Person–Independent Computational Recognition of Emotions Elicited Using Japanese Kanji Words: Machine Learning Approach Using Multimodal Physiological Signals*

Kazuhiko Takahashi[1], Shin-ya Namikawa[2], Masafumi Hashimoto[3]

[1]Information Systems Design, Doshisha University, 1–3 Miyakodani Tatara Kyotanabe Kyoto, Japan
[2]Graduate School of Doshisha University, 1–3 Miyakodani Tatara Kyotanabe Kyoto, Japan
[3]Intelligent Information Engineering and Sciences, Doshisha University, 1–3 Miyakodani Tatara Kyotanabe Kyoto, Japan
[1]katakaha@mail.doshisha.ac.jp; [2]dtk0746@mail4.doshisha.ac.jp; [3]mhashimo@mail.doshisha.ac.jp

Abstract- **In this paper, person–independent computational emotion recognition using multimodal physiological signals - in this study, plethysmogram, skin conductance change, respiration rate and skin temperature - is investigated. Psychophysical experiments are conducted using Japanese kanji words in order to excite three emotions, such as positive, negative and neutral, in subjects and thus elicit physiological signals. The concept of machine learning approaches, such as multilayer neural networks, support vector machines, decision trees and random forests, is used to design emotion recognition systems, and the recognition systems are trained and tested using gathered data under the psychophysical experiments to investigate their characteristics. In experiments of computational emotion recognition, the maximum average recognition rates are 38% using multilayer neural networks, 40% using support vector machines equipped with a Gaussian kernel function, 37% using decision trees and 33% using random forests for all three emotions. The results of the emotion recognition experiments show that using multimodal physiological signals with a machine learning approach is feasible and appropriate for person–independent computational emotion recognition.**

Keywords- Emotion Recognition; Japanese Kanji Words; Physiological Signal; Neural Networks; Support Vector Machines; Decision Trees; Random Forests

I.INTRODUCTION

In human communication, nonverbal information, such as intentions and emotions, plays an important role. In particular, emotional information enables people to communicate with each other more smoothly. Thus, it is evident that the exchange of nonverbal information is pivotal in all forms of communication and is occasionally more important than verbal information. Furthermore, this suggests that nonverbal communication forms the basis of human communication. In addition to human-to-human communication, communication between humans and machines is becoming increasingly common. To achieve more intimate and human–like interactions between humans and machines, the use of both verbal and nonverbal information will be essential in human–machine interface systems and human–computer interactions [1].

Emotion recognition using computers is an interesting but difficult task [2, 3]. People can recognise emotional speech with an accuracy of around 60% and emotional facial expressions with an accuracy of around 70–98%. In studies on computational emotion recognition/affective computing [4], emotional speech was recognised at a rate of 50–60% [5], and facial expressions had a recognition rate of 80–90% [6, 7]. However, a limited number of emotional categories and consciously and purposefully expressed emotions have usually been investigated because such emotions are easier in terms of recognition, control and data collection. To improve the recognition accuracy obtained in such single–modal emotion recognition, many studies have attempted to exploit the advantage of using multimodal information, especially by fusing audio-visual information. On the other hand, physiological signals, such as electrodermal activity, heart rate, electrocardiograms, electromyograms and blood volume pressure, are also useful for evaluating emotions [8–18]; this is because these vital signs are generally controlled by the autonomic nervous system, which is affected by felt emotions. The physiological signal datasets used in most previous studies were obtained using visual elicitation methods in which the subjects look at selected photographs or watch movies in a laboratory setting, and a recognition rate of over 80% on an average has been achieved. However, the recognition rate depends strongly on the

*This article is based on a study "Computational Emotion Recognition Using Multimodal Physiological Signals: Elicited using Japanese Kanji Words" which is first reported in the 35th International Conference on Telecommunications and Signal Processing in July 2012.

datasets; that is, the number of emotional categories and the kinds of emotional categories to use were different on the datasets, moreover the datasets were usually composed of specific subjects with specific stimuli in laboratory conditions although person independence and context independence are important and a desirable quality in emotion recognition. Another possible way of evaluating emotion from physiological information is electroencephalography, or the use of brainwaves, which is an index of the central nervous system [19–21]. Brainwaves might be effective in evaluating emotions because emotions are excited in the limbic system and are deeply related to cognitive processing. However, in such a method, special equipment and/or electrodes are necessary to collect brainwave data, and advanced ability to handle the equipment/sensors and process the data is also pre–requisites.

In this paper, we investigate computational emotion recognition from multimodal physiological signals [22] under the following assumptions. (1) Consciously and purposefully elicited emotions are considered because they are easier to stimulate in humans. (2) Three emotional states (positive, negative and neutral) are selected in light of several studies on computational emotion recognition systems. These emotions are elicited by presenting Japanese kanji words that have a positive, negative or neutral affective valence to human subjects. Because the kanji words are ideographs and have both a morphologic/logographic and phonetic script, their features are different from those of phonetic scripts such as the English alphabet. The attributes of kanji words, e.g. imagery, ease of learning, representation and affective valence, have been investigated; single or compound kanji words have been found to have an affective effect on Japanese [23, 24]. (3) A combination of four physiological signals, the plethysmogram, skin conductance change, respiration rate and skin temperature, is considered because these signals can be easily obtained without any special equipment or expertise in handling/processing the data. In psychology, Russell [25] proposed that emotions could be explained using a two–dimensional plane of emotional valence (pleasure–displeasure) vs. arousal valence (aroused–sleepy). For example, the emotional valence can be evaluated using the heart rate, whereas the arousal valence can be estimated using electrodermal activity. (4) An emotion recognition system achieves person–independence by training with several different human subjects. The final objective of computational emotion recognition is to make it person–independent under unconscious conditions, but this is a difficult issue that will require long–term research. Therefore, in this study, we try to achieve person–independent emotion recognition by gathering data from multiple subjects in limited emotional categories and using machine learning approaches [26]. The rest of this paper is arranged as follows. In Section 2, the collection of physiological signals using multimodal sensors during psychophysical experiments is described. In Section 3, a computational emotion recognition system based on machine–learning approaches and the results of computational emotion recognition experiments are presented.

II.DATA COLLECTION

A.Physiological Signal Sensing System

Figure 1 shows a schematic of a multimodal physiological signal sensing system for gathering physiological signals from a human subject. The sensing system comprises four sensors and two personal computers: one is used to present stimuli to a subject, and the other is used to acquire the physiological signals (plethysmogram, skin conductance change, respiration rate and skin temperature) from the subject.

The plethysmogram is measured by a pulse oximeter that consists of a sensor clip (PP-C012, TEAC Co.) and a physiological signal amplifier. This sensor clip uses a reflective optical sensor and measures the pulse wave pattern of a fingertip. The sensor clip, with a diameter of 1 cm, is mounted on the subject's fingertips and the finger is wrapped in a piece of cloth to secure the sensor clip.

The skin conductance change is measured by a skin conductance metre that consists of two electrodes (PPS-EDA, TEAC Co.), an electrodermal activity (EDA) unit (AP-U030, TEAC Co.) and the physiological signal amplifier. The disposable Ag/AgCl electrodes are mounted on the subject's fingertips. The change in skin conductance at the fingertips is measured by the variation

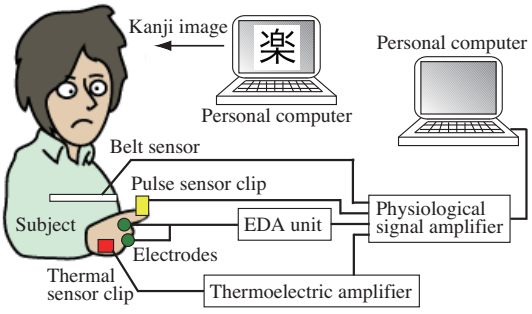

Fig. 1 Schematic of multimodal physiological signal sensing system

negative 死 殺 悲 病 嫌
neutral 野 語 権 図 村
positive 幸 笑 楽 晴 福

Fig. 2 Japanese kanji words (Left to right for 'negative': death, kill, sorrow, illness, disgust. Left to right for 'neutral': field, word, right, figure, village. Left to right for 'positive': happiness, laughter, pleasure, sunny, fortune)

TABLE 1 EVALUATION RESULTS OF KANJI WORD IN THE PSYCHOLOGICAL EXPERIMENT

	Positive					Negative					Neutral				
Subject	幸	笑	楽	晴	福	死	殺	悲	病	嫌	野	語	権	図	村
a	4	4.5	4	5	4	1	1	2	1	1	3	3	3	3	3
b	5	4.5	4	4.5	4.5	1	1.5	2	1.5	2	3	3	4	3	3
c	5	4	4	5	4.5	2	1	2	1	2	3	3	3	3	3
d	5	5	4.5	4	5	1	1	2	2	1.5	3	3	3	3	3
e	5	5	5	5	5	1	1	1	1	1	3	3	3.5	3	3
f	4	4	4.5	4	4	2	2	2.5	2	1.5	3	3	3	3	3
g	2	4.5	3.5	4	5	1	2	2	1	2	4	3	3	4.5	4
h	5	5	5	5	5	1	1	1.5	1	1.5	3	3	3	3	3
i	4	5	4.5	4	5	1.5	1	2	2	2	3	3	3	3	3
j	5	4.5	5	5	5	1	1	2	2	2	3	3	2	3	4
k	5	5	5	4	4	1	1	2	1	1	3	3	3	3	3
l	5	5	5	5	5	1	1	1	1	1	4	3	4	4	4
m	4.5	5	5	4	5	1	1	1	1	2	3	3	4	3	3

in electrical current in the EDA unit.

The respiration rate is measured by a respiration sensor that consists of a hook-and-loop belt sensor (AP-C021, TEAC Co.) and the physiological signal amplifier. The belt extends around the subject's chest cavity and is held in place by a small elastic band, which stretches as the subject's chest cavity expands. The amount of stretch in the belt is measured as a change in voltage.

The skin temperature is measured by a thermal metre that consists of a sensor clip (AP-C050, TEAC Co.), a thermoelectric amplifier (AP-U019, TEAC Co.) and the physiological signal amplifier. The sensor clip is mounted on the subject's fingertip. The change in the skin temperature at the fingertip is measured by the variations in electrical voltage in the thermoelectric amplifier.

The outputs from these sensors are amplified in the physiological signal amplifier (Polymate AP1000, TEAC Co., 32 ch, AD converter: 16 bit, maximum sampling frequency: 2 kHz) and are input to one of the personal computers (CF-R6, Panasonic).

B.Psychological Experiments

In our experiment, three emotions, positive, negative and neutral, were considered. These emotions were stimulated using Japanese kanji words that were selected because their meanings were expected to elicit the corresponding emotions from Japanese people. To evaluate whether a kanji word excites each emotion, questionnaire investigations of subjects who did not take part in the experiment involving the multimodal physiological signal sensing system were conducted. A total of 13 subjects (males, Japanese, age: 21–24) participated in the evaluation experiment. The questionnaire on feeling emotions was evaluated with five grades ranging from 'negative' (score: 1) to 'positive' (score: 5) for each kanji word. Figure 2 shows the 15 kanji words [27] that were chosen to elicit each emotion.

Table 1 shows the results of the questionnaires and Fig. 3 illustrates the average score for each emotion. Each score accurately corresponds to the classifications. Applying the Wilcoxon rank sum test to all the combinations of the scores for kanji words, yielded differences with a significance level of 5%. Applying the Friedman test to the average scores for the emotions, yielded differences with a significance level of 5%. These results indicate that most subjects could feel each emotion when exposed to the kanji words.

C.Psychophysical Experiments

By using the multimodal physiological signal sensing system, physiological signals were collected in psychophysical experiments that used the kanji words as stimuli to excite emotions. By considering a guide to gathering physiological data for affective recognition described in an earlier study [8], the psychophysical experiments were conducted following an *event-excited*, *laboratory setting*, *feeling*, *open-recording* and *emotion-purpose* methodology. The experiments were conducted in a private room in our laboratory where the illumination, sounds and room temperatures were controlled to maintain uniformity.

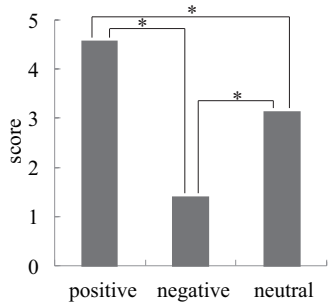

Fig. 3 Average score for each emotion evaluated by human subjects in psychological experiments (*: $p < 0.05$)

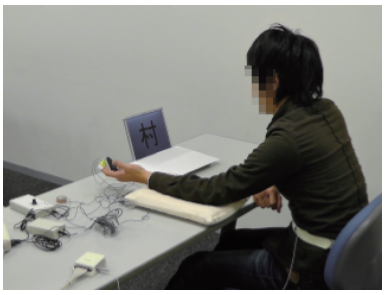

Fig. 4 Experimental setup of physiological signal measurement

In the psychophysical experiment, a total of 15 subjects (14 males and 1 female, Japanese, age: 21–24) participated. A text image illustrating a kanji word drawn in black on a white background was presented on the display of a personal computer. While the text image was presented to the subject, raw physiological signals were collected from each subject twice for each of the three emotions by using the physiological signal measurement system, as shown in Fig. 4. The subject was asked to passively view the text image. One trial of the experiment was conducted as follows.

1. Baseline: Measuring the physiological signals without any stimulus.
2. Stimulating emotion: Measuring the physiological signals for 30 s while/after presenting a kanji word on the display of the personal computer. The displayed kanji word is randomly selected from the stimulus datasets.
3. Evaluating emotion: The subject evaluates her/his emotions by answering a questionnaire that investigates whether she/he can feel the emotion. In the questionnaire, the emotions were evaluated with seven grades ranging from 'negative' (score: 1) to 'positive' (score: 7).
4. Conclude the experiment if all kanji words are evaluated; otherwise, return to Step 2.

In the psychophysical experiment, 450 samples of physiological signals together with the emotional classification labels were collected. Table 2 shows the results of the emotion evaluation obtained using the questionnaires and Fig. 5 shows the average score for each emotion. Applying the Wilcoxon rank sum test to all the combinations of the scores for kanji words, yielded differences with a significance level of 5%. Applying the Friedman test to the average scores among the emotions, yielded differences with a significance level of 5%. Thus, we consider that most subjects could feel each emotion during the psychophysical experiment conducted to gather physiological signals.

III.COMPUTATIONAL EMOTION RECOGNITION EXPERIMENTS

A.Feature Extraction

Figure 6 shows examples of the raw physiological signals measured from a subject during the psychophysical experiments; the sampling frequency is 200 Hz. To conduct computational emotion recognition, the features of each physiological signal must be extracted from the raw signals. In this study, statistical and physiology–dependent features [8] of each physiological signal are

TABLE 2 EVALUATION RESULTS OF KANJI WORD IN THE PSYCHOPHYSICAL EXPERIMENT

	Positive					Negative					Neutral				
Subject	幸	笑	楽	晴	福	死	殺	悲	病	嫌	野	語	権	図	村
A	7	7	7	7	7	1	1	1	1	1	4	4	6	4	4
B	5.5	6	6	5	5.5	1	3.5	2	3	2	4.5	4	4.5	4.5	4
C	6	6.5	6	6	5	1	2	3	2.5	2.5	4	4	4	4.5	4
D	6	5.5	6.5	5.5	6	1	1	2	2	2	3.5	3.5	3	4	4
E	5	4.5	4.5	5	4.5	2	3	4	2.5	3	4	5	3.5	5	4
F	7	6.5	6.5	6	6.5	1.5	1.5	2.5	1	2	4	3	3	2.5	3
G	7	7	7	6	6.5	1	1	1.5	2	1.5	4.5	4	4	4	4
H	6.5	7	7	6	6.5	2	1	2	2	1	4	4.5	5	5	4.5
I	7	5.5	6	6	5	1	1.5	3	2	1.5	5	4.5	5	4.5	4.5
J	6	7	7	6.5	5	1.5	1	1.5	2.5	2	4	4.5	3	4	4.5
K	5	5.5	5.5	5.5	4.4	2	2	3	2.5	2.5	4	4	4	4	4
L	6.5	6.5	6.5	5.5	4.4	1	1	3	2.5	2	4	4	2.5	4.5	4
M	5	5	5	5	6	1.5	2	3	2	3	4	4	4.5	4	4
N	4.5	4	3	3	4	3	3.5	3	3.5	4.5	3.5	4	3	3.5	3.5
O	5.5	7	6.5	6	5.5	1	1	2	2.5	2	4	4	4	3.5	4

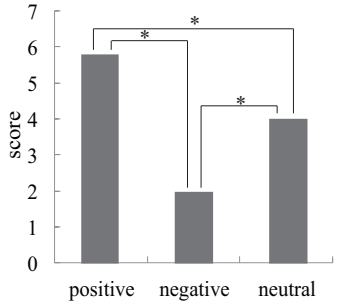

Fig. 5 Average score for each emotion evaluated by human subjects in psychophysical experiments for gathering physiological signals (*: $p < 0.05$)

considered. The statistical features are as follows.

$$\mu_X = \frac{1}{K}\sum_{k=1}^{K} X_k \tag{1}$$

$$\sigma_X = \sqrt{\frac{1}{K-1}\sum_{k=1}^{K}\{X_k - \mu_X\}^2} \tag{2}$$

$$\delta_X = \frac{1}{K-1}\sum_{k=1}^{K-1}|X_{k+1} - X_k| \tag{3}$$

$$\bar{\delta}_X = \frac{\delta_X}{\sigma_X} \tag{4}$$

$$\gamma_X = \frac{1}{K-2}\sum_{k=1}^{K-2}|X_{k+2} - 2X_{k+1} + X_k| \tag{5}$$

$$\bar{\gamma}_X = \frac{\gamma_X}{\sigma_X} \tag{6}$$

$$\alpha_X = \max X_k \tag{7}$$

$$\beta_X = \min X_k \tag{8}$$

Here, X_k indicates the physiological signal $X = \{\mathcal{P}, \mathcal{S}, \mathcal{R}, \mathcal{T}\}$ for sample number k and K is the total number of samples in one trial of the psychophysical experiment. The symbols $\mathcal{P}$, $\mathcal{S}$, $\mathcal{R}$ and $\mathcal{T}$ represent the plethysmogram, skin conductance change, respiration rate and skin temperature, respectively.

The heartbeat rate $\mathcal{H}$ is estimated from the plethysmogram. By applying fast Fourier transform analysis to the pulse waveform $\mathcal{P}_k$ convolved with a Hamming window, the reciprocal of the frequency that has the maximum power spectrum is defined

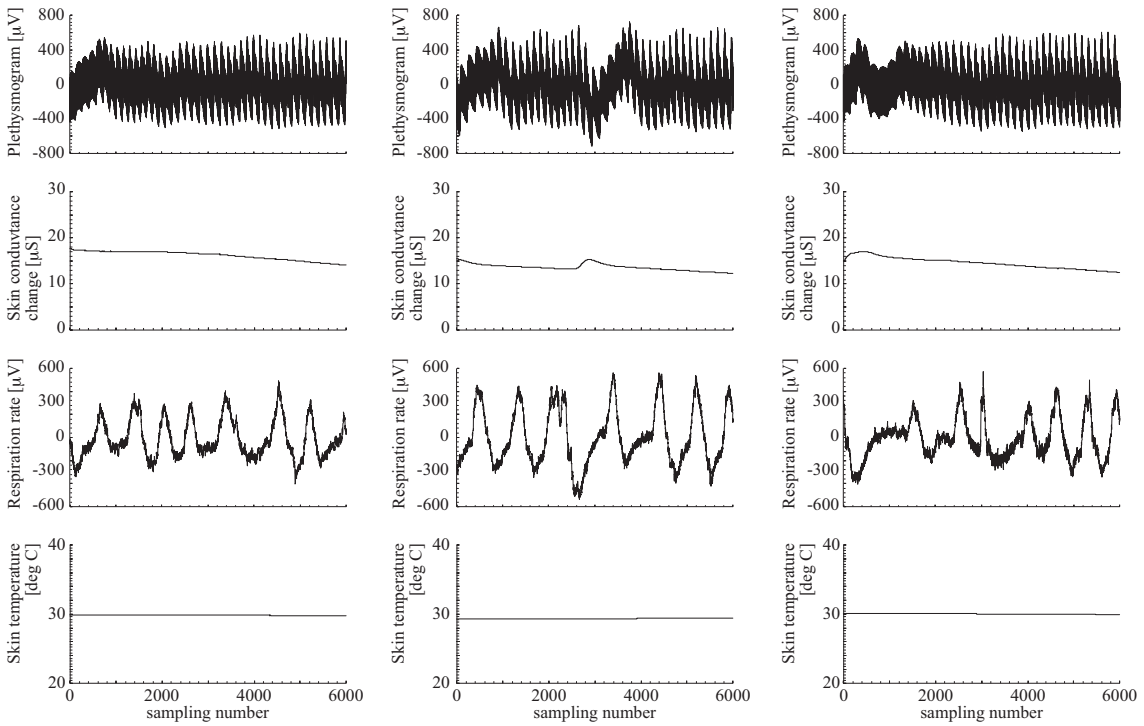

Fig. 6 Examples of raw physiological signals (Top to bottom: plethysmogram, skin conductance change, respiration rate and skin temperature. Left to right: positive, negative and neutral emotions.)

as the heartbeat rate $\mathcal{H}_k$. To reduce the noise and the baseline fluctuations in the skin conductance change, a form of contrast normalisation, $\bar{X}_k = \frac{X_k - \min X_k}{\max X_k - \min X_k}$, is applied to the skin conductance change waveform $\mathcal{S}_k$ convolved with a Hanning window. The skin temperature waveform $\mathcal{T}_k$ is also applied to the form for contrast normalisation. The physiology–dependent features are as follows.

$$f_X = \frac{1}{K_n} \sum_{k=1}^{K_n} X_k \tag{9}$$

$$d_X = \frac{1}{K_n - 1} (X_{K_n} - X_1) \tag{10}$$

Here $X = \{\mathcal{H}, \bar{\mathcal{S}}, \bar{\mathcal{T}}\}$, and K_n is the total number of samples to which the convolution is applied. The respiration waveform $\mathcal{R}_k$ is modified by the mean of the overall respiration data to compensate for day–to–day variations in sensor placement: $\bar{\mathcal{R}}_k = \mathcal{R}_k - \frac{1}{K_a} \sum_{k=1}^{K_a} \mathcal{R}_k$ where K_a is the total number of samples for a subject. Two features, $\mu_{\bar{\mathcal{R}}}$ and $\sigma_{\bar{\mathcal{R}}}$, and the first four 0.1–Hz bands of the power spectral density in the range of 0.0–0.4 Hz in the respiration waveform $p_{i_{\bar{\mathcal{R}}}}$ $(i = 1, \cdots, 4)$ were used as the physiology–dependent features.

The feature vector utilised in the experiments for computational emotion recognition is defined using 32 statistical features and 12 physiology–dependent features.

B.Emotion Recognition System

This study investigated and compared the following standard machine learning approaches to achieve computational emotion recognition systems.

1. Multilayer neural networks (MNNs) with standard back-propagation learning
2. Support vector machines (SVMs) with varying kernel functions
3. Decision trees (DTs) with a Gini diversity index
4. Random forests (RFs) for ensemble learning

As a reference for these machine-learning approaches, a learning vector quantization method (LVQ), a linear classifier based on a Fisher linear discriminant analysis (FLDA), and a nonlinear classifier based on the Bayes rule (naive Bayes classifier, NBC) were conducted.

TABLE 3 CONFUSION MATRIX OF COMPUTATIONAL EMOTION RECOGNITION USING PHYSIOLOGICAL FEATURES (THREE EMOTIONS)

		All features			Principal components			Selected features		
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
MNNs	negative	**0.30**	0.31	0.39	**0.38**	0.35	0.27	**0.33**	0.31	0.36
	neutral	0.37	**0.34**	0.29	0.35	**0.36**	0.26	0.24	**0.39**	0.37
	positive	0.35	0.31	**0.34**	0.34	0.35	**0.41**	0.27	0.36	**0.37**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
SVMs$^{(g)}$	negative	**0.38**	0.33	0.29	**0.41**	0.27	0.32	**0.17**	0.49	0.34
	neutral	0.38	**0.31**	0.31	0.42	**0.31**	0.27	0.13	**0.57**	0.30
	positive	0.38	0.29	**0.33**	0.45	0.25	**0.30**	0.12	0.41	**0.47**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
SVMs$^{(p)}$	negative	**0.41**	0.31	0.28	**0.42**	0.30	0.28	**0.26**	0.53	0.21
	neutral	0.39	**0.29**	0.32	0.41	**0.31**	0.28	0.33	**0.40**	0.27
	positive	0.42	0.26	**0.32**	0.44	0.27	**0.29**	0.41	0.38	**0.21**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
DTs	negative	**0.23**	0.37	0.40	**0.30**	0.28	0.42	**0.29**	0.42	0.29
	neutral	0.29	**0.46**	0.25	0.28	**0.35**	0.37	0.33	**0.40**	0.27
	positive	0.35	0.42	**0.23**	0.36	0.29	**0.35**	0.27	0.32	**0.41**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
RFs	negative	**0.25**	0.35	0.40	**0.30**	0.38	0.32	**0.29**	0.40	0.31
	neutral	0.34	**0.27**	0.39	0.34	**0.35**	0.31	0.37	**0.30**	0.33
	positive	0.35	0.39	**0.26**	0.40	0.30	**0.30**	0.24	0.36	**0.40**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
LVQ	negative	**0.32**	0.38	0.30	**0.37**	0.37	0.26	**0.32**	0.41	0.27
	neutral	0.37	**0.35**	0.28	0.34	**0.33**	0.33	0.33	**0.37**	0.30
	positive	0.38	0.30	**0.32**	0.28	0.32	**0.42**	0.26	0.33	**0.41**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
FLDA	negative	**0.27**	0.40	0.33	**0.25**	0.42	0.33	**0.17**	0.45	0.38
	neutral	0.38	**0.29**	0.33	0.29	**0.41**	0.30	0.17	**0.52**	0.31
	positive	0.39	0.33	**0.28**	0.37	0.33	**0.30**	0.16	0.37	**0.47**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
NBC	negative	**0.43**	0.35	0.22	**0.28**	0.39	0.33	**0.37**	0.31	0.32
	neutral	0.51	**0.21**	0.28	0.43	**0.35**	0.22	0.41	**0.31**	0.28
	positive	0.51	0.34	**0.15**	0.41	0.36	**0.23**	0.36	0.31	**0.33**

C.Emotion Recognition Results

In the experiments on computational emotion recognition, first the parameters for each recognition system were tuned using bootstrap datasets generated from the collected samples of physiological signals so as to achieve a higher recognition rate. The size of the bootstrap dataset was the same as that of the samples of physiological signals, and the total number of dataset was 100. Consequently, the MNNs were tuned to a 44–5–3 network topology. In the SVMs, the inverse kernel width of the Gaussian kernel function was 10^{-2}, and the margin parameter was 10^3. The margin parameter in the second–order polynomial kernel function was also 10^3. In the DTs, the complexity parameter was tuned to 10^{-5}. The number of trees generated in the RFs was 800, and 50 variables were randomly sampled as candidates at each split. Next, the emotion recognition system was trained and tested by leave–one–out cross validation. The left column in Table 3 shows the results of computational emotion recognition in which all the features were used in the feature vector. The averaged recognition rates were 33% using the MNNs, 34% using the SVMs with the Gaussian kernel (SVMs$^{(g)}$), 34% using the SVMs with the second polynomial kernel (SVMs$^{(p)}$), 31% using the DTs, 26% using the RFs, 33% using the LVQ, 28% using the FLDA and 26% using the NBC. Subjects self–evaluated that they could feel each emotion in the psychophysical experiments, as shown in Fig. 5; however, the emotion recognition system had some difficulty in recognising the three emotions. To pre–process the feature vector, principal component analysis (PCA) was applied to the extracted features. Figure 7 shows the contribution ratio; the principal components are arranged in order to decrease eigenvalue. The contribution ratio exceeds 90% for the 18th principal component. Thus, the feature vector was composed of the first 18 principal components. In the MNNs, there was an 18–10–3 network topology. In the SVMs, the inverse kernel width of the Gaussian kernel function was 10^{-2}, and the margin parameter was 10^4, whereas the margin parameter was 10^2 in the second–order polynomial kernel function. In the DTs, the complexity parameter was 10^{-5}. The number of trees generated in the RFs was 600, and 20 variables were randomly sampled as candidates at each split. The centre column in Table 3 shows the results of computational emotion recognition using the 18 principal components. The averaged recognition rates were 38% using the MNNs, 34% using the SVMs$^{(g)}$, 34% using the SVMs$^{(p)}$, 33% using the DTs, 32% using the RFs, 37% using the LVQ, 32% using the FLDA and 29% using the NBC. Although the dimension of the feature vector in the emotion recognition system is reduced by using the PCA and the averaged recognition rates are improved in the MNNs and RFs, the pre–processing by PCA has little influence on the other emotion recognition systems.

To investigate the features, which are relevant to classification of each emotional state and reduce the dimension of the

TABLE 4 CONFUSION MATRIX OF COMPUTATIONAL EMOTION RECOGNITION USING PHYSIOLOGICAL FEATURES (TWO EMOTIONS)

		All features		Principal components		Selected features	
In \ Out		negative	positive	negative	positive	negative	positive
MNNs	negative	**0.51**	0.49	**0.52**	0.48	**0.45**	0.55
	positive	0.54	**0.46**	0.57	**0.43**	0.46	**0.54**
In \ Out		negative	positive	negative	positive	negative	positive
SVMs$^{(g)}$	negative	**0.50**	0.50	**0.51**	0.49	**0.61**	0.39
	positive	0.46	**0.54**	0.52	**0.48**	0.49	**0.51**
In \ Out		negative	positive	negative	positive	negative	positive
SVMs$^{(p)}$	negative	**0.49**	0.51	**0.52**	0.48	**0.63**	0.37
	positive	0.47	**0.53**	0.50	**0.50**	0.61	**0.39**
In \ Out		negative	positive	negative	positive	negative	positive
DTs	negative	**0.49**	0.51	**0.46**	0.54	**0.50**	0.50
	positive	0.35	**0.65**	0.51	**0.49**	0.47	**0.53**
In \ Out		negative	positive	negative	positive	negative	positive
RFs	negative	**0.47**	0.53	**0.43**	0.57	**0.54**	0.46
	positive	0.58	**0.42**	0.56	**0.44**	0.53	**0.47**
In \ Out		negative	positive	negative	positive	negative	positive
LVQ	negative	**0.53**	0.47	**0.56**	0.44	**0.55**	0.45
	positive	0.67	**0.33**	0.60	**0.55**	0.69	**0.31**
In \ Out		negative	positive	negative	positive	negative	positive
FLDA	negative	**0.47**	0.53	**0.46**	0.54	**0.69**	0.31
	positive	0.57	**0.43**	0.60	**0.40**	0.55	**0.45**
In \ Out		negative	positive	negative	positive	negative	positive
NBC	negative	**0.61**	0.39	**0.59**	0.41	**0.61**	0.39
	positive	0.68	**0.32**	0.61	**0.39**	0.63	**0.37**

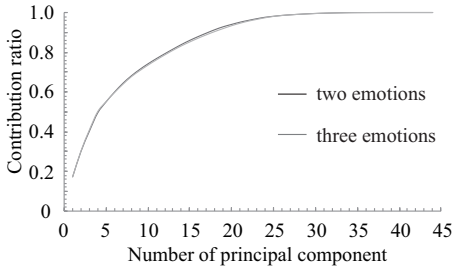

Fig. 7 Relationship between the number of principal components and contribution ratio

feature vector in the emotion recognition system, features were selected using the sequential forward selection (SFS) method; this method starts with an empty set and the feature that fits best is inserted at each step. The FLDA was used to calculate the classification rate in the SFS evaluation. As shown in Fig. 8, six features that achieve the maximum classification rate were selected, as follows: $f_{\bar{\mathcal{S}}}, \sigma_{\mathcal{R}}, d_{\bar{\mathcal{H}}}, \mu_{\mathcal{R}}, p_{3_{\bar{\mathcal{R}}}}$ and $p_{4_{\bar{\mathcal{R}}}}$. In the MNNs, there was a 6–7–3 network topology. In the SVMs, the margin parameter was 1 in both kernel functions, whereas the inverse kernel width of the Gaussian kernel function was 10^{-1}. In the DTs, the complexity parameter was 10^{-5}. The number of trees generated in the RFs was 600, and 30 variables were randomly sampled as candidates at each split. The right column in Table 3 shows the results of computational emotion recognition using the six selected features. The averaged recognition rates were 36% using the MNNs, 40% using the SVMs$^{(g)}$, 29% using the SVMs$^{(p)}$, 37% using the DTs, 33% using the RFs, 37% using the LVQ, 39% using the FLDA and 34% using the NBC. The averaged recognition rates of all the emotion recognition systems were improved slightly by using feature selection; however, the emotion recognition systems have some difficulty in recognising negative emotions.

The left column in Table 4 shows the results of computational emotion recognition considering only two emotions (positive and negative); all features were used in the feature vector. In the MNNs, there was a 44–9–2 network topology. In the SVMs, the inverse kernel width of the Gaussian kernel function was 10^{-3}, and the margin parameter was 10^4, whereas the margin parameter was 10^2 in the second–order polynomial kernel function. In the DTs, the complexity parameter was 10^{-2}. The number of trees generated in the RFs was 10^3, and 40 variables were randomly sampled as candidates at each split. The averaged recognition rates were 49% using the MNNs, 52% using the SVMs$^{(g)}$, 51% using the SVMs$^{(p)}$, 57% using the DTs, 45% using the RFs, 43% using the LVQ, 45% using the FLDA and 47% using the NBC. The centre column in Table 4 shows the results of computational emotion recognition with the feature vector reduced by PCA. As shown in Fig. 7, the variation in the contribution rate for two emotions is almost the same as that for three emotions. Thus, the first 18 principal components, which yield a contribution ratio

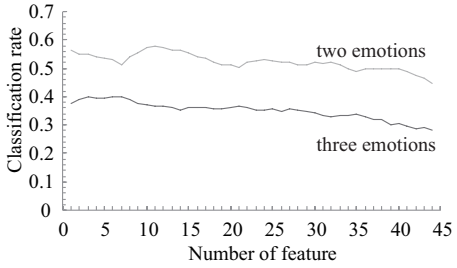

Fig. 8 Relationship between the number of features and classification rate obtained by using SFS

TABLE 5 RESULTS OF BINOMIAL TEST FOR THE RECOGNITION RATES BETWEEN MACHINE LEARNING APPROACHES AND RANDOM APPROACH

Emotions	Feature vector	**MNNs**	**SVMs$^{(g)}$**	**SVMs$^{(p)}$**	**DTs**	**RFs**
Three	All features	-0.095	0.222	0.222	-0.739	-2.398
	Principal components	1.472***	0.222	0.222	-0.096	-0.416
	Selected features	0.851	2.086*	-1.393	1.162	-0.096
Two	All features	-0.245	0.490	0.245	1.719**	-1.226
	Principal components	-0.490	0.000	0.245	-0.490	-1.472
	Selected features	0.000	1.472***	0.245	0.490	0.245

(*: $p < 0.05$, **: $p < 0.1$, ***: $p < 0.2$)

of more than 90%, were selected. In the MNNs, there was an 18–8–2 network topology. In the SVMs, the inverse kernel width of the Gaussian kernel function was 10^{-2}, and the margin parameter was 10^4, whereas the margin parameter was 10^2 in the second–order polynomial kernel function. In the DTs, the complexity parameter was 10^{-5}. The number of trees generated in the RFs was 500, and 50 variables were randomly sampled as candidates at each split. The averaged recognition rates were 48% using the MNNs, 50% using the SVMs$^{(g)}$, 51% using the SVMs$^{(p)}$, 48% using the DTs, 44% using the RFs, 56% using the LVQ, 43% using the FLDA and 49% using the NBC. The right column in Table 4 shows the results of computational emotion recognition with the features selected by SFS. Here 11 features that yield the maximum classification rate, as shown in Fig. 8, were selected, as follows: $d_{\bar{\mathcal{T}}}$, $p_{2_{\bar{\mathcal{R}}}}$, $d_{\mathcal{H}}$, $p_{4_{\bar{\mathcal{R}}}}$, $p_{3_{\bar{\mathcal{R}}}}$, $\mu_{\mathcal{R}}$, $\bar{\delta}_{\mathcal{T}}$, $\delta_{\mathcal{S}}$, $\sigma_{\mathcal{T}}$, $\delta_{\mathcal{T}}$ and $\sigma_{\bar{\mathcal{R}}}$. In the MNNs, there was an 11–5–2 network topology. In the SVMs, the margin parameter was 10 in both kernel functions, whereas the inverse kernel width of the Gaussian kernel function was 10^{-1}. In the DTs, the complexity parameter was 10^{-2}. The number of trees generated in the RFs was 500, and 10 variables were randomly sampled as candidates at each split. The averaged recognition rates were 50% using the MNNs, 56% using the SVMs$^{(g)}$, 51% using the SVMs$^{(p)}$, 52% using the DTs, 51% using the RFs, 43% using the LVQ, 57% using the FLDA and 49% using the NBC. In all the emotion recognition systems, the emotion recognition rates are higher than those obtained in the experiments for three emotions because the number of emotional classes is smaller. The effect of pre–processing by PCA or feature selection by SFS is almost the same as that in the case of recognising three emotions. These results indicate that using multimodal physiological signals with machine learning approaches has the potential to achieve person–independent computational emotion recognition, however the emotion recognition rates are inadequate because those are slightly better than a random approach (the random approach is expected to have the emotion recognition rate of 33.3% in three emotions and 50% in two emotions). Table 5 shows the test statistics that are calculated in the binomial test [28] between the recognition rates obtained by the machine learning approaches and the random approach. The SVM with Gaussian kernel function, which achieved the best rate of emotion recognition by using the selected feature vector, yielded differences with a significance level of 5% in three emotions and of 20% in two emotions.

In this study, kanji words were used as stimuli to induce emotions in subjects. Every kanji word generally has multiple parts of speech in Japanese. The meaning of the kanji word sometimes changes according to the part of speech. In the cognitive process of understanding the meaning of the kanji word and experiencing emotion, which part of speech the subject assumes when interpreting the kanji word would influence what the subject associates with the word, even though this depends greatly on her/his culture, education, memory and experiences. Although there is obviously no direct relationship between the affective valence and the parts of speech in kanji words, computational emotional recognition was tested by adding information on the parts of speech to the feature vector. The kanji words shown in Fig. 2 can be identified in terms of the parts of speech; e.g. the word for 'happiness' is used as three parts of speech (noun, verb and adverb), as shown in Table 7 where '1' indicates that the kanji is used as the part of speech, and '0' indicates otherwise. In computational emotion recognition, three emotions were considered and the parameters of each emotion recognition system were the same as those in the case of Table 3. The left column of Table 6 shows the results of computational emotional recognition where the feature vector contained only the information on the parts of speech. All of the emotion recognition systems yielded similar recognition results; however, there were still some difficulties in recognising three emotions from the parts of speech. The middle column in Table 6 shows the results of

TABLE 6 CONFUSION MATRIX OF COMPUTATIONAL EMOTION RECOGNITION USING PHYSIOLOGICAL FEATURES AND PARTS OF SPEECH (THREE EMOTIONS)

		Parts of speech			All features, parts of speech			Selected features, parts of speech		
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
MNNs	negative	**0.40**	0.20	0.40	**0.63**	0.15	0.22	**0.61**	0.14	0.25
	neutral	0.07	**0.80**	0.29	0.16	**0.61**	0.22	0.16	**0.65**	0.19
	positive	0.40	0.20	**0.40**	0.27	0.21	**0.52**	0.25	0.14	**0.61**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
SVMs$^{(g)}$	negative	**0.40**	0.20	0.40	**0.58**	0.13	0.29	**0.62**	0.18	0.20
	neutral	0.20	**0.80**	0.00	0.21	**0.56**	0.23	0.16	**0.75**	0.09
	positive	0.40	0.20	**0.40**	0.35	0.18	**0.47**	0.36	0.20	**0.44**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
SVMs$^{(p)}$	negative	**0.40**	0.20	0.40	**0.53**	0.19	0.28	**0.50**	0.19	0.31
	neutral	0.00	**0.80**	0.20	0.24	**0.51**	0.25	0.13	**0.73**	0.14
	positive	0.40	0.20	**0.40**	0.35	0.24	**0.41**	0.31	0.20	**0.49**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
DTs	negative	**0.40**	0.20	0.40	**0.54**	0.21	0.25	**0.58**	0.20	0.22
	neutral	0.20	**0.80**	0.00	0.17	**0.66**	0.17	0.26	**0.61**	0.13
	positive	0.40	0.20	**0.40**	0.27	0.16	**0.57**	0.21	0.16	**0.63**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
RFs	negative	**0.43**	0.20	0.37	**0.50**	0.19	0.31	**0.61**	0.16	0.23
	neutral	0.09	**0.80**	0.11	0.15	**0.69**	0.16	0.18	**0.67**	0.15
	positive	0.39	0.20	**0.41**	0.32	0.15	**0.53**	0.27	0.16	**0.57**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
LVQ	negative	**0.54**	0.25	0.21	**0.51**	0.25	0.24	**0.56**	0.26	0.18
	neutral	0.13	**0.71**	0.16	0.12	**0.71**	0.17	0.06	**0.74**	0.20
	positive	0.19	0.30	**0.51**	0.24	0.23	**0.53**	0.18	0.27	**0.55**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
FLDA	negative	**0.80**	0.20	0.00	**0.67**	0.16	0.17	**0.77**	0.16	0.07
	neutral	0.20	**0.80**	0.00	0.20	**0.61**	0.19	0.21	**0.71**	0.08
	positive	0.40	0.20	**0.40**	0.39	0.14	**0.47**	0.39	0.14	**0.47**
In \ Out		negative	neutral	positive	negative	neutral	positive	negative	neutral	positive
NBC	negative	**1.00**	0.00	0.00	**0.71**	0.10	0.19	**0.85**	0.02	0.13
	neutral	0.60	**0.40**	0.00	0.49	**0.31**	0.20	0.43	**0.37**	0.20
	positive	1.00	0.00	**0.00**	0.67	0.14	**0.19**	0.75	0.07	**0.18**

TABLE 7 PARTS OF SPEECH OF KANJI WORDS SHOWN IN FIG. 2

kanji	noun	verb	adjective	adverb	particle
幸	1	1	0	1	0
笑	1	1	0	0	0
楽	1	1	1	0	0
晴	1	1	1	0	0
福	1	1	1	0	0
死	1	1	1	1	0
殺	0	1	0	0	1
悲	1	1	1	0	0
病	1	1	1	0	0
嫌	1	1	0	0	0
野	1	0	1	0	0
語	1	1	0	0	0
権	1	1	0	0	0
図	1	1	1	0	0
村	1	0	1	0	0

computational emotional recognition where the feature vector consists of all the features extracted from the physiological signals and the information on the parts of speech. The right column in Table 6 shows the results where the feature vector consists of the features selected by SFS and the information on the parts of speech. A comparison of Table 6 with Table 3, reveals that using information about the stimuli would be more helpful for improving the emotion recognition rates than either examining more types of physiological signals or extracting other features from them. However, it might not necessarily be suitable for the

objective in this study, which is to achieve computational emotion recognition from physiological signals.

IV.CONCLUSIONS

This paper investigated person–independent computational emotion recognition using multimodal physiological signals. Four physiological signs, the plethysmogram, skin conductance change, respiration rate and skin temperature, were measured to evaluate three emotional states: positive, negative and neutral. Psychophysical experiments using 15 Japanese kanji words to excite emotions in subjects were conducted to gather physiological signals. Statistical and physiology–dependent features were extracted from the signals gathered from 15 subjects. For computational emotion recognition, machine learning approaches, multilayer neural networks, support vector machines, decision trees and random forests, were used to design emotion recognition systems and their characteristics were investigated. In experimental emotion recognition, support vector machines with a Gaussian kernel function using the feature components selected by the sequential forward selection achieved a maximum averaged recognition rate of around 40% for all three emotions and around 56% for two emotions (positive and negative). The results obtained in this study demonstrated that using multimodal physiological signals with a machine learning approach is feasible for person–independent computational emotion recognition.

REFERENCES

[1] A. Jaimes and N. Sebe, "Multimodal Human Computer Interaction: A Survey", *Computer Vision and Image Understanding*, Vol. 108, No. 1–2, pp. 116–134, 2007.

[2] R. D. Ward and P. H. Marsden, "Affective Computing: Problems, Reactions and Intentions", *Interacting with Computers*, Vol. 16, pp. 707–713, 2004.

[3] J. Bullington, "'Affective' Computing and Emotional Recognition Systems: The Future of Biometric Surveillance?", In *Proceedings of the 2nd Annual Conference on Information Security Curriculum Development*, pp. 95–99, 2005.

[4] G. Matthews, M. Zeidner, and R. D. Roberts (ed.), *The Science of Emotional Intelligence : Knowns and Unknowns*, Oxford University Press, 2007.

[5] M. El Ayadi, M. S. Kamel, and F Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases", *Pattern Recognition*, Vol. 44, pp. 527–587, 2011.

[6] R. W. Picard, "Affective Computing: Challenges", *International Journal of Human-Computer Studies*, Vol. 59, No. 1–2, pp. 55–64, 2003.

[7] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 1, pp. 39–58, 2009.

[8] R. W. Picard, E. Vyzas, and J. Healey, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 10, pp. 1175–1191, 2001.

[9] F. Nasoz, K. Alvarez, C. Lisetti, and N. Finkelstein, "Emotion Recognition from Physiological Signals for Presence Technologies", *International Journal of Cognition, Technology and Work*, Vol. 6, No. 1, pp. 4–14, 2004.

[10] A. Faro, D. Giordano, and S. Mineo, "A Fuzzy Processor Based Distributed System for Implementing Affective Human-Computer Interfaces", In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2693–2698, 2004.

[11] A. Haag, S. Goronzy, and J. Williams, "Emotion Recognition Using Bio-Sensors: First Steps Towards an Automatic System", In *Proceedings of Affective Dialogue Systems*, LNCS 3068, pp. 36–48, 2004.

[12] J. Wagner, J. Kim, and F. André, "From Physiological Signals To Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification", In *Proceedings of IEEE international Conference on Multimedia and Expo*, pp. 940–943, 2005.

[13] A. Gračanin, I. Kardum, and J. Hudek-Knežević, "Relations Between Dispositional Expressivity and Physiological Changes During Acute Positive and Negative Affect", *Psychological Topics*, Vol. 16, No. 2, pp. 311–328, 2007.

[14] C. Maaoui, A. Pruski, and F. Abdat, "Emotion Recognition for Human-machine Communication", In *Proceedings of International Conference on Intelligent Robots and Systems*, pp. 1210–1215, 2008.

[15] R. A. Calvo and S. D'Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications", *IEEE Transactions on Affective Computing*, Vol. 1, No. 1, pp. 18–37, 2010.

[16] M. S. Hussain, O. AlZoubi, R. A. Calvo, and S. D'Mello, "Affect Detection from Multichannel Physiology during Learning Sessions with AutoTutor", In *Proceedings of 15th International Conference on Artificial Intelligence in Education*, LNCS 6738, pp. 131–138, 2011.

[17] V. Kolodyazhniy, S. Kreibig, J. Gross, W. T. Roth, and F. Wilhelm, "An Affective Computing Approach to Physiological Emotion Specificity: Toward Subject-Independent and Stimulus-Independent Classification of Film Induced Emotions", *Psychophysiology*, Vol. 48, pp. 908–922, 2011.

[18] S. Jerritta, M. Murugappen, R. Nagarajan, and W. Khairunizam, "Physiological Signals Based Human Emotion Recognition: A Review", In *Proceedings of 7th International Conference on Signal Processing and its Applications*, pp. 410–415, 2011.

[19] K. Ishino and M. Hagiwara, "A Feeling Estimation System Using a Simple Electroencephalograph", In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, pp. 4204–4209, 2003.

[20] K. Takahashi, "Comparison of Emotion Recognition Methods from Bio-Potential Signals", *The Japan Journal of Ergonomics*, Vol. 40, No. 2, pp. 90–98, 2004.

[21] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion Recognition From EEG Using Higher Order Crossings", *IEEE Transactions on Information Technology In Biomedicine*, Vol. 14, No. 2, pp. 186–197, 2010.

[22] K. Takahashi, S. Namikawa, and M. Hashimoto, "Computational Emotion Recognition Using Multimodal Physiological Signals: Elicited using Japanese Kanji Words", In *Proceedings of 35th International Conference on Telecommunications and Signal Processing*, pp. 615–620, 2012.

[23] F. Gotoh and N. Ohta, "Affective Valence of Two-compound Kanji Words", *Tsukuba Psychological Research*, Vol. 23, pp. 45–52, 2001.

[24] T. Ogawa and N. Suzuki, "On the Saliency of Negative Stimuli: Evidence from Attention Blink", *Japanese Psychological Research*, Vol. 46, No. 1, pp. 20–30, 2004.

[25] J. A. Russell, "A Circumplex Model of Affect", *Journal of Personality and Social Psychology*, Vol. 39, pp. 1161–1178, 1980.

[26] M. E. Müller, "Why Some Emotional States Are Easier to be Recognized Than Others: A thorough data analysis and a very accurate rough set classifier", In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, pp. 1624–1629, 2006.

[27] M. Sato and M. Haraguchi, "The Attentional Blink with Single-kanji Stimuli", *Fukuoka University Review of Literature & Humanities*, Vol. 40, No. 1, pp. 35–48, 2009.

[28] S. Nakagawa and H. Takagi, "Statistical Methods for Comparing Pattern Recognition Algorithms and Comments on Evaluating Speech Recognition Performance", *The Journal of the Acoustical Society of Japan*, Vol. 50, No. 10, pp. 849–854, 1994.

Kazuhiko Takahashi received the B.E., M.E. and D.E. degrees in mechanical engineering from Tohoku University, Sendai, Japan, in 1988, 1990 and 1997, respectively. He joined Nippon Telegraph and Telephone Corporation in 1990 and worked on measurement and control of information mechatronics systems. He moved to Doshisha University in 2004 via Yamaguchi University and has studied intelligent control systems and man–machine systems.

Shin-ya Namikawa received the B.E. and M.E. degrees in information engineering from Doshisha University, Kyoto, Japan, in 2010 and 2012, respectively. He has engaged in research on computational emotion recognition at the Graduate School of Doshisha University. He joined YKK Corporation in 2012.

Masafumi Hashimoto received the B.E., M.E. and D.E. degrees in aeronautical engineering from Osaka Prefecture University, Osaka, Japan, in 1979, 1981 and 1988, respectively. Beginning in 1981, he worked for the department of aeronautical engineering, Osaka Prefecture University. He moved to Doshisha University in 2004 via Hiroshima University and has studied intelligent vehicle automation systems.