensors after each participant finished watching each video in order to obtain accurate data.

3.2.3.1. EEG sensor
EEG is a record of the oscillations of brain electric potentials recorded from electrodes attached to the human scalp (Nunez and Srinivasan, 2007). The theta (4-7 Hz), alpha (8-12 Hz), and beta (13-21 Hz) waves were extracted from the raw data, and the power and peak-to-peak values for each of the bands were calculated. It has been reported that alpha waves increase during relaxation, while alpha waves decrease and beta waves and beta/alpha waves increase when the mental load is high or when feeling discomfort (Uwano et al., 2008). In addition, Awang et al. reported a positive correlation between stress and theta waves (Awang, 2011).

The EEG was measured at Fp1 and Fp2 of the frontal cortex as defined by the international 10-20 method. The reason for measuring the EEG of the prefrontal cortex is based on previous studies that have shown that this is the region where evaluative emotions are most likely to be expressed (Mitsukura, 2019). The EEG-Z sensor (Thought Technology Ltd., 2022) from Thought Technology was used as the sensor.

3.2.3.2. BVP sensor
BVP bounces infra-red light against a skin surface and measures the amount of reflected light. The amount varies with the amount of blood in the skin. From the raw data, the Inter-Beat Interval (IBI), which is the interval between R waves that occurs when blood is pumped from the left ventricle to the aorta (Miyata, 1998; Hori, 2008), and the Normal-to-

Normal Interval (NN), which is the IBI without artifacts (Citi et al., 2012), were calculated, and the low frequency (LF) component (0.04-0.15 Hz) and the high-frequency (HF) component (0.15-0.45 Hz) were extracted (Hori, 2008). The HF component appears when the parasympathetic nervous system is dominant, while the LF component appears regardless of whether the sympathetic or parasympathetic nervous system is dominant.

The BVP was measured on ball of index finger of a participant's left hand. Participants were asked to place their left hands palm up on a desk and to move them as little as possible. A BVP-Flex/Pro sensor (Thought Technology Ltd., 2022) from Thought Technology was used to measure BVP.

3.2.3.3. Eye tracker

The eye tracker was used to measure the participants' viewing position on the screen (X-Y coordinate) and their pupil diameters. Sawahata et al. investigated the relationship between viewers' comprehension of a TV program and the direction of their gazes and reported that the variances in gaze direction tended to be lower for scenes for which the participants had better comprehension (Sawahata et al., 2008). Pupil diameter responds to cognitive processing, arousal, and increased interest (Hess and Polt, 1960; de Winter et al., 2021). Typically, the greater the level of arousal or interest, the larger the pupil size. A Euclidean distance of an eye movement between a frame and the previous frame was calculated from the eye coordinate values of the frames.

The eye tracker was placed on the hinge of the laptop computer for video playback, facing the participant's face. Tobii's Tobii pro nano (Tobii AB, 2022) was used as the eye tracker. The sampling rate is 60 Hz, the resolution of the eye position is 1280 x 800 px, and the unit of measurement for pupil diameter is 0.1 mm.

3.3 Stimulus Video

Four videos of different genres, ranging from 5 to 7 minutes in duration, were prepared as stimulus videos to be presented to the participants. The genres are horror, nature, animation, and education. The original resolution of all videos is 1280 x 720.

The independent variables of video quality in this experiment are resolution and frame rate. As a combination of the values of the two variables, we set the conditions for the nine levels of degradation shown in numbers 1 through 9 in Table 2. The condition with no degradation is shown as '0' in Table 2, with a resolution of 1280 x 720 and a frame rate of 30 fps. Video quality other than resolution and frame rate is as follows: video codec is H.264, bit rate is 1800 kbps, and video decompression format is YUY2. The audio quality was not changed as follows: AAC audio codec, 192 kbps audio bit rate, 44.1 KHz sampling rate, and PCM audio format.

Table 2: *Video degradation conditions*

Degradation condition	Resolution [px]	Frame rate [fps]
0 (no degradation)	*1280 × 720*	*30*
1	*1280 × 720*	*25*
2	*854 × 480*	*30*
3	*1280 × 720*	*20*
4	*640 × 360*	*30*
5	*1280 × 720*	*15*
6	*426 × 240*	*30*
7	*1280 × 720*	*10*
8	*320 × 180*	*30*
9	*1280 × 720*	*5*

As shown in Figure 2, each video contains 15 seconds of each of the degradation conditions 1 through 9, with no-degradation time periods before and after the degradation conditions. The 15-second video sequences were generated by re-encoding the degraded conditions at a lower resolution and frame rate based on Table 2.

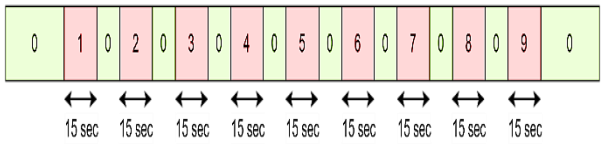


Figure 2: *The order and time of the degradation conditions in a video*

3.4 Procedure and task

One by one, participants came to the laboratory to perform the experimental task. At the beginning, the experimenter explained the experiment to the participant and obtained informed consent from the participant. Next, after the participant was seated facing the laptop computer for video playback, the experimenter attached the physiological sensors and the neck-mounted loudspeaker to the participant and calibrated the eye tracker. For the participant's baseline, the experimenter recorded his/her EEG, BVP, and pupil diameter during 30 seconds of resting with eyes closed, followed by 1 minute of resting while looking at a video-playing laptop with no display.

The experimental task assigned to the participants was to watch the videos to evaluate the quality of the videos using a slider if they noticed any differences in the quality of the videos. The participants were instructed to keep their hands and heads as still as possible, and to input the evaluation values without looking at the slider while looking at the screen.

Before the start of the experimental task, the participant was familiarized with the slider operation by watching a 3-minute video for task practice, which was different from the stimulus videos.

As a within-subjects design, all participants watched all four videos and rated the degree of degradation of the videos several times at their own timing. The order in which the participants watched the videos was the same: horror, nature, animation, and education. After each video, the participant completed a questionnaire about his or her impression of the content of the video. The questionnaire items were five adjectives (actually written in Japanese) on a five-point scale (from A to E) for five items describing impressions of the content of the video shown in Table 3. The table was displayed on the screen immediately after the video ended, and the participant respond verbally.

Table 3: *The answers regarding impressions of a video*

A	B	C	D	E
Very like	*Slightly like*	*Neither*	*Slightly dislike*	*Very dislike*
Very interesting	*Slightly interesting*	*Neither*	*Slightly boring*	*Very boring*
Very tense	*Slightly tense*	*Neither*	*Slightly calm*	*Very calm*
Very pleasant	*Slightly pleasant*	*Neither*	*Slightly afraid*	*Very scared*
Very dynamic	*Slightly dynamic*	*Neither*	*Slightly static*	*Very static*

Participants were 11 college students (7 males and 4 females) between the ages of 20 and 24. After watching all the videos, each participant completed a questionnaire survey about his/her daily video viewing.

4. CONSTRUCTION OF AN ESTIMATION MODEL USING MACHINE LEARNING

We construct models to estimate a subjective evaluation value of video degradation based on the physiological signals, degradation conditions, video genre, and video impressions recorded in the experiment in Section 3. In this study, 33 types of data (hereinafter referred to as “physiological data”), including recorded raw data and data calculated from the raw data, were prepared as explanatory variables, and subjective evaluation values of video degradation were used as objective variables. We prepared five data sets from which we selected types of the explanatory variables to be used, and compared the differences among the datasets.

We built regression models by using Random Forests (RF) and Neural Networks (NN) for supervised machine learnings using subjective evaluation values as the correct data. All recorded data were resampled at 16 Hz. Eighty percent of the data, randomly selected from all the data, was used as training data, and the remaining twenty percent of the data was used as model evaluation data.

4.1 Dataset

Table 4 shows five datasets used for training to build a model and for evaluating the model. The combination of data types in each dataset is based on the test of the difference in estimation accuracy depending on the presence or absence of physiological signals and on the test of whether estimation is possible with a small number of physiological signal types. There are a total of 29 types of raw physiological signal data and data calculated from the raw data. The “top 15 most important physiological data” refers to the 15 types of the most important physiological data among the 29 types calculated by the random forest as described in section 4.3.

Table 4: *Types of data in datasets for machine learning*

No. of dataset	Overview of data types
1	*Degradation condition, video genre, video impression*
2	*Degradation condition, video genre, All types of physiological data*
3	*Degradation condition, video genre, All types of physiological data, video impression*
4	*Degradation condition, video genre, top 15 most important physiological data*
5	*Degradation condition, video genre, top 15 most important physiological data, video impression*

4.2 Pre-processing

4.2.1. Coding of values related to video

To code the degradation conditions, the condition numbers from 1 to 9 shown in Figure 2 were set to one-tenth (i.e., 0.1, 0.2, ..., 0.9). On the other hand, for the zeros representing the no-degradation condition, the zeros were assigned between the degradation conditions to distinguish the respective zeros before and after each degradation condition. For example, degradation condition 2 was coded 0.2, degradation condition 3 was coded 0.3, and 0 between these conditions was coded 0.25.

The four video genres were assigned the values 0.25, 0.50, 0.75, and 1.00. In addition, responses to each of the five types of five-level adjectives (Table 3) regarding impressions of the video content were assigned a score of 0.1, 0.2, 0.3, 0.4, and 0.5.

For the objective variable, the subjective rating value of video degradation, participants rated the video on a 5-point scale (Table 1), but since the slider itself has a resolution of 1000 steps, we attempted to subdivide the responses into a higher resolution of 21 steps,

with values ranging from 0.00 to 5.00 in increments of 0.25. Figure 3 shows the distribution of the subdivided subjective evaluation values of all participants for each degradation condition. The plots outside the box-and-whisker diagrams are outliers.

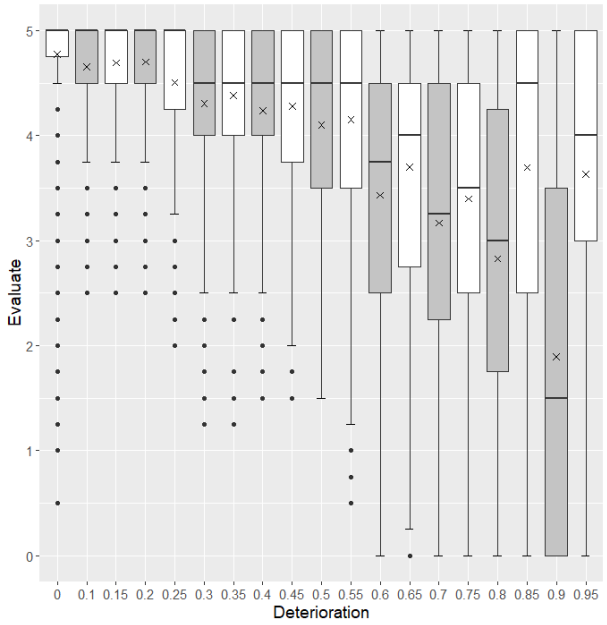


Figure 3: *Box-and-whisker diagram of subjective evaluation values for the subdivision of the degradation conditions*

4.2.2. Coding of values for physiological signals

The percentage of the total LF power values and the percentage of the total HF power values of the BVP were recorded as values between 0 and 100, so they were linearly transformed and standardized to fit between -1.0 and 1.0 after dividing the values by 100.

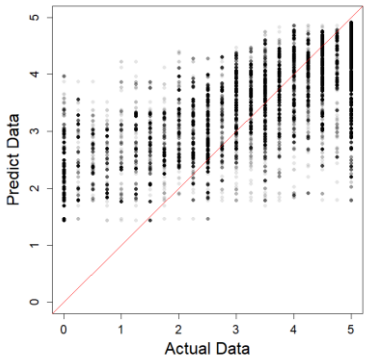
On the other hand, values calculated from physiological signals and physiological signals other than the total LF/HF power value were checked for normality by the Anderson-Darling test (Gross and Ligges, 2015). Since normality could not be confirmed ($p<0.05$), the data were standardized by robust Z-score.

4.3 Random Forest (RF)

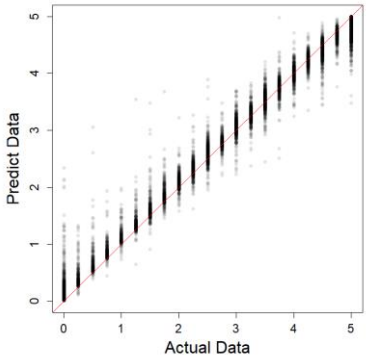
In random forests, it is possible to calculate the importance of explanatory variables using Mean Decrease Gini. In the model building, we used the top 15 physiological data types of importance based on Mean Decrease Gini for datasets 4 and 5 (Table 4), since the number of data for training is not large. The top 15 physiological data in order of importance were: NN interval of BVP, HF power ratio of BVP, LF power ratio of BVP, amplitude of BVP, IBI of BVP, power value of theta wave of EEG Fp1, power value of theta wave of EEG Fp2, pupil diameter of right eye, power value of beta wave of EEG Fp2, pupil diameter of left eye, power value of beta wave of EEG Fp1, power value of alpha wave of EEG Fp1, power value of alpha wave of EEG Fp2, raw data of EEG Fp2, and raw data of EEG Fp1.

The R randomForest package was used for the implementation. All parameters used the default values of the package. The number of features, *mtry*, was set to *n*/3 (*n* being the number of explanatory variable data types), and the number of decision trees to be created, *ntree*, was set to 500.

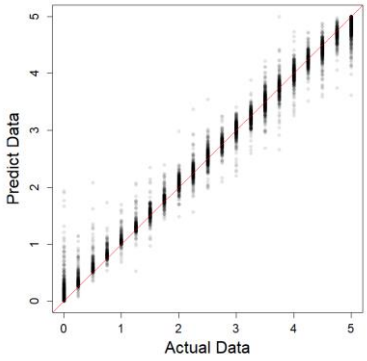
Figure 4 (a) to (e) show graphs comparing the estimated subjective evaluation values by the regression model constructed using the RF for each of the five datasets, with the actual measured subjective evaluation values from the experiment. We find that the plots in the datasets that use physiological data (datasets 2 to 5) converge on the diagonal and improve the accuracy of estimation compared to dataset 1, which does not use physiological data.



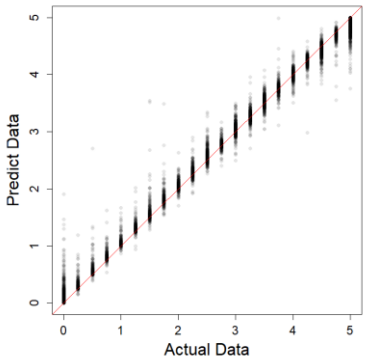
(a) Dataset 1



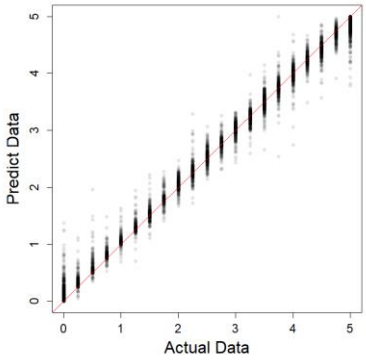
(b) Dataset 2



(c) Dataset 3

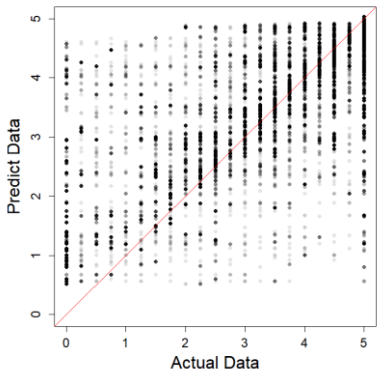


(d) Dataset 4

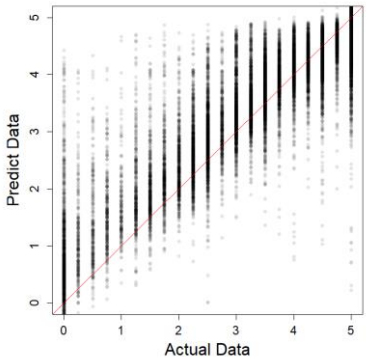


(e) Dataset 5

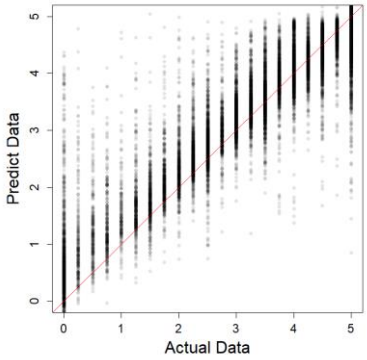
Figure 4: *Estimation results of regression model with random forest*



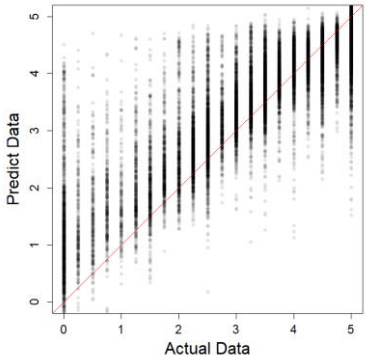
(a) Dataset 1



(b) Dataset 2



(c) Dataset 3

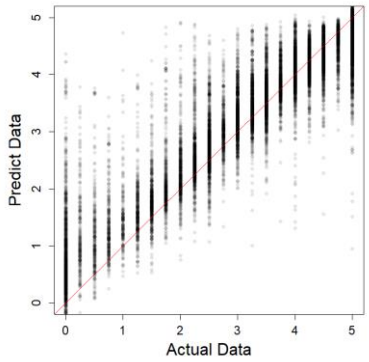


(d) Dataset 4

4.4 Neural Network (NN)

Sony Neural Network Console was used for implementation (Sony, 2022). The basic structure of the layers consists of three layers of Affines and two layers of ReLUs. Additionally, two layers of Dropouts were inserted to prevent over-learning. For the output layer, a HuberLoss layer was used to solve the regression problem. The data used were the same as for the aforementioned RF, including the top 15 most important physiological data.

Figure 5 (a) to (e) show graphs comparing the estimated subjective evaluation values by the regression model constructed using the NN for each of the five datasets with the actual measured subjective evaluation values from the experiment. We find that the plots in the datasets that use physiological data (datasets 2 to 5) converge on the diagonal and improve the accuracy of estimation compared to dataset 1, which does not use physiological data. Furthermore, we find that the convergence of the NN results to the diagonal is weak compared to the graph of the RF results shown in section 4.3.



(e) Dataset 5

Figure 5: *Estimation results of regression model with neural network*

4.5 Accuracy of the estimated model

Coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Mean Error (ME) were used as metrics for estimation accuracy of the estimation models. Table 5 shows the values of the metrics for each dataset in RF and NN. The highest R^2 of all is 0.997 for the RF method using dataset 5. The lowest RMSE is 0.073 for the RF method using dataset 5, and the lowest absolute value of ME is 0.00 for the RF method using dataset 4. For NN, the highest R^2 and RMSE are for dataset 3, and the lowest absolute ME is for dataset 5.

Table 5: *Estimation accuracy by learning method and dataset*

Training model	No. of Dataset	R^2	RMSE	ME
RF	*1*	0.597	0.858	0.004
	2	0.992	0.119	0.003
	3	0.996	0.084	0.003
	4	0.997	0.076	-0.000
	5	0.997	0.073	0.001
NN	*1*	0.652	0.797	-0.037
	2	0.853	0.519	-0.043
	3	0.906	0.414	-0.049
	4	0.825	0.566	-0.056
	5	0.905	0.417	-0.031

5. DISCUSSIONS

5.1 Machine learning results

Table 5 shows that datasets 2 through 5, which include physiological signals, have significantly higher estimation accuracy than dataset 1 which does not include any physiological signals, both in the RF and in the NN. The results validate our hypothesis that physiological signals such as EEG, BVP, and eye gaze are useful for estimating viewers' subjective evaluation of video degradation. The reason for this may be that the physiological responses of humans change in accordance with their stress or discomfort. We believe that the responses of the viewers differ according to the degree of degradation of the video. Therefore, as a future work, it is necessary to analyse the detailed relationships among the degree of video degradation, the degree of stress or discomfort, and the physiological responses.

Physiological data based on BVP ranked higher in importance than EEG and gaze in the calculation of Gini impurity for the RF. The result differs from our hypothesis that EEG is more important than other physiological signals because it is more directly related to visual processing. Although we have not yet elucidated the reason for the result, we hypothesize that the BVP responded to stress caused by the degradation, while the EEG was primarily related to the processing of the video content and did not respond well to the degradation.

Comparison of the dataset of the top 15 most important physiological data with the dataset of all physiological data shows that even the top 15 datasets have sufficient accuracy to obtain an R^2 greater than 0.90. This indicates that physiological data, mainly BVP, may be sufficient to estimate subjective evaluation values for the deterioration. Furthermore, the BVP sensor is simpler in structure and less expensive than EEG and gaze sensors, making it highly valuable both as an implementation and as a use of physiological psychological data for our proposed affective network control system.

5.2 Design of affective network control system

Based on our analyses and discussions described above, this section describes a design proposal for affective network control system, which we have proposed as the basic background for this paper. The Affective network control system is a switching system that estimates subjective responses of network users from their physiological signals and controls the network parameters based on the estimated results, such as the subjective evaluation values in this paper. We believe that this will not only keep user's QoE above a certain standard, but also enable effective use of network resources.

Specifically, for example, as in the experiment, if the system can automatically estimate each user's subjective evaluation of video degradation based on their physiological signals, it is possible to control bandwidth to extent that an evaluation of a user with a high evaluation does not become low, and to use the resulting resources as incremental resources for another user with a low evaluation. Or, even for a single user, for example, if the system can estimate that the user is dissatisfied with degradation of a video he/she is watching on his/her smartphone, the

system automatically switches the playback device to another device with higher bandwidth and resolution that is closer to the user.

6. CONCLUSION

We conclude that physiological signals are useful by experimentally validating that the use of physiological signals can estimate the subjective evaluation of video degradation by video viewers with a higher accuracy of a coefficient of determination of 0.997 than the case without physiological signals. Moreover, we experimentally derived the result that BVP is more important for this estimation among BVP, EEG, and eye gaze. We also derived that random forests can estimate subjective evaluation values with higher accuracy than simple neural networks. In our analysis, we showed that sufficient estimation accuracy can be obtained in random forests by using only the physiological data of high importance instead of all the physiological data.

As our future work, we plan to conduct a more detailed analyses of relationships among video quality, subjective evaluation values, and physiological data. Then, we plan to build a machine learning model with higher estimation accuracy using fewer types of data. Finally, based on the findings, we plan to implement our proposed affective network control system and validate its usefulness.

REFERENCES

Arndt, S., Brunnström, K., Cheng, E., Engelke, U., Möller, S., and Antons, J. (2016) Review on using physiology in quality of experience. IS&T International Symposium on Electronic Imaging 2016, Human Vision and Electronic Imaging 2016, HVEI-125.

Awang, S.A., Pandiyan, P.M., Yaacob, S., Ali, Y.M., Ramidi, F., Mat, F. (2011) Spectral Density Analysis: Theta Wave as Mental Stress Indicator. Signal Processing, Image Processing and Pattern Recognition (SIP 2011). Communications in Computer and Information Science, vol 260. 103–112. Springer, Berlin, Heidelberg.

Bentaleb, A., Begen, A. C. and Zimmermann, R. (2016) SDNDASH: Improving QoE of HTTP Adaptive Streaming Using Software Defined Networking. Proceedings of the 24th ACM international conference on Multimedia (MM '16). 1296–1305. Association for Computing Machinery, New York, NY, USA.

Cao, Z., Fitschen, J. and Papadimitriou, P. (2015) FreeSurf: Application-Centric Wireless Access with SDN. Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, 357–358. Association for Computing Machinery, New York, NY, USA.

Citi, L., Brown, E., and Barbieri, R. (2012) A Real-Time Automated Point-Process Method for the Detection and Correction of Erroneous and Ectopic Heartbeats. IEEE transactions on bio-medical engineering. vol. 59. 2828-2837.

de Winter, J. C. F., Petermeijer, S. M., Kooijman, L., and Dodou, D. (2021) Replicating five pupillometry studies of Eckhard Hess. International Journal of Psychophysiology, Vol.165, 145-205.

Gross, J., and Ligges, U. (2015) Tests for Normality. Version1.0-4, from https://cran.r-project.org/web/packages/nortest/nortest.pdf.

Hayashi, T. (2015) QoE-centric Operation for Optimizing User Quality of Experience. NTT Technical Review, Vol. 13 No. 9.

Hess, E. H., and Polt, J. M. (1960) Pupil Size as Related to Interest Value of Visual Stimuli. Science, Vol.132, No.3423, 349-350.

Hori, T. (2008) Physiological Psychology. Baifukan, Tokyo.

Huang, Z. (2015) The Research of SDN Group Policy Model Based on Application Centric Environment. 2015 Sixth International Conference on Intelligent Systems Design and Engineering Applications (ISDEA), 884-889.

International Telecommunication Union (ITU), (2019) Methodologies for the subjective assessment of the quality of television images, BT.500-14 (10/2019).

Kolb, J., Chaudhary, P., Schillinger, A., Chandra, A. and Weissman, J. (2015) Cloud-Based, User-Centric Mobile Application Optimization. IEEE International Conference on Cloud Engineering, 26-35.

Liotou, E., Tseliou, G., Samdanis, K., Tsolkas, D., Adelantado ,F., and Verikoukis, C.(2015) An SDN QoE-service for dynamically enhancing the performance of OTT applications. Seventh International Workshop on Quality of Multimedia Experience (QoMEX), 1-2.

Liotou, E., Tsolkas, D., and Passas, N. (2016) A roadmap on QoE metrics and models. 23rd International Conference on Telecommunications (ICT), 1-5.

Mitsukura, Y. (2019) KANSEI Detection and Its Application Using the EEG: Widespread of KANSEI Research in Society. IEICE ESS Fundamentals Review, Vol. 13, Issue 3, 180-186.

Miyata, Y. ed. (1998) New Physiological Psychology. 1, Kitaoji-shobo, Kyoto.

Nunez, P. L., and Srinivasan, R. (2007) Electroencephalogram. Scholarpedia, Vol. 2, No. 2:1348. Retrieved Feb. 15, 2022 from http://www.scholarpedia.org/article/Electroencephalogram.

Porcu, S., Floris, A., Voigt-Antons, J. -N., Atzori L., and Möller, S. (2020) Estimation of the Quality of Experience During Video Streaming From Facial Expression and Gaze Direction. IEEE Transactions on Network and Service Management, vol. 17, no. 4, 2702-2716.

Robitza, W., Kara, P. A., Martini, M. G., and Raake, A. (2016) On the Experimental Biases in User Behavior and QoE Assessment in the Lab. IEEE Globecom Workshops (GC Wkshps), 1-6.

Sawahata, Y., Komine, K., Hiruma, N., Itou, T., Watanabe, S., Suzuki., Y., Hara, Y., Issiki, N. (2008) Determining Relationship between Eye-Gaze Distribution and Comprehension of TV Programs. The Journal of The Institute of Image Information and Television Engineers, Vol. 62, Issue 4, 587-594.

Seddiki, M.S., Shahbaz, M., Donovan, S.P., Grover, S., Park, M., Feamster, N., and Song, Y. (2015). FlowQoS: Per-Flow Quality of Service for Broadband Access Networks. Georgia Institute of Technology, SCS Technical Report; GT-CS-15-02.

Skorin-Kapov, L., Varela, M., Hoßfeld, T., and Chen, K. (2018) A Survey of Emerging Concepts and Challenges for QoE Management of Multimedia Services. ACM Trans. Multimedia Comput. Commun. Appl. 14, 2s, Article 29.

Sony Network Communications Inc. (Retrieved Feb. 15, 2022) Neural Network Console. from https://dl.sony.com/.

Thought Technology Ltd. (Retrieved Feb. 15, 2022) EEG-Z Sensor - T9305Z. from https://thoughttechnology.com/eeg-z-sensor-t9305z/.

Thought Technology Ltd. (Retrieved Feb. 15, 2022) Blood Volume Pulse (BVP) Sensor - SA9308M. from https://thoughttechnology.com/blood-volume-pulse-bvp-sensor-sa9308m/.

Tobii AB (Retrieved Feb. 15, 2022) Tobii Pro Nano. from https://www.tobiipro.com/product-listing/nano/.

Uwano, H., Ishida K., Matsuda, Y., Fukushima, S., Nakamichi, N., Ohira, M., Matsumoto K., Okada, Y. (2008) Evaluation of Software Usability Using Electroencephalogram: Comparison of Frequency Component between Different Software Versions. Journal of Human Interface Society, human interface Vol.10, No.2, 233-242.