A Bayesian Algorithm for Anxiety Index Prediction Based on Cerebral Blood Oxygenation in the Prefrontal Cortex Measured by Near Infrared Spectroscopy

**YUKIKATSU FUKUDA[1], YASUTOSHI IDA[1], TAKASHI MATSUMOTO[1], (Life Fellow, IEEE),
NAOHIRO TAKEMURA[2], AND KAORU SAKATANI[2]**

[1]Department of Electrical Engineering and Bioscience, Waseda University, Tokyo 169-8050, Japan
[2]College of Engineering, Nihon University, Tokyo 101-8308, Japan
CORRESPONDING AUTHOR: T. MATSUMOTO (takashi@matsumoto.elec.waseda.ac.jp)

ABSTRACT   Stress-induced psychological and somatic diseases are virtually endemic nowadays. Written self-report anxiety measures are available; however, these indices tend to be time consuming to acquire. For medical patients, completing written reports can be burdensome if they are weak, in pain, or in acute anxiety states. Consequently, simple and fast non-invasive methods for assessing stress response from neurophysiological data are essential. In this paper, we report on a study that makes predictions of the state-trait anxiety inventory (STAI) index from oxyhemoglobin and deoxyhemoglobin concentration changes of the prefrontal cortex using a two-channel portable near-infrared spectroscopy device. Predictions are achieved by constructing machine learning algorithms within a Bayesian framework with nonlinear basis function together with Markov Chain Monte Carlo implementation. In this paper, prediction experiments were performed against four different data sets, i.e., two comprising young subjects, and the remaining two comprising elderly subjects. The number of subjects in each data set varied between 17 and 20 and each subject participated only once. They were not asked to perform any task; instead, they were at rest. The root mean square errors for the four groups were 6.20, 6.62, 4.50, and 6.38, respectively. There appeared to be no significant distinctions of prediction accuracies between age groups and since the STAI are defined between 20 and 80, the predictions appeared reasonably accurate. The results indicate potential applications to practical situations such as stress management and medical practice.

INDEX TERMS   Anxiety index, blood oxygenation, health and safety, near infrared spectroscopy, neuronal activity, oxyhemoglobin, prediction methods, prevention medicine, regional cerebral blood flow, translational engineering.

I. INTRODUCTION

Stress-induced psychological and somatic diseases are becoming endemic in many contemporary societies. A first step in the prevention of stress-induced diseases could be to quantify anxieties that individuals have, although quantification *per se* may not prevent the diseases. While a variety of written self-report anxiety measures provide valid and reliable measures of anxiety, these indices tend to be time consuming to acquire. For medical patients, completing a written report can be burdensome if they are weak, in pain, or in acute anxiety states. Cumbersomeness is another aspect of such written tests that needs improvement. Thus, simple and fast non-invasive methods for assessing stress response from neurophysiological data are essential.

In [1]–[7], we used a near-infrared spectroscopy (NIRS) device to investigate neurophysiological mechanisms of stress responses. One of our findings was that the prefrontal cortex (PFC) plays an important role in stress response as the

asymmetry of PFC activity measured by NIRS correlated with behavioral and somatic responses to mental stress. It has been observed that there is left/right asymmetry of oxy hemoglobin concentrations in PFC activities during tasks involving mental stress. We measured these activity patterns using a two-channel NIRS device instead of fMRI or EEG and found that they correlated with the systemic stress responses of the autonomic nervous system and the hypothalamic-pituitary-adrenal (HPA) axis system [3]–[7].

Our present study attempted to go beyond the study of such asymmetries of PFC activities in two directions:

A) The first direction was to consider the resting state of a subject instead of the subject being under a stress task. This makes task design easier because it is nontrivial and often difficult to design a stress task that is effective for every subject. It should be noted that research on the brain under resting conditions is an active study area [8]–[14], and it has been reported that the brain at rest consumes a significant amount of energy and much activities with distinctive patterns are reported.

B) The other direction was to make predictions about anxiety index from oxy and deoxy hemoglobin concentration changes of PFC using a two-channel portable NIRS device. We achieved this by constructing a machine learning algorithm within a Bayesian framework and implemented it with Markov Chain Monte Carlo (MCMC). A brief overview of MCMC is given in Appendix B.

Specifically the study proceeded in the following steps:
i) Acquire the anxiety index via the state-trait anxiety inventory (STAI) test.
ii) Use a two-channel portable NIRS device to measure oxy and deoxy hemoglobin concentration changes of PFC in four different groups, where two of them comprise young subjects, whereas the other two comprise elderly subjects.
iii) Place each subject in a resting state in which no task is performed.
iv) Build a Hierarchical Bayesian machine learning algorithm for anxiety index predictions by taking into account both oxy and deoxy hemoglobin concentration changes. Because the amount of data was limited, the predictive capability of the algorithm was evaluated by excluding one set of data and using the remaining set for learning. Following the learning phase, the set of data previously withheld was used to examine the predictive capability of the algorithm. The Hierarchical Bayesian algorithm does not need a validation set since it does not utilize leave-one-out cross validation.
v) Evaluate the prediction capabilities of the algorithm against four different datasets from each of which one set of data was withheld for testing.

The remainder of this paper is organized as follows: Section II reviews work related to our study. Section III describes the materials and methods used as well as the experimental settings, subjects, machine learning algorithm, and feature extraction method. Section IV presents our prediction results. Section V discusses the overall study and the results obtained. Section VI concludes this paper and outlines future work. The appendix outlines the measurement principle of the NIRS device employed in the study.

II. RELATED WORK

NIRS is a relatively well established method of measuring oxy hemoglobin (oxy-Hb) and deoxy hemoglobin (deoxy-Hb) concentrations in cerebral vessels [15]. Changes in oxy-Hb are correlated with changes in regional cerebral blood flow (rCBF) [16], whereas changes in total hemoglobin (sum of oxy-Hb and deoxy-Hb; t-Hb) reflect regional cerebral blood volume (rCBV) changes [17]–[19].

The results of a number of studies indicate that correlations exist between the electrical neural activity and changes in the blood oxygenation levels measured by NIRS. Simultaneous recordings of NIRS and visual evoked potential signals in humans during visual stimulation reveal a linear correlation between hemodynamic changes and evoked potential amplitude [20]. Further, simultaneous measurements of NIRS and EEG at rest indicate that an increase in oxy-Hb is associated with an increase in neuronal activity and vice versa [21], [22]. fMRI measurements in the resting state also display the same phenomena [23].

Various written tests are available for assessing anxieties. A verbal evaluation method attempted in [24] proved faster than the written method and demonstrated a high correlation with STAI. In [25], a visual analog scale was considered and compared with STAI.

We conclude this section by stating that this paper is a significantly improved and expanded version of [26] in that four datasets from different age groups are analyzed here, whereas only one target dataset was used in [26].

III. MATERIALS AND METHODS
A. EXPERIMENTAL SETTINGS
We acquired four datasets from different locations and on different occasions in 2010 and 2011. Each individual participated only once.
i) Dataset 1
Population: 19 subjects
(six women; 13 men), age range: 19–26 years.
ii) Dataset 2
Population: 19 subjects
(13 women; six men), age range: 20–24 years.
iii) Dataset 3
Population: 17 subjects
(10 women; seven men), age range: 61–79 years.
iv) Dataset 4
Population: 20 subjects
(16 women; four men), age range: 60–79 years.

All the subjects were healthy, with no history of psychiatric or neurological disorders. They all gave written informed consent on forms approved by the ethical committees of the Nihon University School of Medicine and Waseda University.
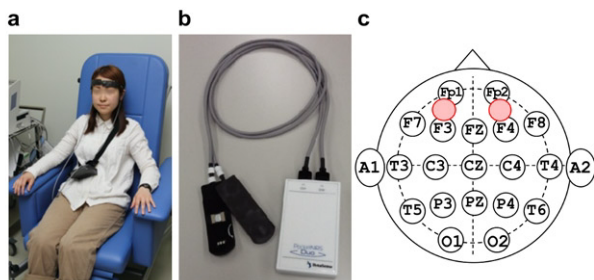
FIGURE 1. (a) Experimental setup; (b) Two-channel pocket NIRS; (c) International 10–20 system—red circles indicate the measurement areas.

In our experiments, each subject was seated in a comfortable chair in a dimly lit room (Fig. 1(a) and we measured the oxy- and deoxy-Hb concentration changes using a portable two-channel NIRS system (PNIRS-10, Hamamatsu Photonics K.K., Japan) (Fig. 1(b). The NIRS probes were set symmetrically on the forehead of each subject; the positioning is similar to the midpoint between the electrode positions Fp1/F3 (left) and Fp2/F4 (right) of the international 10-20 system, as demonstrated in Fig. 1(c) [6].

The sensor part (with weight approximately 100 g, resulting in only a very small burden on the subject) communicated with a PC via Bluetooth (class 2).

FIGURE 2. Experimental protocol observed.

Fig. 2 illustrates the experimental protocol observed. It consisted of the following steps. First, each subject completed the STAI questionnaire before NIRS measurements. Second, we calibrated the equipment. Third, the subject was told that the preparation period would begin. Fourth, the subject was instructed to rest quietly: rest period. This corresponded to the analysis period.

B. MACHINE LEARNING

In general, the machine learning paradigm comprises the following three ingredients:

  i) preprocessing
  ii) feature extraction
  iii) learning/prediction algorithm.

The nature of NIRS data is not well understood; consequently, the procedure that needs to be followed in each step is nontrivial. This paper reports on our attempts to carry out (ii) and (iii), with (i), which is no less important, being left for future work. Although the results are not included in this paper, we attempted to devise third and fifth order Butterworth bandpass filters (passband 0.01–1.0 Hz) and LOF outlier detection/deletion on the NIRS raw data for (i).

This preprocessing improved the performance only for two of the four datasets. We will report on the effects of preprocessing in a future paper; in this paper, our focus is on (ii) and (iii). The following subsections discuss (ii) and (iii) above.

C. FEATURE EXTRACTION

Using all the available features is information rich; however, less relevant features often degrade predictive performance. This is due to the fact that as the number of features increases, the number of unknown parameters also increases, such that poorly learned parameters degrade prediction capabilities.

Note that each channel of the target device acquires the concentration changes of oxy- and deoxy-Hb so that the acquired data is a four-dimensional vector. Fig. 3 shows typical NIRS data.
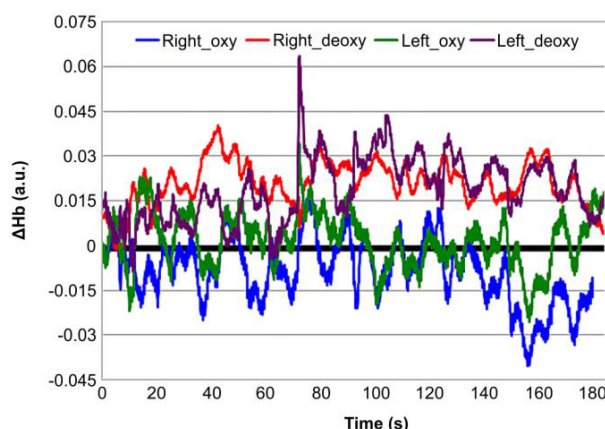

FIGURE 3. Typical NIRS data consisting of four quantities: right/left △ oxy and right/left △ deoxy. Note that the measured quantities from the device used in this study are the concentration changes △ oxy and △ deoxy, instead of absolute quantities.

In extracting appropriate features, we first noted that the measured quantities from the device used in the study were the concentration changes △ oxy and △ deoxy, instead of absolute quantities [27], [28] (see the appendix also). Therefore, the first order statistics, the mean values, should be avoided for prediction purposes. In the study, we used the following three Pearson correlation coefficients that were reported as effective in [26]:

  i) △ oxy (left)/△ deoxy (left)
  ii) △ oxy (left)/ △ deoxy (right)
  iii) △ deoxy (left)/△ deoxy (right)
across the four datasets.

The three quantities are illustrated in Fig. 4 in a schematic manner, with an arrow indicating Pearson correlation coefficient between the two associated quantities. Note that Pearson correlation coefficients are normalized in the range $[-1, +1]$. All the arguments in the ensuing section are based on these three second-order features.

Since the acquired NIRS data was four-dimensional, there are six Pearson correlation coefficients so that there were $\sum_{n=1}^{6} {}_6C_n = 63$ possibly combinations. By exhaustive search, we found [26] that the three-dimensional features
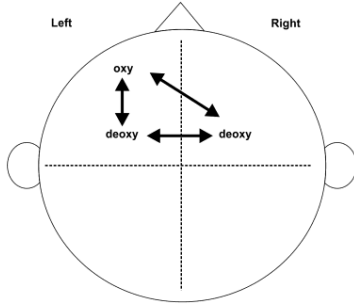
FIGURE 4. Schematic showing the three extracted features.

demonstrated in Fig. 4 gave good performance in terms of prediction errors.

D. ALGORITHM USED

Let $x^{(i)} := (x_1^{(i)}, \ldots, x_K^{(i)}) \in R^K, \ i = 1, \ldots, N$ be the NIRS feature vector of the *i-th* individual. For our selected features, $K = 3$, as described in Fig. 4. Let $y^{(i)}, \ i = 1, \ldots, N$ be the STAI state index obtained from the questionnaire, where N is the number of subjects in an experiment. Of the N available pieces of data, we reserved one $(x_{reserved}, \ y_{reserved})$ and attempted to fit the $N - 1$ remaining pieces of data with the following machine learning model:

$$P(y^{(i)}|x^{(i)}; \omega, \beta) = \sqrt{\frac{\beta}{2\pi}} \exp(-\frac{\beta}{2}(y^{(i)} - f(x^{(i)}; \omega))^2)$$
$$i = 1, \ldots, N - 1 \quad (1)$$

where f is a basis function for data fitting, and ω denotes the associated unknown parameter to be learned. β is another parameter to be learned that is often called the hyperparameter, and corresponds to the uncertainty level associated with the acquired STAI value. In order to capture the potential nonlinear relationship between the NIRS data and the STAI, we considered the following nonlinear basis function:

$$f(x^{(i)}; \omega) := \sum_{h=1}^{H} \left(\omega_{(K+1)h} \sigma \left(\sum_{k=1}^{K} \omega_{kh} x_k^{(i)} + \omega_{0h} \right) \right) + \omega_{0(K+1)}$$
$$(2)$$

where σ is a sigmoidal function. This basis function is nonlinear with respect not only to x but also ω, which makes the learning algorithm nontrivial. It is known that this basis function (perceptron) can approximate any nonlinear function with arbitrary precision [29]. As in [26], the unknown parameter vector ω is decomposed as $\omega = \{\{\omega_k\}_{k=1}^{K}, \ \omega_{(K+1)}, \ \omega_0, \ \omega_{0(K+1)}\}$ with $\omega_k := (\omega_{k1}, \ldots, \omega_{kH}) \in R^H, \omega_{(K+1)} := (\omega_{(K+1)1}, \ldots, \omega_{(K+1)H}) \in R^H, \ \omega_0 := (\omega_{01}, \ldots, \omega_{0H}) \in R^H$ and $\omega_{0(K+1)} \in R$. This decomposition amounts to the fact that each ω_k is the first-layer weights, associated with each feature $x_k^{(i)}$, $k = 1, \ldots, K$. $\omega_{(K+1)}$ are the second-layer weights, associated with hidden units. The remaining two parameters $\{\omega_0, \ \omega_{0(K+1)}\}$ were the bias parameters. The number of hidden units H can be estimated within the present Hierarchical Bayesian framework. This endeavor, however, requires

a more complicated algorithm and much more time to execute. From our own experiences, as well as those reported by other researchers, we used eight as the choice for H. Our study formulated the prediction problem within a Bayesian framework, where a prior distribution was assumed about the unknown parameters and that information incorporated into the data fitting model given by (1). The prior distribution for ω_k was assumed to be specified by

$$P(\omega_k|\alpha_k) = N\left(0, \left(\frac{1}{\alpha_k}\right)I\right) \quad (3)$$

which is the Gaussian distribution with a zero mean vector and covariance matrix $(1/\alpha_k)I; \alpha_k$ is the hyperparameter, associated with first-layer weights ω_k, and I denotes the identity matrix, $k = 1, \ldots, K$. This prior distribution often prevents overfitting. $\alpha_{(k+1)}$ is associated with second-layer weights $\omega_{(k+1)}$. There were two other hyperparameters $\{\alpha_0, \alpha_{0(K+1)}\}$, which were associated with the bias parameters $\{\omega_0, \omega_{0(K+1)}\}$. The prior distributions for β and α were assumed to follow the gamma distribution. Let

$$\alpha := \{\{\alpha_k\}_{k=1}^{K}, \alpha_{(K-1)}, \alpha_0, \alpha_{0(K-1)}\}, \ x := \{x^{(i)}\}_{i=1}^{N-1},$$
$$y := \{y^{(i)}\}_{i=1}^{N-1}. \quad (4)$$

The rationale behind the gamma distribution for prior is explained in Appendix C.

Assuming the data from each individual to be independent, the Bayes formula gives the posterior distribution:

$$P(\omega, \alpha, \beta|x, y)$$
$$= \frac{\prod_{i=1}^{N-1} P(y^{(i)}|x^{(i)}; \omega, \beta)P(\omega|\alpha, \beta)P(\alpha, \beta)}{\iiint \prod_{i=1}^{N-1} P(y^{(i)}|x^{(i)}; \omega, \beta)P(\omega|\alpha, \beta)P(\alpha, \beta)d\omega d\alpha d\beta} \quad (5)$$

Prior distribution for hyperparameter α and β are not explicitly written in (5), for the purpose of clarity.

Given the reserved NIRS data $x_{reserved}$, the prediction of the STAI index associated with the reserved data $y_{reserved}$ was performed by computing the predictive distribution:

$$P(y_{reserved}|x_{reserved}, x, y)$$
$$= \iiint P(y_{reserved}|x_{reserved}, \omega, \beta)P(\omega, \alpha, \beta|x, y)d\omega d\alpha d\beta. \quad (6)$$

E. PREDICTION CAPABILITY EVALUATION

As described in subsection IIID, one piece of data $(x_{reserved}, \ y_{reserved})$ out of N was reserved, whereas the remaining $N - 1$ pieces of data were used for learning the unknown parameters. After the parameters were learned, the reserved data were input to the algorithm for STAI prediction. Because one STAI value was also reserved, we were able to compute the difference between the predicted and the reserved STAI values, which can signify prediction error. Since hyperparameters (α, β) can also be learned via (5), the method is called the Hierarchical Bayesian algorithm.
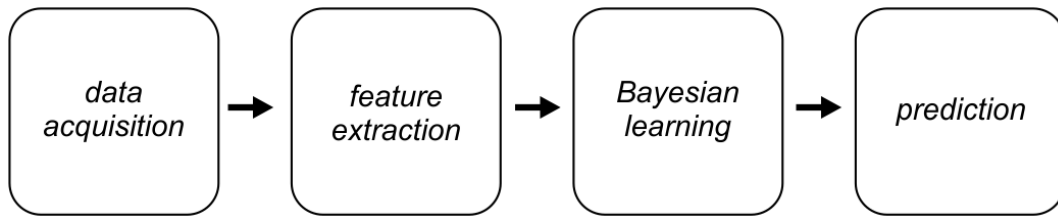
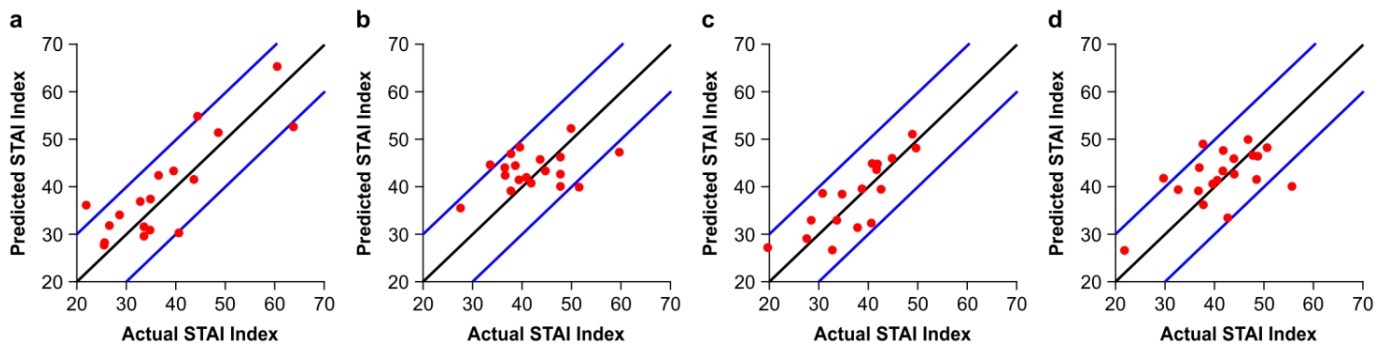**FIGURE 5.** Overall flow of activities carried out in the project.

FIGURE 6. Scatter plots of actual STAI and predicted STAI values with the proposed algorithm based on the nonlinear model: (a) dataset 1; (b) dataset 2; (c) dataset 3; (d) dataset 4.

There were four datasets, each comprising 17-20 sets of data. We performed four prediction experiments, one experiment for each of the four datasets. Within each dataset, one set of data was reserved for testing while the rest were used for learning. The reserved data were not used in the learning process. It should be noted that in the STAI prediction of each subject, the algorithm had never seen the data from the particular subject to be tested. It should also be noted that this is not leave-one-out cross validation. The Hierarchical Bayesian method used in this study does not use a validation dataset.

Only the state anxiety index was used in the present study, even though the trait anxiety index was available. We will attempt to make predictions of the trait anxiety index in our future project.

Fig. 5 demonstrates an overall flow of the prediction algorithm.

IV. EXPERIMENTAL RESULTS
In this section, we evaluate the prediction capabilities of the proposed algorithm.

A. PREDICTION CAPABILITY EVALUATION
Fig. 6 depicts scatter plots of the actual STAI values and the predicted values with the nonlinear model. Figs. 6(a)–(d) are the scatter plots corresponding to datasets 1 to 4. It should be noted that the figures are not the results of linear regression between the quantities of the horizontal and the vertical axes (see subsection IIID). The blue lines signify the RMSE of 10, which is an empirical value proposed

by one of the coauthors of this paper who is a clinician in neurosurgery in the sense that an RMSE of 10 for STAI could be tolerable in a clinical practice where the STAI score is between 20 and 80. We note that only a limited amount of data was available for STAI values near 20 and 80. As a result, the algorithm had very little chance of seeing very low/high STAI scores and it was difficult to make good predictions; however, the results appear to be reasonable, including the low/high STAI values.

Table 1 summarizes the results with multiple evaluation indices: RMSE, Pearson Correlation Coefficients, and the p-values of two tailed t-tests between the actual STAI values and the predicted STAI values. Note that datasets 1 and 2 were acquired from young subjects (19–26 years and 20–24 years, respectively) whereas datasets 3 and 4 were obtained from elderly subjects (61–79 years and 60–79 years, respectively). The effect of aging did not appear to be critical within this prediction framework.

The table also compares nonlinear and linear models. The nonlinear model of our proposed method gave better results in all cases. It can be seen that the linear predictions were much less accurate than those with the proposed nonlinear predictions. Significant prediction improvements are discernible with the nonlinear model. Table 2 summarizes the proportions of RMSE upper bounds for each dataset.

B. INTRA-SUBJECT-VARIATIONS
To assess the ''intra-subject'' variation of the proposed algorithm, we used the following multiple NIRS measurement data. Following acquisition of dataset 1, we asked the

TABLE 1. Prediction results. Performance indices are RMSE, pearson correlation, and the associated P-value of two tailed T-test between the actual STAI and the predicted values. It also compares linear and nonlinear models.

Dataset	RMSE		Correlation Coefficient		P Value	
	nonlinear	linear	nonlinear	linear	nonlinear	linear
1	**6.20**	9.99	**0.832**	0.447	**9.94E-06**	5.11E-02
2	**6.62**	8.52	**0.326**	-0.610	**1.73E-01**	9.83E-01
3	**4.50**	7.59	**0.821**	0.392	**5.35E-05**	4.74E-02
4	**6.38**	6.80	**0.565**	0.485	**9.44E-03**	1.89E-02

For each dataset and each performance criterion, there are two numbers; nonlinear and linear model. Of these two numbers, better results are in bold type. The proposed method based on the nonlinear model shows better results in all cases.

TABLE 2. Summary of the proportions of RMSE for differing upper bounds with the proposed algorithm based on the nonlinear model.

Dataset	Proportion of dataset (%)		
	RMSE < 5	RMSE < 10	RMSE min
1	73.7	84.2	1.5300
2	42.1	84.2	0.9070
3	70.6	100.0	0.4490
4	60.0	85.0	0.0216

The rightmost column gives the minimum RMSE for each dataset.

19 subjects to repeat three more trials subsequent on the trial described in IIIA. There was a 3-5 min resting period between two trials. The Bayesian learning was conducted using the (18 out of 19) first set of trial data, which was in fact the same as those used in IIIA and tested against the remaining data. Using the parameters learned from the first set of trial data, tests were performed against the second, third, and fourth set of trial data.
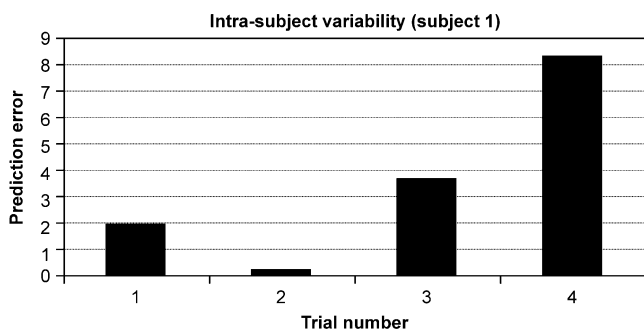

FIGURE 7. Intra-subject variability of the prediction capability of subject 1 over four trials. Learning was performed with the data from the first set of trial data for 18 subjects other than subject 1, whereas prediction capability was tested against the withheld data for subject 1. With the same learned data, tests were performed on the second, third, and fourth sets of data from subject 1. In 10 of the 19 subjects, the prediction error of the fourth trial was greater than that of the first trial.

Fig. 7 demonstrates the prediction error variations over the four trials of subject 1.

The prediction error for the fourth trial was greater than that for the first trial in 10 subjects out of 19. However, in nine subjects, the prediction error for the fourth trial was less than that for the first trial. Fig. 8 shows the results for subject 10.
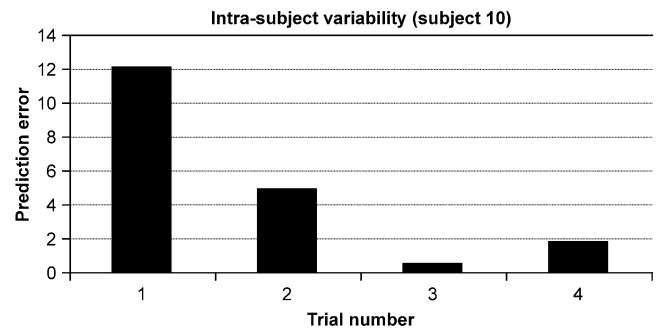

FIGURE 8. Intra-subject variability of the prediction capability of subject 10 over four trials. Learning and prediction are similar to those in Fig. 7. In nine of the 19 subjects, the prediction error for the fourth trial was less than that for the first trial.

The overall average intra-subject prediction error was 7.638, which is greater than those listed in Table 1.

V. DISCUSSION

A. THE BRAIN AT REST

The brain at rest has been an active research area and numerous activities with distinctive patterns have been reported [8]–[14]. Our study considered the NIRS data in the resting condition. The method used in the study did not need task design, which made experimentation easier. With NIRS measurements for stress task [3]–[7], a subject's NIRS data started changing when the subject was told that the task would begin in 30 s, even before the actual stress task began. In the resting condition assumed in our study, that effect was less than in the stress task; however, we could not exclude the potential effect of the information on the subject when the subject was informed of the beginning of the measurement 30 s before the measurements began.

B. BEYOND OXY HEMOGLOBIN ASYMMETRY

The present study went beyond investigating the left/right asymmetry of oxy-Hb concentrations in the PFC activity

during tasks involving mental stress [3]–[7] such that the STAI index predictions were performed based on the four-dimensional NIRS values instead of the mere oxy-Hb concentration changes. The three-dimensional features were extracted from the oxy- and deoxy-Hb concentrations changes acquired from a two-channel NIRS device placed on the PFC.

C. PREDICTIONS ON FOUR DIFFERENT DATASETS

Note that the prediction experiments were performed on four different datasets that were acquired from different individuals on different occasions. Since the STAI values varied between 20 and 80, the RMSE values 6.20, 6.62, 4.50, and 6.38 appear reasonable.

Table 2 summarizes the proportions of the RMSE values that were less than 10 as well as those less than five in each dataset. It also gives the minimum prediction errors as well. There are three facts that should be mentioned here:

   i) The algorithm had never seen the NIRS values as well as the associated STAI of a particular individual before. The parameters were learned from the data of other subjects only, suggesting that the STAI values of unknown subjects may be predictable once the machine learned.

   ii) There were several subjects whose STAI values could be predicted by the proposed algorithm with reasonable accuracy. Table 2, for instance, shows that there were 84.2–100% of the young as well as elderly subjects whose STAI prediction errors were less than 10, which appears reasonable since the STAI varied between 20 and 80. The rightmost column of Table 2 gives the minimum RMSE for each dataset.

   iii) In machine learning prediction problems, it is often critical to choose appropriate features. Our study employed the three features depicted in Fig. 4 throughout the four datasets, irrespective of age.

D. COMPARISON WITH LINEAR MODEL

Recall that our basis function (2) is nonlinear not only with respect to feature vector x, but also with respect to unknown parameters. We tested a linear model instead of the nonlinear model and found that the linear model was much less accurate than the proposed nonlinear model, as summarized in Table 1.

E. AGE GROUPS

Two groups comprised young subjects (19–26 years and 20–24 years, respectively), whereas the remaining two comprised elderly subjects (61–79 years and 60–79 years, respectively). The results in Table 1 appear to indicate that age does not play any significant role in the STAI prediction problem.

F. GENDER DIFFERENCE

We note that a gender difference is discernible in that the highest RMSE was 6.62, followed by 6.38 in datasets 2 and 4,

respectively, where the result for the female subjects dominates that for the male subjects.

It is not clear whether this is a limitation of the model or not. Potential differences in the results due to gender difference are not surprising because it is a known fact that gender differences exist in brain functions. In language processing tasks, for example, the right PFC is often activated in male subjects, whereas both the right and left PFC are activated in female subjects [30].

G. INTRA-SUBJECT PREDICTION ERRORS

Fig.7 appears to suggest that the subject could have started becoming tired as the trial proceeded. However, there were other cases where the prediction error decreased instead of increased over the four trials, as was shown in Fig.8. It is difficult at this point to draw any conclusion.

H. PORTABLE NIRS DEVICE

On comparing our experiences with EEG-based experiments, both portable and non-portable, we determined that the two-channel portable NIRS device is easier to equip than EEG devices. This is mainly because when conducting EEG measurements, one needs to carefully reduce the contact impedance between the head and the probe, whereas there is no such intricacy associated with the NIRS device used in our study. The burden on the part of the subjects was also light compared with EEG measurements because the subjects could blink and carry out minor movements, whereas blinking and head movement significantly distort EEG signals. The fact that the NIRS device was portable and lightweight also helped subjects to participate in the experiments with relative ease.

I. POTENTIAL APPLICATIONS

The results of our prediction experiments, which we report in Section IV, appear to indicate that the proposed algorithm is functional at least in the target datasets because the average prediction error is relatively small. Further, the prediction capabilities do not appear to depend on age. The device is portable, weighs 100 g, and is easy to attach; therefore, there could be potential applications in practical situations.

One of the coauthors of this study, who is a neurosurgeon and routinely examines the STAI index of patients before major operations, found that it helps when reasonable anxiety is predictable with this kind of simple device without a written self-report.

The proposed method can be utilized in cases where patients are weak, in pain, or in acute anxiety states such that it is difficult for them to complete written questionnaires.

One of the coauthors of this paper is a neurosurgery clinician and routinely acquires STAI questionnaires from patients prior to major operations. Sometimes, however, a patient is not able to answer the questionnaire in a credible manner. If the prediction method reported in this paper is functional, it could help in such scenarios.

J. LIMITATIONS

Our predictions are based on a Bayesian machine learning algorithm using the features extracted in terms of prediction errors. The fact that our predictions were reasonable means that the acquired NIRS data contained information about the STAI index via a complicated nonlinear manner provided appropriate features are found. One of the limitations of our study is that physiological mechanisms were not investigated. For instance, deciphering why the three selected features given in Fig. 4 are important is still an outstanding research question. This endeavor, however, necessitates a significant amount of effort and will therefore be left for future research.

Another limitation is that NIRS measures the blood oxygen changes within illuminated pathways such that measured values of hemoglobin changes are also affected by extracranial tissues, not only intracranial tissues. This means that the NIRS measurements could contain multiple sources of hemoglobin changes [11]. In addition, it was suggested that the experiment itself could affect PFC activity at rest since the subject might feel the strain during NIRS measurements, depending on their anxiety levels.

K. FUTURE PROJECTS

The following possible future research projects are under consideration:

   i) It would be interesting to study if the capability of the proposed prediction algorithm depends on gender and handedness, among other characteristics.
   ii) It would also be worth investigating the prediction capability of the STAI trait index, which is the feelings of stress, worry, discomfort, etc. that one experiences on a daily basis, instead of the STAI state index, which is the feeling solely at the time of the experiment.
   iii) NIRS measurements are the building blocks for the proposed algorithm; therefore, we need to eliminate possible uncertainties and examine its prediction capabilities. For instance, a project worth pursuing would be to consider a particular subject performing the experiment over several time intervals, e.g., one or several weeks or several months, and examine the prediction capabilities of the proposed method. Another is to examine how the measurement environment might affect the prediction capabilities. Environmental factors worth considering include room temperature, humidity, lighting, time of a day, among others.
   iv) An advantage of the present method is that it is not restricted to the STAI data and the NIRS data, as long as targeted written test data and associated sensor data are available.
   v) Providing experimental data considering both cases, i.e., answering the questions before or after acquisition, could reveal useful information about NIRS measurement sensitivity at resting protocols.

VI. CONCLUSION

In this paper, we reported on a study in which we attempted to make predictions of the anxiety index of subjects from data acquired by a two-channel NIRS device placed on the PFC. The method utilized comprised several steps:

   i) data acquisition
   ii) feature extraction
   iii) construction of machine learning algorithm
   iv) prediction

Data were acquired from four different groups—two comprising young subjects and the remaining two comprising elderly subjects. Care had to be taken in constructing the machine learning algorithm because the amount of data acquired was highly limited to 17–20 subjects in each dataset, in addition to the potentially significant amount of uncertainties contained. The proposed algorithm was formulated within a Bayesian framework and MCMC. Prediction errors were 4.50–6.62 among the four datasets. Since the STAI score varies between 20 and 80, the prediction accuracies appeared reasonably accurate. There appeared no significant difference in prediction capabilities between the young and the elderly subjects. The device is portable, weighs only 100 g, and is easy to position; therefore, the method can be used in practical situations.

In one of our ongoing research projects, a new algorithm for improving prediction accuracies is being developed. We will report on this in the future.

APPENDIX A
MEASUREMENT PRINCIPLE OF THE TARGET NIRS DEVICE

Let us now briefly look at the measurement principle of the target device. First, assume that one is interested in a single substance. Let $\varepsilon(\lambda)$, $C(\lambda, t)$, and $L(\lambda)$ denote absorption concentration coefficient ($\mu M^{-1} cm^{-1}$), concentration of the absorbent (mL) at time t, and the optical path length (mm), respectively, at wavelength λ. Then, the modified Beer-Lambert law [28], [30], [31] asserts that the optical density $OD(\lambda, t)$ of the quantity of interest is given by

$$OD(\lambda, t) = \varepsilon(\lambda)C(\lambda, t)L(\lambda) + OD(\lambda, t)_R \qquad (7)$$

where the last term represents the term that cannot be detected by the sensor because of photon scattering. Consider the change defined by

$$\Delta OD(\lambda, t, t_0) := OD(\lambda, t) - OD(\lambda, t_0) \qquad (8)$$

where t_0 is the reference time. If one assumes that $OD(\lambda, t)_R$ is approximately constant over the target time interval, then this term is removed from (8). Consequently, given the measurement $\Delta OD(\lambda, t, t_0)$, one can compute

$$(C(\lambda, t) - C(\lambda, t_0))L(\lambda) = \frac{\varepsilon(\lambda)}{\Delta OD(\lambda, t, t_0)} \qquad (9)$$

provided that $\varepsilon(\lambda)$ is available. Therefore, the measured quantity is the change instead of the absolute quantity in an arbitrary unit. Since there are more than one quantity of interest, i.e., oxy- and deoxy-Hb, the equipment utilized

in our study used equations similar to (8) at several differing wavelengths and computed a two-dimensional vector of changes.

APPENDIX B
MARKOV CHAIN MONTE CARLO (MCMC)

Since analytical closed form equations are not available for equations (5) and (6), an approximation method is needed. MCMC constructs a Markov Chain over the parameter space and regards the resulting Markov Chain samples as the samples from the posterior and predictive distributions.

Theoretical results of MCMC guarantee convergence of the MCMC samples to the target distribution, and methods that check how close the samples are to their target are available. These convergence check methods are themselves ongoing vital research projects [32]; however, it appears that no comprehensive method currently exists. In practical situations, one often decides when to end the burn in phase and how many samples to draw after the burn in, from experience. In our study, the burn in period was 3000, after which 3000 samples were drawn.

APPENDIX C
PRIOR DISTRIBUTIONS

The prior distribution for the parameters specified by (3) discourages the parameters becoming large, thereby preventing overfitting, which is reasonable, and has been successfully used in many practical problems. In order to explain prior distributions for hyperparameters, let α_i be one of the components of the hyperparameter vector defined in (4). Since this hyperparameter appears in (3) and is the inverse of the variance of a Gaussian distribution, ω_i controls how ω_i is scattered. There are at least three desired properties for α_i:

(a) it should be positive,
(b) it should have a large range for variations,
(c) it should be tractable for drawing posterior samples.

These are fairly standard requirements in many practical problems, including the present study. One such candidate is Beta distribution:

$$P(\alpha_i) := \frac{(\psi_i/2\kappa_i)^{\psi_i/2}}{\Gamma(\psi_i/2)}\alpha^{(\psi_i/2-1)}e^{-\alpha_i(\psi_i/2\kappa_i)},$$

where κ_i and ψ_i are the width and the shape parameters. In addition to (a) and (b) above, this distribution has one additional desirable property—it is equipped with *natural conjugacy* [33]. Specifically, the posterior distribution belongs to the same distribution family as the prior distribution, which enables easier posterior sampling. It should be noted that even though the hyperparameter prior distributions are independent, their posterior distributions are generally dependent on each other. Theoretically, κ_i and ψ_i can be further learned with another hierarchy; however, it is often sufficient to assume empirical fixed values in many practical problems. In this study, $\kappa_i = 0.02$ and $\psi_i = 0.01$, for all i, were assumed. For hyperparameter β, which appears in (1), $\kappa_\beta = 60$ and

$\psi_\beta = 720$ were used. These are based on our previous experiences with other data.

REFERENCES

[1] K. Sakatani *et al.*, "Effects of occlusal disharmony on working memory performance and prefrontal cortex activity induced by working memory tasks measured by NIRS," *Adv. Experim. Med. Biol.*, vol. 765, pp. 239–244, 2013.

[2] A. Amemiya, T. Takeda, K. Nakajima, K. Ishigami, T. Tsujii, and K. Sakatani, "Effects of experimentally deviated mandibular position on stress response," *Adv. Experim. Med. Biol.*, vol. 765, pp. 1–7, 2013.

[3] K. Sakatani, "Optical diagnosis of mental stress: Review," *Adv. Experim. Med. Biol.*, vol. 737, pp. 89–95, 2012.

[4] K. Sakatani, M. Tanida, and M. Katsuyama, "Effects of aging on activity of the prefrontal cortex and autonomic nervous system during mental stress task," *Adv. Experim. Med. Biol.*, vol. 662, pp. 473–478, 2010.

[5] M. Tanida, M. Katsuyama, and K. Sakatani, "Effects of fragrance administration on stress-induced prefrontal cortex activity and sebum secretion in the facial skin," *Neurosci. Lett.*, vol. 432, no. 2, pp. 157–161, Feb. 2008.

[6] M. Tanida, M. Katsuyama, and K. Sakatani, "Relation between mental stress-induced prefrontal cortex activity and skin conditions: A near-infrared spectroscopy study," *Brain Res.*, vol. 1184, pp. 210–216, Dec. 2007.

[7] M. Tanida, K. Sakatani, R. Takano, and K. Tagai, "Relation between asymmetry of prefrontal cortex activities and the autonomic nervous system during a mental arithmetic task: Near infrared spectroscopy study," *Neurosci. Lett.*, vol. 369, no. 1, pp. 69–74, Oct. 2004.

[8] W. Ishikawa *et al.*, "New method of analyzing NIRS data from prefrontal cortex at rest," *Adv. Experim. Med. Biol.*, vol. 789, pp. 391–397, 2013.

[9] Y. Hoshi and M. Tamura, "Fluctuations in the cerebral oxygenation state during the resting period in functional mapping studies of the human brain," *Med. Biol. Eng. Comput.*, vol. 35, no. 4, pp. 328–330, Jul. 1997.

[10] M. L. Pierro, A. Sassaroli, P. R. Bergethon, B. L. Ehrenberg, and S. Fantini, "Phase-amplitude investigation of spontaneous low-frequency oscillations of cerebral hemodynamics with near-infrared spectroscopy: A sleep study in human subjects," *NeuroImage*, vol. 63, no. 3, pp. 1571–1584, Nov. 2012.

[11] L. Minati, I. U. Kress, E. Visani, N. Medford, and H. D. Critchley, "Intra- and extra-cranial effects of transient blood pressure changes on brain near-infrared spectroscopy (NIRS) measurements," *J. Neurosci. Methods*, vol. 197, no. 2, pp. 283–288, Apr. 2011.

[12] Y. Tong, L. M. Hocke, S. C. Licata, and B. D. Frederick, "Low-frequency oscillations measured in the periphery with near-infrared spectroscopy are strongly correlated with blood oxygen level-dependent functional magnetic resonance imaging signals," *J. Biomed. Opt.*, vol. 17, no. 10, pp. 1060041–10600410, Oct. 2012.

[13] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI," *Magn. Reson. Med.*, vol. 34, no. 4, pp. 537–541, Oct. 1995.

[14] A. Anand, Y. Li, Y. Wang, M. J. Lowe, and M. Dzemidzic, "Resting state corticolimbic connectivity abnormalities in unmedicated bipolar disorder and unipolar depression," *Psychiatry Res.*, vol. 171, no. 3, pp. 189–198, Mar. 2009.

[15] A. Villringer, J. Planck, C. Hock, L. Schleinkofer, and U. Dirnagl, "Near infrared spectroscopy (NIRS): A new tool to study hemodynamic changes during activation of brain function in human adults," *Neurosci. Lett.*, vol. 154, no. 2, pp. 101–104, May 1993.

[16] Y. Hoshi, N. Kobayashi, and M. Tamura, "Interpretation of near-infrared spectroscopy signals: A study with a newly developed perfused rat brain model," *J. Appl. Physiol.*, vol. 90, no. 5, pp. 1657–1662, May 2001.

[17] O. Pryds, G. E. Andersen, and B. Friis-Hansen, "Cerebral blood flow reactivity in spontaneously breathing, preterm infants shortly after birth," *Acta Paediatrica*, vol. 79, no. 4, pp. 391–396, Apr. 1990.

[18] M. Ferrari, D. A. Wilson, D. F. Hanley, and R. J. Traystman, "Effects of graded hypotension on cerebral blood flow, blood volume, and mean transit time in dogs," *Amer. J. Physiol., Heart Circulatory Physiol.*, vol. 262, no. 6, pp. 1908–1914, Jun. 1992.

[19] K. Sakatani, M. Ohtaki, M. Kashiwasake, and K. Hashi, "Effects of hyperventilation and CO_2 inhalation on cerebral oxygen metabolism of moyamoya disease measured by near-infrared spectroscopy," in *Neurochemical Monitoring in the Intensive Care Unit*. Tokyo, Japan: Springer-Verlag, 1995, pp. 226–234.

[20] L. Rovati, G. Salvatori, L. Bulf, and S. Fonda, "Optical and electrical recording of neural activity evoked by graded contrast visual stimulus," *BioMed. Eng. Online*, vol. 6, p. 28, Jul. 2007.

[21] Y. Hoshi, S. Kosaka, Y. Xie, S. Kohri, and M. Tamura, "Relationship between fluctuations in the cerebral hemoglobin oxygenation state and neuronal activity under resting conditions in man," *Neurosci. Lett.*, vol. 245, no. 3, pp. 147–150, Apr. 1998.

[22] M. Butti *et al.*, "Multimodal analysis of a sustained attention protocol: Continuous performance test assessed with near infrared spectroscopy and EEG," in *Proc. 28th Annu. Int. Conf. IEEE EMBS*, Aug./Sep. 2006, pp. 1040–1043.

[23] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fMRI signal," *Nature*, vol. 412, no. 6843, pp. 150–157, Jul. 2001.

[24] E. G. Benotsch, S. K. Lutgendorf, D. Watson, L. J. Fick, and E. V. Lang, "Rapid anxiety assessment in medical patients: Evidence for the validity of verbal anxiety ratings," *Ann. Behavioral Med.*, vol. 22, no. 3, pp. 199–203, 2000.

[25] C. H. Kindler, C. Harms, F. Amsler, T. Ihde-Scholl, and D. Scheidegger, "The visual analog scale allows effective measurement of preoperative anxiety and detection of patients' anesthetic concerns," *Anesthesia Analgesia*, vol. 90, no. 3, pp. 706–712, Mar. 2000.

[26] Y. Fukuda, W. Ishikawa, R. Kanayama, T. Matsumoto, N. Takemura, and K. Sakatani, "Bayesian prediction of anxiety level in aged people at rest using 2-channel NIRS data from prefrontal cortex," presented at the 41st Meeting Int. Soc. Oxygen Transp. Tissue, Dartmouth, NH, USA, pp. 303–308, Jun. 2013.

[27] K. Morinaga *et al.*, "Anticipatory anxiety-induced changes in human lateral prefrontal cortex activity," *Biol. Psychol.*, vol. 74, no. 1, pp. 34–38, Jan. 2007.

[28] D. T. Delpy, M. Cope, P. van der Zee, S. Arridge, S. Wray, and J. Wyatt, "Estimation of optical pathlength through tissue from direct time of flight measurement," *Phys. Med. Biol.*, vol. 33, no. 12, pp. 1433–1442, 1988.

[29] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer-Verlag, 2006.

[30] M. Hirnstein, R. Westerhausen, M. S. Korsnes, and K. Hugdahl, "Sex differences in language asymmetry are age-dependent and small: A large-scale, consonant–vowel dichotic listening study with behavioral and fMRI data," *Cortex*, vol. 49, no. 7, pp. 1910–1921, Jul./Aug. 2013.

[31] S. Angelo and F. Sergio, "Comment on the modified Beer–Lambert law for scattering media," *Phys. Med. Biol.*, vol. 49, no. 14, pp. 255–257, Jul. 2004.

[32] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. London, U.K.: Chapman & Hall, 1996.

[33] R. Neal, *Bayesian Learning for Neural Networks*. New York, NY, USA: Springer-Verlag, 1996.

YUKIKATSU FUKUDA received the B.E. degree from the Department of Electrical Engineering and Bioscience, Waseda University, Tokyo, Japan, in 2012, where he is currently pursuing the M.S. degree.

His current research interests include Bayesian machine learning for prediction problems with Monte Carlo implementations using near-infrared spectroscopy data.

YASUTOSHI IDA received the B.E. degree from the Department of Electrical Engineering and Bioscience, Waseda University, Tokyo, Japan, in 2012, where he is currently pursuing the M.S. degree.

His current research interests include sentiment analysis and relational model by Bayesian method. He is working on handling relational data with side-information via Bayesian model.

TAKASHI MATSUMOTO (M'71–SM'83–F'86–LF'10) received the B.S. degree from the Department of Electrical Engineering, Waseda University, Tokyo, Japan, the M.S. degree from the Department of Applied Mathematics, Harvard University, Cambridge, MA, USA, and the Ph.D. degree from the Department of Electrical Engineering, Waseda University.

He is currently a Professor with the Faculty of Advanced Science and Engineering, Waseda University. He held visiting positions with the University of California at Berkeley, Berkeley, CA, USA (1977–1979), and Cambridge University, Cambridge, U.K. (2003–2004). His research interests are Bayesian machine learning with Monte Carlo implementations and Bootstrap method, with target data, including biological data, image data, and time series data. He also has interests in nonparametric priors for Bayesian learning.

Prof. Matsumoto was a recipient of the 1994 JNNS Best Paper Award. He is a member of the Japanese Neural Network Society.
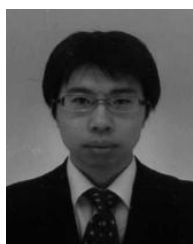
NAOHIRO TAKEMURA received the B.Sc. degree in engineering and the M.Sc. degree in informatics from Kyoto University, Kyoto, Japan, in 2003 and 2005, respectively.

He is currently a Researcher with the College of Engineering, Nihon University, Tokyo, Japan. His research interests include neuroscience and biomedical engineering.

KAORU SAKATANI received the M.D. degree from Osaka Medical College, Osaka, Japan, in 1981, the Ph.D. degree in medicine from the Graduate School of Osaka Medical College, Takatsuki, Japan, in 1987, and the Ph.D. degree in engineering from the Graduate School of Hokkaido University, Sapporo, Japan, in 1998.

He is currently a Professor with the School of Medicine, College of Engineering, Nihon University, Tokyo, Japan. He is also a Board-Certified Neurosurgeon in Japan. His research interests include biomedical engineering, optical engineering, neuroimaging, and neuroscience.

Dr. Sakatani was a recipient of awards, such as the Medical Research Encouragement Prize of the Japan Medical Association (2010) and the Asian Computer Tomography Award (1998). He is the Vice President of the Japan Optical Brain Functional Imaging, and an executive member of the Japanese Society of Integrative Medicine, the Japanese Society of Neuromonitoring, the Japanese Society of Neurosurgery, the Japanese Society of Stroke, and the Japanese Society of Cerebral Blood Flow and Metabolism.