Multimodal Stress Estimation Using Multibiological Information: Towards More Accurate and Detailed Stress Estimation

Takumi Nagasawa†, Ryo Takahashi†, Keiko Ogawa-Ochiai†† and Norimichi Tsumura† (member)

Abstract In this paper, we propose a method for estimating stress levels using multibiological information obtained from red-green-blue (RGB) facial videos. The most popular method for estimating stress is by questionnaire, which is not highly objective or accurate. To estimate stress levels more accurately, it is desirable to establish the method which can estimate with objective and accurate information. In recent years, many methods for using biological information have been proposed. In this study, we used three types of biological information to estimate stress: pulse rate variability, blink rate, and pupil size, which were captured using an RGB camera. Results showed that the use of these biological signals improved the accuracy of stress estimation. Moreover, using biological information in a multimodal manner enabled the most accurate and detailed stress estimation.

Keywords: stress estimation, RGB camera, remote measurement, pulse wave, eye blink, pupil diameter.

1. Introduction

Modern society has made a great leap forward in economic and technological development; people are more affluent now than ever before. According to Slavich [1], social stress is a major variable in the complex and multifactorial etiology of psychiatric disorders. Work-related stress as well as stress that arises from human relations, such as loneliness, romantic breakups, or marital conflicts, are associated with an increased risk of psychopathology. If stress in daily life is left untreated, it can lead to psychological reactions, such as anxiety, tension, and insomnia; physical reactions such as headache, abdominal pain, and fatigue; and accumulation of stress, which may result in serious diseases, such as depression, autonomic dysregulation, stroke, and heart disease. To prevent the mental and physical problems caused by stress, proper evaluation of accumulated stress is necessary, which will allow the adoption of effective coping strategies, such as exercise and hobbies. Therefore, developing a rapid and accurate method of stress estimation is vital.

Current mainstream stress estimation methods are based on questionnaires, which are easy to administer and enable periodic surveying. However, stress estimation using questionnaires are not highly objective or accurate because of the subjective evaluation of stress. Methods that involve expert evaluations ensure objectivity at the cost of ease and continuity of measurements.

The use of biological information to estimate stress has been explored as an alternative to using questionnaires. Biological information can be measured objectively and continuously without the subject's intentional manipulation. A biometric measure that is typically used to estimate stress is heart rate; stress-induced changes in the activity of the autonomic nervous system are reflected in the R-wave interval (RRI), of which the amplitude of the heartbeat is highest. Medical devices based on contact-type sensors, such as electrocardiography and photoplethysmography are conventionally used to measure the RRI, but such devices are not practical for use in the home or workplace. Moreover, stress can be affected by the wearing of contact sensors; therefore, a non-contact measurement method would be more ideal.

In recent years, non-contact stress estimation methods using camera have been investigated, which involve capturing biological information in the form of facial and skin videos. Specific biological information that can be
obtained using a camera include pulse waves based on pixel value changes due to blood flow, facial expressions based on the motion of facial muscles, and eye blinks. These measures reflect the aforementioned autonomic nervous system activity as well as drowsiness and can be measured continuously in subjects. In a previous study by Mitsuhashi et al. [2], pulse waveforms were extracted from facial videos recorded using a red-green-blue (RGB) camera and classified into four levels of stress. The accuracy of stress estimation was around 80% for each level of stress.

In this paper, we propose a new method for estimating stress using multiple types of biological information to improve the accuracy of stress classification. Moreover, we applied the method to estimate stress that changes continuously over time.

2. Related work

When we encounter a stressor, our body responds in a specific way. Stress stimuli are transmitted to the hypothalamus, which integrates the autonomic, endocrine, and immune systems to maintain balance between the mind and body (homeostasis). However, when excessive stress is experienced over prolonged periods, there is a breakdown of homeostasis and the disease (e.g., melancholia, cerebral hemorrhage, autonomic ataxia) manifests itself. Although acceptable levels of stress vary between individuals, identifying current levels of stress is vital for preventing stress-induced mental illnesses.

The most common method of subjective stress assessment is a questionnaire-based assessment. The advantage of this method is that it is easy for anyone to use once the question format and evaluation method have been established. The downside is the large number of responses that is required for a detailed estimation of stress. This presents challenges for frequent and continuous data collection and analysis. Additionally, because data obtained using questionnaires are based on subjective evaluations, it is difficult to identify symptoms and stress that the subject may not be aware of, which may reduce the accuracy of the estimation. Thus, a simpler estimation method that is more detailed and accurate is needed.

Several biometric methods are currently being investigated as alternatives to questionnaire-based stress estimation. A valuable characteristic of biometric methods is that the intention of the subject does not interfere with the index being acquired; moreover, continuous measurement and analysis are easy. In particular, heart rate-related indices for measuring signs of mental stress, such as heart rate variability (HRV), are now well established. These measures reflect changes in the balance of autonomic nervous system activity. For example, heartbeat has been used to measure stress levels of doctors performing surgeries using contact sensors [3]. The study showed that stress levels were successfully differentiated using multiple unimodal features. However, because the subjects and measurement environment of this study were highly specific, their stress estimation method requires verification in a more generalized experimental setting. A more generalized setting for stress estimation has been employed previously, where stress levels were measured using a contact sensor and a head-mounted display while healthy subjects carried out mental arithmetic tasks [4]. Physiological measures, such as pulse waves, galvanic skin responses, and pupil size were used to extract features and measure stress levels. Such multimodal methods that use multiple biometric data ensure robust and accurate measurements, which enable more detailed analyses of stressors and their effects. However, many contact sensors and probes cause discomfort and thus cannot be used by people with irritated or damaged skin.

Furthermore, the simultaneous use of multiple sensors has shown to increase stress levels in subjects.

To solve the contact sensor problem, the image-based photoplethysmography (iPPG) method was developed, which measures subtle color changes in skin caused by changes in heartbeat. The iPPG method allows non-contact estimation by analyzing video data acquired using a camera.

First, we describe a method that uses the color channels directly. The principle that underlies iPPG is based on the fact that the sensitivity of the green channel of an RGB camera matches the spectral response of oxyhemoglobin [5]. This method has fast implementation by spatially averaging the green channel of the skin image captured by the RGB camera and has the advantage of real-time processing. Although pulse waves extracted from a single channel are susceptible to noise effects, such as body motion, skin surface reflections, and illumination variations, we have demonstrated that pulse rate variability (PRV) measurements are effective [6]. However, these types of noise can cause analysis errors, especially in identifying the peak position of the pulse wave, which can result in major problems in the analysis of autonomic nervous system activity using pulse wave peak intervals.
In stress estimation, it is particularly important to use a high-precision pulse wave detection method that does not cause deviations of the peak position of the pulse wave.

To achieve higher accuracy in pulse wave detection, an optics-based modeling method was proposed. A concrete example of this approach is the two-layer skin model proposed by Fukunishi et al. [7, 8]. They constructed a two-layer skin model that consisted of one melanin and one hemoglobin layer. Assuming that the pigments represented by the melanin and hemoglobin components are distributed independently of each layer, an image of hemoglobin, which is sensitive to blood volume, can be extracted, and spatial principal and spatial independent component analyses can be applied to the face image using a logarithmic transformation. This method is robust against illumination variations because it is represented by a combination of bias and shading components in logarithmic space [9]. After extracting the hemoglobin image from the RGB image, the hemoglobin components in the region of interest (RoI) are spatially averaged to generate a pulse wave. The details of Fukunishi et al.’s approach can be found in [7, 9]. This method of PRV measurement has shown to be more effective than that of Fukunishi, Mcduff, and Tsumura [8].

The study by Mitsuhashi [2] applied this iPPG method to estimate the stress levels step-by-step, not binary classes by extracting multiple features from PRV parameters. Their results showed that the features obtained using the iPPG method were effective for each level of the stress level classification.

3. Proposed method

As mentioned earlier, stress affects autonomic nervous system balance, and that these effects are manifested as physiological changes, such as increased heart rate and pupil dilation. Because these biological changes occur unconsciously to the subject, measuring these changes allows objective observations of autonomic nervous system activity. In our study, we followed the non-contact measurement method used in previous research [2] and aimed to estimate stress more precisely and comprehensively by using multiple biometric data as features.

3.1. Relationship between stress and biosignals

Stress-induced modulation of autonomic nervous system balance is reflected in various parts of the body, such as heartbeat, electroencephalogram (EEG), skin surface potential, and eye movements. In this section, we describe the relationship between the pulse wave, eye blink, and pupil diameter, which are the indices we used in our study for stress estimation.

The pulse wave is defined as a waveform that captures blood pressure or volume changes in the peripheral vascular system that occur as the heart pumps blood. Because the dominance of the sympathetic nervous system by stress stimuli (e.g., anxiety, tension, relationship) causes changes in heart movement, the most important indices for quantifying mental stress— as represented by the activity of the autonomic nervous system—are heart rate and HRV [10, 11, 12]. People who feel stressed have a higher heart rate. HRV represents the variation in the time interval between the peaks of the heart rate waveform. It is well known that this peak fluctuates depending on a person’s mental state. When stressed, the sympathetic nervous system is activated, which is reflected in low frequency fluctuations. Conversely, when in a relaxed state, the parasympathetic nervous system is activated, which is reflected high frequency fluctuations [13, 14, 15]. Based on these observations, mental stress can be classified into two categories: relaxed and stressed states that correspond to HRV frequency components.

Eye blinks have been shown to be an important indicator of drowsiness. Several previous studies [16, 17] have examined techniques for estimating sleepiness by analyzing eyelid movements from facial videos. The study by Haak et al. showed that blinking frequency increases during stressful situations [18].

Several studies have shown that pupils are reliable indicators of cognition and emotion [19, 20]. Pupil size and movement are controlled by the pupillary sphincter and pupil dilator muscles; the former contracts and the latter dilates the pupil. Like pulse rate, the human pupil has been shown to reflect the activity of the autonomic nervous system. Specifically, the pupil dilates (mydriasis) when mental stress is applied [21, 22, 23]. Marco et al. found a significant correlation between pupil diameter and stress in subjects performing a stressful driving task, which demonstrated that pupil diameter has discriminative power in stress detection [24].

3.2. Pulse waveform estimation

In our study, we used a non-contact pulse wave extraction method using skin pigment component separation [8], which was also used in Mitsuhashi et al. [2]. Using this method, the pulse waveform can be estimated by calculating the average pixel value of the facial skin area using the hemoglobin component of the video, which fluctuates with changes in heart rate and blood volume. Furthermore, by applying a band-pass filter to the estimated pulse waveform and extracting and
applying a band of 45–180 beats per minute (the normal range of human heart rate), a more accurate pulse waveform can be obtained. Figure 1 shows the timeseries variation of the pulse waveform obtained from a facial video.

3.3. Feature extraction from the pulse waveform

As mentioned above, changes in biological information due to stress appear as fluctuations in the intervals between the peaks of the pulse waveform. Because pulse fluctuations are calculated based on the difference between adjacent peak times of the pulse waveform, we first calculated pulse fluctuations as shown in Fig. 2. Subsequently, the pulse waveform was analyzed in the time and frequency domains to determine the characteristics of other pulses.

3.3.1. Time-domain analysis

The time-domain features, which were obtained from the pulse waveform intervals shown in Fig. 2, were analyzed by calculating their values directly. Specifically, the value of the pulse wave interval, meanRRI, and the standard deviation, stdRRI, as well as the average value of the pulse rate waveform, meanHR, and the standard deviation, stdHR, were calculated from one pulse wave interval. The stdRRI of the pulse peak interval reflects the overall fluctuation of the pulse wave, whereas the root mean square of the pulse waveform sequential difference, RMSSD, reflects short-term pulse waveform fluctuations. The method used for calculating RMSSD is shown in Eq. (1).

$$RMSSD = \sqrt{\frac{1}{N-1} \sum_{j=1}^{N-1} (RR_{j+1} - RR_j)^2}$$  \hspace{1cm} (1)

where $N$ represents the total number of consecutive pulse waveform intervals and $RR_j$ refers to the $j$th pulse waveform interval. For the time domain, we extracted the five abovementioned features.

3.3.2. Frequency-domain analysis

Features in the frequency domain were obtained by analyzing the power spectral density (PSD) of the pulse peak interval. The PSD was calculated using the periodogram proposed by Lomb and Scargle [25, 26, 27, 28]. The changes in the high- (HF: 0.15 to 0.40 Hz) and low-frequency (LF: 0.04 to 0.15 Hz) components of the pulse waveform fluctuations reflect the activity of the autonomic nervous system.

We calculated $LF$ as the integrated values of the LF band of the PSD, and $HF$ was calculated as the integrated values of the HF band. Furthermore, $nLF$ was calculated as normalized $LF$, $nHF$ was calculated as normalized $HF$, and $LF/HF$ was calculated as the ratio of $LF$ to $HF$; thus, acquiring a total of ten feature types from the pulse wave.

3.4. Blink estimation

Blinks can be captured visually with ease and is an indicator that is highly suitable for extraction from face videos. Soukupová and Cech [29] proposed a method that detects eye blinks from facial video footage using landmarks. We used the same method to acquire 68 landmarks from a facial video using machine learning and determined the degree of eye-opening/closing based on the vertical and horizontal distance fluctuations of six points on the eye. Figure 3 shows an example of the six
landmarks and the vertical and horizontal distance fluctuations. Equation (2) shows the calculation formula for the biological signal of a blink.

\[
\text{EAR} = \frac{p_2 p_4 + p_3 p_5}{2(p_1 p_3)}
\]

The aspect ratio of the eye (EAR) was calculated using Eq. (2) and represents the ratio of the vertical to horizontal distances of the eye based on the six aforementioned landmarks. The horizontal distance of the eye remains the same whether the eye is open or closed, whereas the vertical distance is constant when the eye is open but rapidly decreases when the eye is closed, such as during a blink. These differences are used to determine whether the eye is opened or closed. Figure 4 shows the blink waveform, which is the timeseries change of the EAR value. The region showing a rapid decrease in value signifies a blink.

### 3.5. Feature extraction from blinks

Figure 5 shows the blink waveform and the feature points that were obtained from the waveform. The waveform in Fig. 5 is an enlarged schematic diagram of where the EAR changes sharply in Fig. 4. In Fig. 5, \(P_s\) and \(P_e\) represent the start and endpoint of the blink, respectively, whereas \(P_{ob}\) and \(P_{eb}\) represent the endpoint of the eyelid closing and the start point of the eyelid opening, respectively. \(P_{max}\) represents the time at which the value of EAR reaches the maximum. Based on each of these observations, we extracted 10 feature values.

The amplitude at the point of eyelid closure, \(A_{ci}\), was defined by the difference between the EAR values of \(P_s\) and \(P_{ob}\). The amplitude, \(A_{op}\), at the point of eyelid opening was determined by the difference between the EAR values of \(P_{eb}\) and \(P_e\). The maximum amplitude, \(A_{max}\), was defined by the difference between the average of \(P_{ob}\) and \(P_s\) and the EAR of \(P_{max}\). The eyelid closing speed, \(V_{cl} = A_{ci}/T_{cl}\), was defined by the time difference, \(T_{cl}\), between \(P_s\) and \(P_{ob}\), whereas the eyelid opening speed, \(V_{op} = A_{op}/T_{op}\), was defined by the time difference, \(T_{op}\), between \(P_{eb}\) and \(P_e\). The means and standard deviations of these parameters were obtained as feature values. Additionally, we determined the number of blinks that occur multiple times in a short period of time, \(EB_{numall}\), which was the number of times \(P_{ob}\) and \(P_e\) were detected as individual peak detection points; the number of blinks when the peak of blink is counted as one time until the EAR value drops sharply from the baseline (the section where the value does not change significantly) and stabilizes again, \(EB_{numepoch}\), which was calculated as one blink from \(P_s\) to \(P_e\); and \(Eye\_Closed\_Time\), which was the sum of the time the eyes were closed from \(P_s\) to \(P_e\). In total, we obtained 14 features from the blink waveform.

### 3.6. Pupil estimation

Here, we explain how pupil diameter changes were acquired by fitting circles to the pupil and iris that were segmented from the facial video.

The eye region was extracted from the facial video, and circles were fitted to the pupil and iris, which were segmented using deep learning. The ratio of pupil and iris size was then used to obtain pupil diameter change. Using the ratio of the diameter of the circle fitted to the pupil and iris as an index, we obtained the change in pupil diameter, which was independent of the distance between the camera and the face. Figure 6 shows the procedure for determining the size of the pupil and iris from the facial video. Segmentation of the pupil/iris region of the facial video footage was performed using the U-Net deep learning network, which is a full-layer convolutional network that is specialized for image segmentation. We selected U-net because it requires fewer images for learning, can be trained quickly, and has high segmentation accuracy. The pupil/iris ratio (PIR) was calculated by comparing the diameters of the fitted circles shown in Fig. 6. The PIR value is important
because the iris size remains constant under stress, while the pupil size varies. For example, if the pupil dilates under stress, the PIR value increases because the iris size does not change. Equation (3) was used to calculate the PIR.

\[
P_{IR} = \frac{a_p}{a_i} = \frac{\text{Diameter of Circle Fitted to Pupil}}{\text{Diameter of Circle Fitted to Iris}} \tag{3}
\]

Figure 7 shows an example of a PIR timeseries change calculated using Equation (2). Figure 7 illustrates a similarly sharp decrease that was observed in the blink waveform in Fig. 4. This is likely to be due to the pupil not being detected during a blink, which resulted in the loss of the numerical value. To continuously analyze pupil fluctuations, Hermitian interpolation was performed to compensate for the numerical value loss. Figure 8 shows the result of interpolating the PIR values provided in Fig. 7. As shown in Fig. 8, it was possible to eliminate extreme PIR numerical value changes by performing numerical interpolation.

### 3.7. Feature extraction from the pupil

In this subsection, we explain the extraction of features related to the pupil. As shown in Figs. 7 and 8, we acquired a total of six pupil features. These were meanPIRori, which was the average value of the original PIR shown in Fig. 7; stdPIRori, which was the standard deviation of the original PIR; meanPIRip, which was the average value of the interpolated PIR values shown in Fig. 8; stdPIRip, which was the standard deviation of the interpolated PIR values; minPIR, which was the minimum PIR value; and maxPIR, which was the maximum PIR value. These values did not differ between before and after interpolation.

### 3.8. Feature selection and estimation model by using multiple regression

By the feature extraction described above, 26 types of features are obtained. Here, we describe feature selection and the multiple regression model used for stress prediction. In the stress estimation, various methods are used for prediction, but a method with low computational cost is required when considering future practical applications. We applied multiple regression as the prediction model to estimate the stress. In the prediction model, the explanatory variables are the features obtained from the biometric data, and the dependent variables are the discrete labeled values of the four-step stress task by difficulty level. Since overfitting is expected to occur in multiple regression if all the acquired features are used, it is necessary to select the effective features to be used. To select features, the top 10 features are calculated using the univariate feature extraction of scikit-learn. Then, we calculate Pearson’s product-moment correlation coefficient for each selected feature, and combinations with high values are highly relevant biometric combinations. For example, meanHR and meanRRI are selected in high rank. However, their correlation coefficient value are −0.99, which means they have strong negative correlation (because meanHR is the reciprocal of B). Therefore, we selected meanHR as one
of the explanatory variables from pulse and don’t use meanRRI. The top 5 features in each biometric combination (pulse, blink, pupil, pulse + blink, pulse + pupil, blink + pupil, pupil + blink + pupil) are used as explanatory variables.

4. Experiment

In this study, we measured four levels of stress states. We verified the stress classifications and accuracy using multiple regression analysis-based classifiers based on all the biological features obtained.

Figures 9 and 10 show the experimental setting and procedures. Nine students in their 20s (the age range is 22-24, six men and three women) participated as test subjects. Video footage was captured in a dark room using an RGB camera with a frame rate of 60 fps (frames per second.) and a resolution of $1024 \times 768$ pixels. Artificial sunlight was provided as lighting. Each test subject stabilized their face on a chin rest to prevent noise from face and body movements from affecting the pulse waveform. We recorded a 2-minute facial video of each subject during four stress states (a relaxed state and three states that involved solving mental arithmetic tasks of varying difficulty). There was a 6-minute rest interval between recordings to eliminate any stress effects from the previous task. The mental arithmetic tasks were divided into three levels. For the easiest task, participants were asked to perform multiplication of two single-digit numbers. The moderate difficulty level involved multiplication of a double-digit number with a single-digit number. For the most difficult task, participants were asked to multiply two double-digit numbers. During each recording, subjects were instructed to perform the task while looking directly at the camera and keeping their eyes open as much as possible to ensure accurate recordings of blink and pupil data. Subjects completed the questionnaire based on simplified version of the STAI designed to measure state anxiety after each recording session to measure subjective levels of stress experienced by participants. A questionnaire has 6 questions about emotions which subject feel in each session. Figure 11 shows the questionnaire used in this study. The evaluation of the questionnaire is performed by summing the responses of the subjects for each session. The questionnaire was designed with the expectation that the total value would increase as stress increased, and the responses of the subjects to the items of "Calm", "Relax", and "Satisfaction", which were expected to decrease as stress increased, were inverted when adding up the total (e.g. if the subject answered 1, 2, 3, or 4 for "Calm", we inverted it to 4, 3, 2, or 1 before summing.). Figure 12
shows the summing result of questionnaire for each session. In the task in which the participants are performing mental arithmetic indicated by the index 1-3, we can see that the total value of the questionnaire also increases with each increase in difficulty. On the other hand, the sum value of 0: Relax state, is higher than the value of 1: Easy mental arithmetic. When the responses to each item in these two tasks were compared, the "Relax" and "Satisfaction" items had larger values in the 0: Relax state, while the other four items remained unchanged. When we asked subjects after answering the questionnaire, more of them answered that they were more accustomed to the shooting environment for "Relax" than for "1". As for "Satisfaction," more subjects felt a sense of fulfillment to answer a simple 2-minute mental calculation task. So, we judged that the difference between task "0" and "1" does not come from mental stress. This provided the relationship between mental stress caused by mental arithmetic and the result of questionnaire.

5. Results

We compared classification accuracies of the biological features obtained from the facial videos using multiple regression, both alone and in various combinations (i.e., pulse only, blinks only, pulse and blinks, and all biosignals). Five features were selected based on the Pearson's product-moment correlation coefficient for each combination of biological signals, and these were then used for the stress estimations. We also performed cross validation using the leave-one-out method to verify the accuracies of our stress estimations.

To estimate stress levels, we created a model formula by performing multiple regression analysis on training data and calculating a predictive label value using the test data. The multiple regression analysis estimation method calculates the estimation value as a constant value, so it shows the possibility of estimating the changes in stress state timeseries as the future work. Moreover, the use of a multiple regression model is ideal for future developments of real-time stress estimation systems owing to its low computational cost.

As an indicator of classification accuracy, we compared the average value of the difference between the correct label and the prediction value using root mean square error (RMSE). Specifically, we expressed the correct label by discretely assigning numerical values to each state when the video was taken, as shown with red letters in Fig. 10. Equation (4) was used to calculate the RMSE for the correct answer label and the predicted value.

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 }
\]

where \(N\) presents the number of data, \(y_i\) presents the prediction value, and \(\hat{y}_i\) refers to the correct label value. A lower RMSE value indicates better estimation accuracy.

Table 1 shows 5 selected feature values in each set of biological information. We select these combinations of feature values as the result of feature selection. Note the feature selection for the case of combining all biometric data in the bottom row of Table 1. It is worth noting that LF/HF, which has been shown to be an effective index for estimating stress, was not selected as a feature. This demonstrates that features were selected from each of pulse waves, blinks, and pupil data.

Figure 13 shows the RMSE for each combination of biological information. The RMSE value is 0.828 when all three biological information are combined, whereas the RMSE value is 0.914 when only the pulse used in previous studies is used for stress estimation. We also calculate the RMSE for the other combinations of biological information; The RMSE values do not become smaller when any single biometric is used (Blink: 0.980, Pupil: 0.987), and when two biometrics are combined,
The combination of pulse and pupil has the lowest RMSE value (Pulse + Blink: 0.862, Pulse + Pupil: 0.831, Blink + Pupil: 0.933).

We also compared the coefficient of determination of the multiple regression analysis using all data. The coefficient of determination is an index that indicates goodness of fit of the model to the data. Higher values indicate better model fit. Figure 14 shows the coefficients for each biological information combination. The coefficient of determination is also low when using a single biometric (Pulse: 0.344, Blink: 0.220, Pupil: 0.294), and increases as the number of biometric combinations increases, from two biometric combinations (Pulse + Blink: 0.413, Pulse + Pupil: 0.443, Blink + Pupil: 0.347) to all three (0.503).

We verified the accuracy of the estimation using the leave-one-out method, and the best accuracy was still 0.611 when pulse, blink, and pupil were combined.

As shown in Figs. 13 and 14, we found that estimation accuracy improved by including more biological indicators and that the estimation accuracy was highest when pulse, blinking rate, and pupil dilation values were combined.

Figure 15 shows the difference in the numerical value of A for each subject for each stress task. In both cases, the number of blinks during the stress task tended to be higher than the number of blinks during relaxation. For all subjects, the number of times they blinked during middle difficulty mental arithmetic increased compared to the relaxed state.

6. Discussion

In this study, $LF/HF$, which has been shown to be an effective index for estimating stress, was not selected as a feature. We believe that factors other than the stress task should also be considered, such as the environmental load imposed by the video footage being recorded in a dark room while the test subject's face was illuminated by a bright light, or long-term stress associated with the daily mental state of the test subjects. To eliminate the effects of such factors, we considered that the environmental stress imposed during video recording can be minimized by providing sufficient time for the test subject to adapt to the environment before acquiring measurements. Moreover, the effect of long-term stress can be mitigated by providing time before taking measurements in they stress level for each subject. In most of the subjects, the number blinks increased as stress level increased, which is consistent with the research of Haak et al. [18]. In general, there are three types of blinks: spontaneous blink, reflex blink, and voluntary blink [30]. Spontaneous blinks are physiological blinks, while reflexive blinks are caused by stimuli such as light shining into the eye. In this study, the experiment was conducted in a laboratory environment, and no lighting variation occurred between tasks. To equalize the effect of spontaneous eye blink as much as possible between tasks, the subjects were instructed to be in a sufficient state without thirst before the start of the shooting. It is known that voluntary eye blinks increase with tension and decrease with thought and concentration. For the "nervousness" item of the questionnaire, the mean value of each subject increased as the task difficulty increased (0: 1.67, 1: 1.78, 2: 1.89, 3: 2.89). This is similar to the trend of increasing blink rate shown in Figure 15. Though spontaneous blinks and reflexive blinks may also have an effect, we speculate that the increase in blinking was the strong influence of voluntary blink, which reflect human's mental state. Based on these results, our future plan is to use chronologically varying blinks as an indicator of a given task, which will allow us to more specifically infer the source of stress.
7. Conclusion

In this study, we proposed a method for acquiring three types of biological information from RGB facial videos to estimate four levels of stress. We combined multiple types of biological information and the result shows the possibility to improve stress estimation accuracy. In future studies, we plan to verify the effectiveness of our method by applying appropriate stressors to test subjects. There is overwhelming evidence that stress decreases the activity of immune system cells, such as lymphocytes, macrophages, granulocytes, and natural killer cells. In particular, cytokines have been suggested to be closely related to stress. Identifying the association between biomarkers and the features selected in our study will strengthen the reliability of our measurement model. We also plan to apply this method to more complex stress situations. Furthermore, to provide more detailed analysis using pupil information, such as that using pulse waveforms, it will be necessary to develop better methods for analyzing pupil changes. Note that the method used in this study is multiple regression, in which we assumed the label value are linear. However, in order to demonstrate the usefulness of using multimodal biometric data, it is necessary to compare nonlinear methods such as k-nearest neighbor method, SVM, machine learning, etc. In addition, the subjects were stressed by performing a mental arithmetic task. Stress can be classified according to the way the stress stimulus is given and the factors involved. The stress felt by the subjects in this study was a temporary mental stress. One of our future projects is to classify chronic stress and temporary stress by using multibiological information. Finally, this study was conducted in a laboratory environment. In future studies, we will conduct measurements using an experimental setup that is similar to that of a typical office in subjects with a wider range of ages, so that the stress estimation model is more generalizable.

Acknowledgments

We thank Sarina Iwabuchi, PhD, from Edanz Group (https://en-author-services.edanz.com/ac) for editing a draft of this manuscript.

References

1) G.M. Slavich: "Life Stress and Health: A Review of Conceptual Issues and Recent Findings", Teaching of Psychology. 43, 4, pp.346-353 (2016)
2) R. Mitsuhashi, K. Iuchi, T. Goto, A. Matsubara, T. Hirayama, H. Hashizume and N. Tsumura: "Video-based stress level measurement using imaging photoplethysmography", IEEE International Conference on Multimedia & Expo Workshops (ICMEW). pp.90-95 (2019)
3) M. Weenk, A. Aiken, L. Engelen, S. Bredie, T. Belt and H. Goor.: "Stress measurement in surgeons and residents using a smart patch", The American Journal of Surgery. 216, pp.361-368 (2018)
4) R. Abdi, A. Alhiatary, E. Hay and A. Al-bashir: "Objective detection of chronic stress using physiological parameters", Medical & Biological Engineering & Computing. 56, 12, pp.2273-2286 (Dec. 2018)
5) W. Verkruysse, L.O. Svassand and J.S. Nelson: "Remote plethysmographic imaging using ambient light", Optics Express. 16, 26, pp.21341-21445 (2008)
6) K. Alghoul, S. Alharthi, H. Al Osman and A. El Suddik: "Heart rate variability extraction from videos signals: ICA vs. EVM comparison", IEEE Access. 5, pp.4711-4719 (2017)
7) M. Fukunishi, K. Kurita, S. Yamamoto and N. Tsumura: "Non-contact video-based estimation of heart rate variability spectrogram from hemoglobin concentration", Artificial Life and Robotics. 22, 4, pp.457-463 (Dec. 2017)
8) M. Fukunishi, D. Mcduff and N. Tsumura: "Improvements in remote video based estimation of heart rate variability using the Welch FFT method", Journal of Artificial Life and Robotics. 23, 1, pp.15-22 (Mar. 2018)
9) N. Tsumura, N. Ojima, K. Sato, M. Shiraishi, H. Shimizu, H. Nabeshima, S. Akazaki, K. Hori and Y. Miyake: "Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin", ACM Transactions on Graphics. 22, 3, pp.770-779 (July 2003)
10) A. Moriguchi, A. Otsuka, K. Kohara, H. Mikami, K. Katoriha, T. Tsumetoshi, K. Higashimori, M. Ohishi, Y. Ya and T. Ogihara: "Spectral change in heart rate variability in response to mental arithmetic before and after the beta-adrenoceptor blocker, carteolol", Clinical Autonomic Research. 2, 4, pp.287-290 (1992)
11) N. Hjortskov, D. Rissen, A.K. Blangsted, N. Fallentin, U. Lundberg and K. Sigaard: "The effect of mental stress on heart rate variability and blood pressure during computer work. European journal of applied physiology", 92, 1-2, pp.84-89 (2004)
12) Z.B. Moses, L.J. Lueckcn and J.C. Eason: "Measuring task-related changes in heart rate variability", Engineering in Medicine and Biology Society 29th Annual International Conference of the IEEE. pp.644-647 (2007)
13) F. Macer, K. Fagard and G. A. Armitage: "Power spectral analysis of RR interval and blood pressure short-term variability at rest and during dynamic exercise: comparison between cyclists and controls", Int J Sports Med. 17, pp.157-161 (1996)
14) M. Pagani, F. Lombardi, S. Guzzetti, O. Rimoldi, R. Furlan, P. Pizzinelli, G. Sandrone, G. Maloffto, S. Dell’Orto, E. Piccaluga.; "Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog", Cire Res. 9, pp.178-193 (1986)
15) M. Malik and A.J. Camm: "Heart rate variability", Clin Cardiol. 13, pp.570-576 (1990)
16) D.F. Dinges, R. Grace., "PERCLOS: A valid psychophysiological measure of alertness as assessed by psychomotor vigilance", U.S. Department of Transportation, Federal Highway Administration, FHWA-MCRT-98-006, 10 (1998)
17) M. Tsujikawa, Y. Onishi, Y. Kiuchi, T. Ogatsu, A. Nishino, S. Hashimoto: "Drowsiness Estimation from Low-Frame-Rate Facial Videos using Eyelid Variability Features", 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 7 (2018)
18) M. Haak, S. Bos, S. Panic and L.J.M. Rothkrantz: "Detecting stress using eye blinks and brain activity from EEG signals", Proceeding of the 1st driver car interaction and interface (DCII 2008). pp.35-60 (2009)
19) E. Gram hod and S.R. Stein hauer: "Pupilometric measures of cognitive and emotional process", International Journal of
20) J. Kuhlmann and M. Böttcher: “Pupillography: Principles, methods and applications”, W. Zuckschwerdt Verlag, München (1999)

21) T. Partala and V. Surakka: “Pupil size variation as an indicator of affective processing”, International Journal of Human-Computer Studies. 59, pp.185-198 (2003)

22) M.M. Bradley, L. Miccoli, M.A. Escrig and P.J. Lang: “The pupil as a measure of emotional arousal and autonomic activation”, Psychophysiology. 45, pp.602-607 (2008) doi: 10.1111/j.1469-8986.2008.00655

23) M.L.H. Vo, A.M. Jacobs, L. Kuchinke, M. Hofmann, M. Conrad, A. Schacht and F. Hutzel: “The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect”, Psychophysiology. 45, pp.130-140 (2008) doi: 10.1111/j.1469-8986.2007.00600

24) P. Marco, A.M. Mohammad, T. Adrien, C. Jean-Rémy, M. Frédéric, B. Simone and B. Thierry: “Automatic stress classification with pupil diameter analysis”, Journal of Human-Computer Interaction. 30, pp.1-17 (2014)

25) J.H. Horne and S.L. Baliunas: “A Prescription for Period Analysis of Unevenly Sampled Time Series”, Astrophysical Journal. Vol. 302, pp.757-763 (1986)

26) N.R. Lomb: “Least-squares frequency analysis of unequally spaced data”, Astrophysics and Space Science. Vol. 39, pp.447-462 (1976)

27) W.H. Press and G.B. Rybicki: “Fast Algorithm for Spectral Analysis of Unevenly Sampled Data”, Astrophysical Journal. Vol. 338, pp.277-280 (1989)

28) J.D. Scargle: “Studies in Astronomical Time Series Analysis. Ö†. Statistical Aspects of Spectral Analysis of Unevenly Spaced Data”, Astrophysical Journal. Vol. 263, pp.835-853 (1982)

29) T. Soukupova and J. Čech: “Real-Time Eye Blink Detection using Facial Landmarks”, 21st Computer Vision Winter Workshop (2016)

30) A.A. V., Cruz, D.M. Garcia, C.T. Pinto and S.P. Cechetti: “Spontaneous Eyeblink Activity”, The Ocular Surface. 9, 1, pp.29-41 (2011)

31) H. Anisman: "Stress, immunity, cytokines and depression", Blackwell, Munksgaard, 14, pp.251-261 (2002)