PAPER

Error Control for Performance Improvement of Brain-Computer Interface: Reliability-Based Automatic Repeat Request

Hiromu TAKAHASHI[†a)], ***Student Member***, **Tomohiro YOSHIKAWA**[†], ***Nonmember***, ***and*** **Takeshi FURUHASHI**[†], ***Member***

SUMMARY Brain-Computer Interfaces (BCIs) are systems that translate one's thoughts into commands to restore control and communication to severely paralyzed people, and they are also appealing to healthy people. One of the challenges is to improve the performance of BCIs, often measured by the accuracy and the trial duration, or the information transfer rate (ITR), i.e., the mutual information per unit time. Since BCIs are communications between a user and a system, error control schemes such as forward error correction and automatic repeat request (ARQ) can be applied to BCIs to improve the accuracy. This paper presents reliability-based ARQ (RB-ARQ), a variation of ARQ designed for BCIs, which employs the maximum posterior probability for the repeat decision. The current results show that RB-ARQ is more effective than the conventional methods, i.e., better accuracy when trial duration was the same, and shorter trial duration when the accuracy was the same. This resulted in a greater information transfer rate and a greater utility, which is a more practical performance measure in the P300 speller task. The results also show that such users who achieve a poor accuracy for some reason can benefit the most from RB-ARQ, which could make BCIs more universal.

key words: *brain-computer interfaces, information transfer rate, error control, automatic repeat request*

1. Introduction

Brain-Computer Interfaces (BCIs) are promising technologies that translate one's thoughts into commands to restore control and communication to severely paralyzed people such as those with amyotrophic lateral sclerosis (ALS), and also appealing to healthy people. Among various ways to record brain activities, Electroencephalogram (EEG) is said to be the most practical way due to its non-invasiveness and relatively small cost [1]. In fact, EEG-based BCIs have been considerably researched [2]–[4]. The typical performance measures for BCIs are the accuracy and the speed, and there is a trade-off between them in general. Hence, the information transfer rate (ITR), the amount of information transferred from a user to a BCI system per unit time, is often employed alternatively. Much research aiming to improve the accuracy have been reported. For instance, adaptive learning is to re-train a classifier to adapt to changing features of EEGs [5], [6]; feedback training is to feed some information, e.g., classification results, back to users and help them adjust their EEGs [7]. By contrast, the most straightforward way to improve the accuracy is to lengthen the duration per classification, e.g., the more EEG samples are averaged, the greater the accuracy becomes. Nonetheless, the accuracy and the speed need to be balanced to maximize the ITR [8], [9].

In the field of data transmission, there are some strategies against errors in noisy channels. For instance, forward error correction (FEC) allows the receiver to detect and correct errors, automatic repeat request (ARQ) asks the sender to re-transmit, and Hybrid ARQ is a combination of FEC and ARQ. These strategies do not decrease errors per se, but correct or detect them to recover corrupted data. If the transmission is between a human and a computer, other types of error controls are normally used; for example, confirmation is to ask users if their inputs are correctly interpreted [10]. In fact, the confirmation has been long used for BCIs [11], [12]. Additionally, the error-related potentials (ErrPs), occurring on erroneous feedbacks, have been reported to be usable for error detection in BCIs [13]. This ErrP-based error detection requires a few hundred milli-seconds to see if the ErrPs have occurred, which is much shorter than confirmation requires.

In the past several years, reliability-based hybrid ARQ (RB-HARQ) has been proposed [14]. This is a variant of hybrid ARQ, in which the re-transmissions are requested based on the reliability of transmitted data. It has been reported that RB-HARQ can provide performance close to the channel capacity. This paper presents reliability-based ARQ (RB-ARQ), which uses the reliability, or the maximum posterior probability, as the criterion for the request like RB-HARQ. It is worth noting that the proposed method does not include FEC since humans, i.e., the senders in BCIs, can hardly handle complicated error control codings. The proposed method needs only a fraction of time in the repeat decision, and is very simple because users only have to keep doing the required task until being told to stop. Although there exist some other error reduction methods such as the rejection rule for the mental imagery task [15] and the dynamic subtrial limiting for the P300 speller task [9], [16], this paper presents that the proposed method effectively works for both the P300 speller and the mental imagery task. Moreover, this study is the first attempt to explicitly apply the ARQ strategy to BCIs, and one of the advantages of the idea to apply error controls developed in the data transmission area to BCIs is that various existing methods could also possibly be applied to BCIs, e.g., a simple FEC might be applicable and more effective.

Manuscript received June 22, 2010.

Manuscript revised March 4, 2011.

[†]The authors are with the Department of Computational Science and Engineering, Graduate School of Engineering, Nagoya University, Nagoya-shi, 464–8603 Japan.

a) E-mail: takahashi@cmplx.cse.nagoya-u.ac.jp

DOI: 10.1587/transinf.E94.D.1243

2. Methods

2.1 Proposed Method

There exists three types of ARQ; and the stop-and-wait ARQ is the simplest one [17], where the receiver sends back an ACK (Acknowledge) or a NAK (Not Acknowledge) frame to ask the sender to transmit the next frame or the same frame, respectively (see Fig. 1). It is normally necessary to send extra information in order for the receiver to detect errors. RB-ARQ, however, utilizes solely the reliability of data as criterion for the request decision not to increase the burden on the sender, i.e., the user in BCIs. Moreover, the sender keeps sending the same data until receiving an ACK in the proposed method (see Fig. 2), as it would not be wise to disturb the user by telling a NAK.

Let $\mathcal{K}$ be a set of possible states, e.g., motor imagery of left or right hand, and $\boldsymbol{x}$ be the corresponding EEGs, belonging to $u \in \mathcal{K}$. The *Bayes classifier* classifies $\boldsymbol{x}$ as such $k \in \mathcal{K}$ that the posterior probability $P(k|\boldsymbol{x})$ is maximum:

$$\hat{u} = \underset{k\in\mathcal{K}}{\operatorname{argmax}}\, P(k|\boldsymbol{x}), \tag{1}$$

where $\hat{u}$ denotes a predicted label [18]. The maximum of the posterior probability, which can be seen in Eq. (1), is equivalent to the probability of correct classification; thus, it can be regarded as the reliability of data. Given that $\boldsymbol{x}_i^n$ can be obtained at time n in the ith of N_{trial} trial and belongs to u_i, each of which is independent of each other, and let $\mathcal{X}_i^N = \{\boldsymbol{x}_i^1, \boldsymbol{x}_i^2, \ldots, \boldsymbol{x}_i^N\}$ be a set of data at time N, then the reliability λ_i^N can be defined and calculated as follows:

$$\begin{aligned}\lambda_i^N &\equiv \max_{k\in\mathcal{K}} P(k|\mathcal{X}_i^N),\\ &= \max_{k\in\mathcal{K}} \frac{P(k)\prod_n P(\boldsymbol{x}_i^n|k)}{\sum_{l\in\mathcal{K}} P(l)\prod_n P(\boldsymbol{x}_i^n|l)},\\ &= \max_{k\in\mathcal{K}} \frac{P(k)\prod_n \frac{P(k|\boldsymbol{x}_i^n)}{P(k)}}{\sum_{l\in\mathcal{K}} P(l)\prod_n \frac{P(l|\boldsymbol{x}_i^n)}{P(l)}}.\end{aligned} \tag{2}$$

If each class has the same prior probability, the reliability can be expressed as

$$\lambda_i^N = \max_{k\in\mathcal{K}} \frac{\prod_n P(k|\boldsymbol{x}_i^n)}{\sum_{l\in\mathcal{K}}\prod_n P(l|\boldsymbol{x}_i^n)}. \tag{3}$$

Consequently, $P(k|\mathcal{X}_i^N)$ can be calculated using each $P(k|\boldsymbol{x}_i^n)$ and the prior probabilities. Note that $P(k|\mathcal{X}_i^N)$ is equivalent to $P(k|\bar{\boldsymbol{x}}_i^N)$ when normal distributions are assumed, where $\bar{\boldsymbol{x}}_i^N$ denotes $\frac{1}{N}\sum_n \boldsymbol{x}_i^n$. In RB-ARQ, a user keeps thinking the same thought, e.g., in mental tasks, or keeps being exposed to stimuli, e.g., in the P300 speller, unless the following condition is satisfied:

$$\lambda_i^N > \lambda, \tag{4}$$

where λ is a given threshold. On the other hand, the rejection rule and the standard averaging employ the stopping condition, $\max_{k\in\mathcal{K}} P(k|\boldsymbol{x}_i^N) > \lambda$, and $N = N_{\text{target}}$, respectively, where N_{target} denotes the number of samples to be averaged (see Table 1). Let N_i and T_s be the time point when the stopping criterion is satisfied and the duration of EEG data needed to obtain $\boldsymbol{x}$, respectively. A trial duration d is defined as the average length of time per class estimation in this paper; thus, $d = \frac{1}{N_{\text{trial}}}\sum_i^{N_{\text{trial}}} T_i = T_s \times \frac{1}{N_{\text{trial}}}\sum_i^{N_{\text{trial}}} N_i$, and $d = \frac{1}{N_{\text{trial}}}\sum_i^{N_{\text{trial}}} T_i + T_r$ when the resting duration is taken into considered. An algorithmic description of the proposed method at the ith trial is shown in Algorithm 1.

The authors have proposed two methods, which apply the binary erasure model to BCIs [19], which are essentially equivalent to the rejection rule and RB-ARQ, respectively. However, the methods presented in [19] are applicable neither to multi-class problems nor the support vector machine (SVM) since they are based on two-class normal distributions. On the other hand, RB-ARQ employs the maximum posterior probability as the reliability; thus, it can be applied to both multi-class problems [20] and the SVM [21].

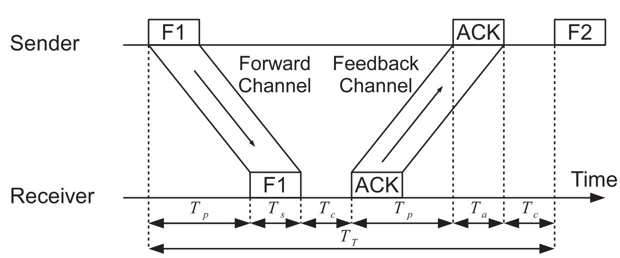


Fig. 1 Stop-and-wait ARQ. The receiver sends back a NAK on detecting an error, otherwise it sends an ACK. T_p and T_c are the delays in propagation and in processing, respectively. T_s and T_a denote the transmission durations of a data frame and of an ACK/NAK frame, respectively. T_T denotes the total time, i.e., $T_T = T_p + T_s + T_c + T_p + T_a + T_c$.

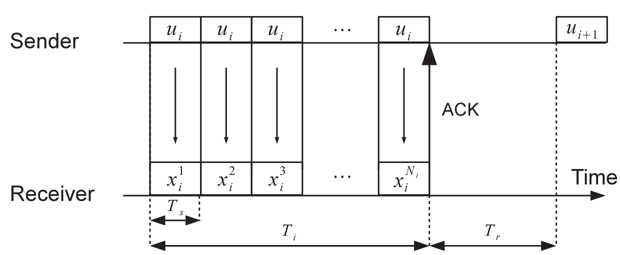


Fig. 2 Reliability-based ARQ. The sender keeps sending the same data until receiving an ACK. N_i, T_i and T_r denote the overall number of sending u_i, the overall duration, i.e., $T_i = T_s \times N_i$, and the time for a rest, respectively. Given that the forward channel is electrodes, cables, etc., and that the feedback one is via a monitor, T_p and T_c are negligibly shorter than T_s and T_r; thus, they were omitted.

Table 1 Stopping criterion and class estimation in each method.

Method	Stopping criterion	Class estimation
Averaging	$N = N_{\text{target}}$	$\hat{u}_i = \operatorname{argmax}_k P(k\|\mathcal{X}_i^N)$
Rejection	$\max_k P(k\|\boldsymbol{x}_i^N) > \lambda$	$\hat{u}_i = \operatorname{argmax}_k P(k\|\boldsymbol{x}_i^N)$
RB-ARQ	$\max_k P(k\|\mathcal{X}_i^N) > \lambda$	$\hat{u}_i = \operatorname{argmax}_k P(k\|\mathcal{X}_i^N)$

Algorithm 1 The algorithm of the proposed method

$N \leftarrow 0$
$\mathcal{X}_i^N \leftarrow \{\}$
repeat
  $N \leftarrow N + 1$
  get $\boldsymbol{x}_i^N$
  $\mathcal{X}_i^N \leftarrow \mathcal{X}_i^{N-1} \cup \{\boldsymbol{x}_i^N\}$
until (4) holds
estimate the class label $\hat{u}_i$
$N_i \leftarrow N$
return $\hat{u}_i$ and N_i

2.2 Performance Evaluation

The information transfer rate is defined as

$$\mathrm{ITR} = \frac{\log_2|\mathcal{K}| + p\log_2 p + (1-p)\log_2 \frac{1-p}{|\mathcal{K}|-1}}{d}, \tag{5}$$

where $|\mathcal{K}|$ is the number of elements in $\mathcal{K}$, e.g., $|\mathcal{K}| = 3$ for the first and the second experiments and $|\mathcal{K}| = 36$ for the last experiment, p is the classification accuracy, and d is the trial duration defined in Sect. 2.1. ITR is one of the popular performance measures in BCI research [1]. However, ITR defined in Eq. (5) does not represent the actual performance in practice since it represents the maximum rate, i.e., the channel capacity, and ITR should be below the upper limit to make the error rate arbitrary small [26]. Therefore, note that Eq. (5) is the "maximum" ITR. A new measure called *Utility* has been proposed for the P300 speller [27], defined as follows:

$$U = \frac{(2p-1)\log_2(|\mathcal{K}|-1)}{d} \tag{6}$$

if $p > 0.5$, otherwise $U = 0$. This measure assumes that the backspace is contained in the interface matrix to delete a previously misspelled letter if needed, and that the user tries to spell perfectly using it; thus, the transferred information is $\log_2(|\mathcal{K}|-1)$ and the expected length of time per letter is $d/(2p-1)$. That is, the utility is the ITR when the backspace is available and letters are perfectly spelled. The utility seems to be more practical than the ITR; thus, this paper utilized it as a performance measure for the P300 speller experiment in addition to the ITR.

3. Experiments

To evaluate the effectiveness of the proposed method, three experiments were carried out: a simulation using normal samples, and applications to mental imagery tasks including two motor imagery tasks and to the P300 speller task. The first experiment aimed at grasping the theoretical relationship between the accuracy and the average trial duration, especially when data were normally distributed since linear discriminant analysis (LDA), which is a popular classifier for BCIs and was used in all the experiments, assumed that data were normally distributed with equal covariances. The rest two experiments aimed at evaluating the proposed method when applied to the most famous BCI paradigms: motor imagery tasks and the P300 speller task [25].

The three experiments focused on a comparison of ARQs for BCIs. In the first two experiments, the proposed method, RB-ARQ, was compared with the standard averaging and the rejection rule because both could also be regarded as ARQs as described in Sect. 2.1 and the rejection rule had been proposed for the mental imagery task in the second experiment [15]. In the last experiment, the proposed method was compared with the standard averaging and the dynamic subtrial limiting method, hereafter referred to as Lenhardt's method [16]. Lenhardt's method is also an ARQ proposed for the P300 speller and it employs two criteria for repeat decision. Through the following experiments, it can be shown that the proposed method is the most effective ARQ for BCIs.

3.1 Experiment I: Simulation Using Normally Distributed Samples

3.1.1 Experimental Settings

Let $f_k(\boldsymbol{x})$ be the kth class-conditional probability and be a two-dimensional Gaussian distribution, each of which is equally-spaced on a circle of radius r:

$$f_k(\boldsymbol{x}) = \frac{1}{2\pi|\Sigma_k|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_k)'\Sigma_k^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)\right\} \tag{7}$$

with the following parameters,

$$\boldsymbol{\mu}_k' = r\,(\cos(2\pi k/|\mathcal{K}|), \sin(2\pi k/|\mathcal{K}|))\,, \tag{8}$$
$$\Sigma_k = I, \tag{9}$$

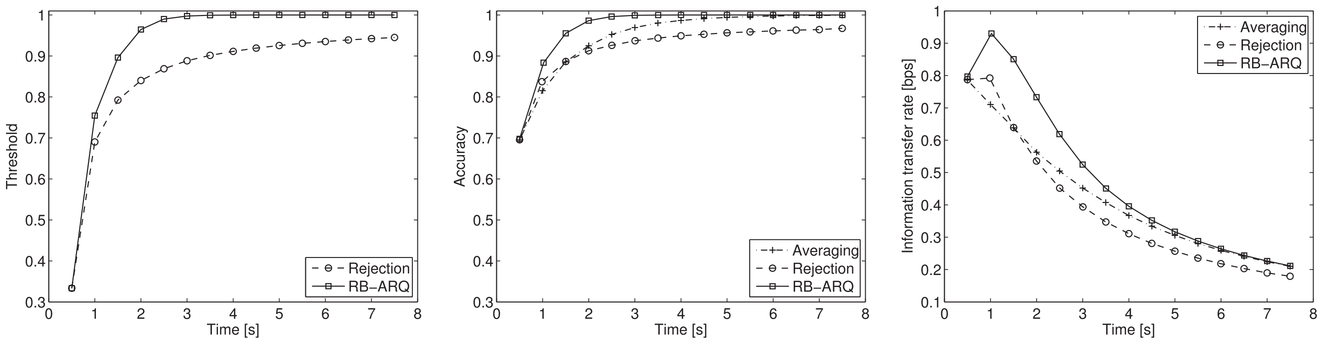
where $\boldsymbol{a}'$ and I denote the transpose of $\boldsymbol{a}$ and an identity matrix, respectively. It was assumed that the set of classes was $\mathcal{K} = \{0, 1, 2\}$ and that each sample $\boldsymbol{x}$ took a half second to measure, i.e., $T_s = 0.5$ [s]; thus, $d = 0.5 \times \frac{1}{N_{\mathrm{trial}}} \sum_i^{N_{\mathrm{trial}}} N_i$. Note that the posterior probability can be calculated as

$$P(k|\boldsymbol{x}) = \frac{\pi_k f_k(\boldsymbol{x})}{\sum_{l\in\mathcal{K}} \pi_l f_l(\boldsymbol{x})}, \tag{10}$$

where π_k denotes the prior probability and was set to be one-third. For each threshold, 1000 trials, $N_{\mathrm{trial}} = 1000$, were performed; and the thresholds were determined by the false position method so that the average trial durations were equal to a target duration within a tolerance of 0.05 s. For each target duration of 0.5 s, 1.0 s, …, 7.5 s, the procedure above was repeated for 100 times to calculate the ultimate average accuracy and trial duration. The algorithm is described in Algorithm 2.

3.1.2 Result and Discussion

Figures 3 (a), 3 (b) and 3 (c) compare the trial duration with the threshold, the accuracy and the information transfer rate,



(a) Time vs. threshold (b) Time vs. accuracy (c) Time vs. information transfer rate

Fig. 3 Theoretical relationships between the trial duration and the thresholds (Fig. 3 (a)), the trial duration and the accuracy (Fig. 3 (b)), and the trial duration and the ITR (Fig. 3 (c)), obtained using samples from normal distribution with the radius $r = 1$.

Algorithm 2 The algorithm to acquire the target trial duration and its corresponding accuracy

---

for $j = 1$ to 100 **do**
  initialize λ
  repeat
    for $i = 1$ to 1000 **do**
      $u_i \leftarrow \mod(i, 3)$
      estimate u_i by Algorithm 1
    end for
    calculate the accuracy p using the returned $\{\hat{u}_i\}$, and update λ by the false position method
  until $|d - d_{\text{target}}| \leq 0.05$
  store the accuracy p and the trial duration d
end for
calculate the ultimate accuracy and the trial duration

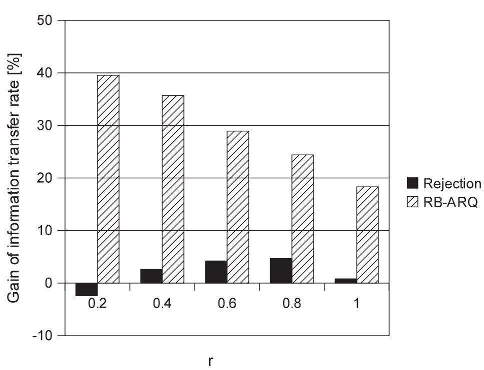


Fig. 4 Theoretical percentage gain of the information transfer rate with respect to Averaging, when the radius r changes.

respectively, with the radius (Eq. (8)) $r = 1$. Figure 4 shows the relationship between the radius r and the gain of the ITR with respect to Averaging. The gain was defined as

$$\frac{\max(\text{each method's ITR}) - \max(\text{averaging's ITR})}{\max(\text{averaging's ITR})}. \tag{11}$$

Figures 3 (a) and 3 (b) tell that as the threshold increases, the trial duration lengthens and the accuracy improves, and that RB-ARQ is superior to the others. It also tells that the rejection is superior to the averaing when the time is small, and vice versa when the time is large. It is worth noting that when the threshold is transformed as Eq. (12), the trial duration is roughly proportional to the transformed threshold, which simplifies the search for the threshold appropriate for a target trial duration using the false position method (see Algorithm 2).

$$\tilde{\lambda} = \begin{cases} \operatorname{arctanh}(\lambda) & \text{(RB-ARQ)} \\ \frac{1}{1-\lambda} & \text{(rejection)} \end{cases} \tag{12}$$

Figure 3 (c) tells that the ITR was maximized when the trial duration was 1 s using RB-ARQ, which was greater than the maximum ITR obtained using the averaging by about 20 percents. According to Fig. 4, as the radius r decreases, the gain of the ITR increases in the case of RB-ARQ, while the gain is almost zero in the case of the rejection. Three distributions were equally-spaced on a circle of radius r; thus, the smaller r was, the more the distributions were overlapped, leading to a poor accuracy. Therefore, this result suggests that untrained users could benefit more from RB-ARQ.

3.2 Experiment II: Mental Imagery Tasks

3.2.1 Data Description

The data used in this experiment was the data set V in BCI Competition III[†] [15]. It contains EEG data recorded from three subjects, i.e., subject 1, 2, and 3, when they were doing one of three different mental tasks. The three classes were left hand movements (class 2), right hand movements (class 3) and word generation (class 7), i.e., $\mathcal{K} = \{2, 3, 7\}$. Each data set of each subject includes four sessions: the first three as training data, and the last as test data. Each session lasted about 4 mins, in which the subject was doing a given task for about 15 s, then switched to another task on the operator's request. The data sets are provided in two styles: raw EEG signals, and precomputed features; and the latter one was employed in this experiment. Every 62.5 ms, i.e., $T_s = 62.5$ [ms], the power spectral density of the previous one second was estimated with a resolution of 2 Hz for eight

[†]Available: http://www.bbci.de/competition/ (accessed Aug. 12th, 2009)

channels, i.e., C3, Cz, C4, CP1, CP2, P3, Pz, and P4 [22] and then twelve frequency components: 8 Hz, 10 Hz, ..., 30 Hz were extracted and concatenated into a feature vector; therefore, the feature vector lied on 96 dimensional space. The number of the samples in one session was around 3,500.

3.2.2 Further Preprocessing and Application of Proposed Method

Considering the rule in the competition, requiring outputs of classifications every 0.5 s, i.e., $T_s = 0.5$ [s], consecutive eight samples in the provided data were averaged; thus, the number of the averaged samples per session was around 440. Then a variable selection was carried out, where the jth ($j = 1, 2, \ldots, 20$) subset candidate contained the most j separable variables in terms of the *Fisher's ratio* [23], and the subset with the best classification accuracy of five-hold cross validation with LDA as a classifier was selected. The reason for limiting the maximum number of variables to 20 was because the resultant number was about 10 in a preceding pilot study and it was desirable to reduce the computational cost. The dimension of the feature vector was reduced down to 2 by linear projection based on LDA [18], in which the projection matrix was different for each subject. Three normal distributions with equal covariances were estimated, and the posterior probabilities of test data was calculated by Eq. (10) with the prior probabilities being one-third for each class. Note that the variable selection and the distribution estimations were carried out separately for each subject. The rest procedure followed Algorithm 2; thus, $N_{\mathrm{trial}} = 1000$ and $d = 0.5 \times \frac{1}{N_{\mathrm{trial}}} \sum_i^{N_{\mathrm{trial}}} N_i$. It should be noted that new data sets were re-sampled with replacement from the original data sets, since the experimental task was performed in a way such that the proposed method was not applicable to the data sets as they were, i.e., the subject was doing the same task for 15 s regardless of the reliability

3.2.3 Result and Discussion

Figure 5 shows the scatter plots in the canonical variate space, i.e., LD1 and LD2, (Figs. 5 (a)–(c)), and comparisons of the methods in terms of the accuracy (Figs. 5 (d)–(f)), and of the ITR (Figs. 5 (g)–(i)). Table 2 shows the results of the Shapiro-Wilk's normality test [28], testing whether the samples in the canonical variate space were from the normal distributions. Figure 6 is the Q-Q plot of subject 2's class 3, whose W statistic is the lowest in Table 2, based on the fact that the Mahalanobis distance of a D-dimensional multivariate normal sample follows the chi-square distribution of D degrees of freedom [29].

According to Figs. 5 (a)–(c) and Table 2, it is hard to say from the statistical viewpoint that the samples were from normal distributions with equal covariance. However, Figs. 5 (d)–(f) show that the accuracy improved as the trial duration lengthened and the threshold became higher in the cases of RB-ARQ and the averaging, whereas it was not the case for the rejection. According to the central limit theorem, the distribution of the sample mean from any distribution approaches the normal distribution as the sample size increases [30]. In RB-ARQ and the averaging, where the reliability is $\max_k P(k|\mathcal{X}_i^N)$, or equivalently $\max_k P(k|\bar{\boldsymbol{x}}_i^N)$, the number of samples to be averaged increases as the trial duration and the threshold becomes larger; thus, the distribution of the sample mean $\bar{\boldsymbol{x}}_i^N$ approaches the normal distribution. This would be the reason why RB-ARQ and the averaging improved the accuracy, though each $\boldsymbol{x}_i^n$ is not necessarily from the normal distribution. As shown in Fig. 6, almost all plots were on the regression line except for several outliers, indicating that some technique for outlier removal could make the distribution more Gaussian. In terms of ITR, the maximum ITR obtained by using RB-ARQ was larger than those obtained by using other methods in the cases of the subjects 2 and 3, whereas they were almost equal in the case of the subject 1. This result is consistent with the experiment I, since the three distributions of the subject 1 were least overlapped as shown in Fig. 5 (a), while Fig. 4 shows the least overlapped case, i.e., $r = 1$, resulted in the least ITR gain.

Table 2 Results of Shapiro-Wilk normality test for experiment II.

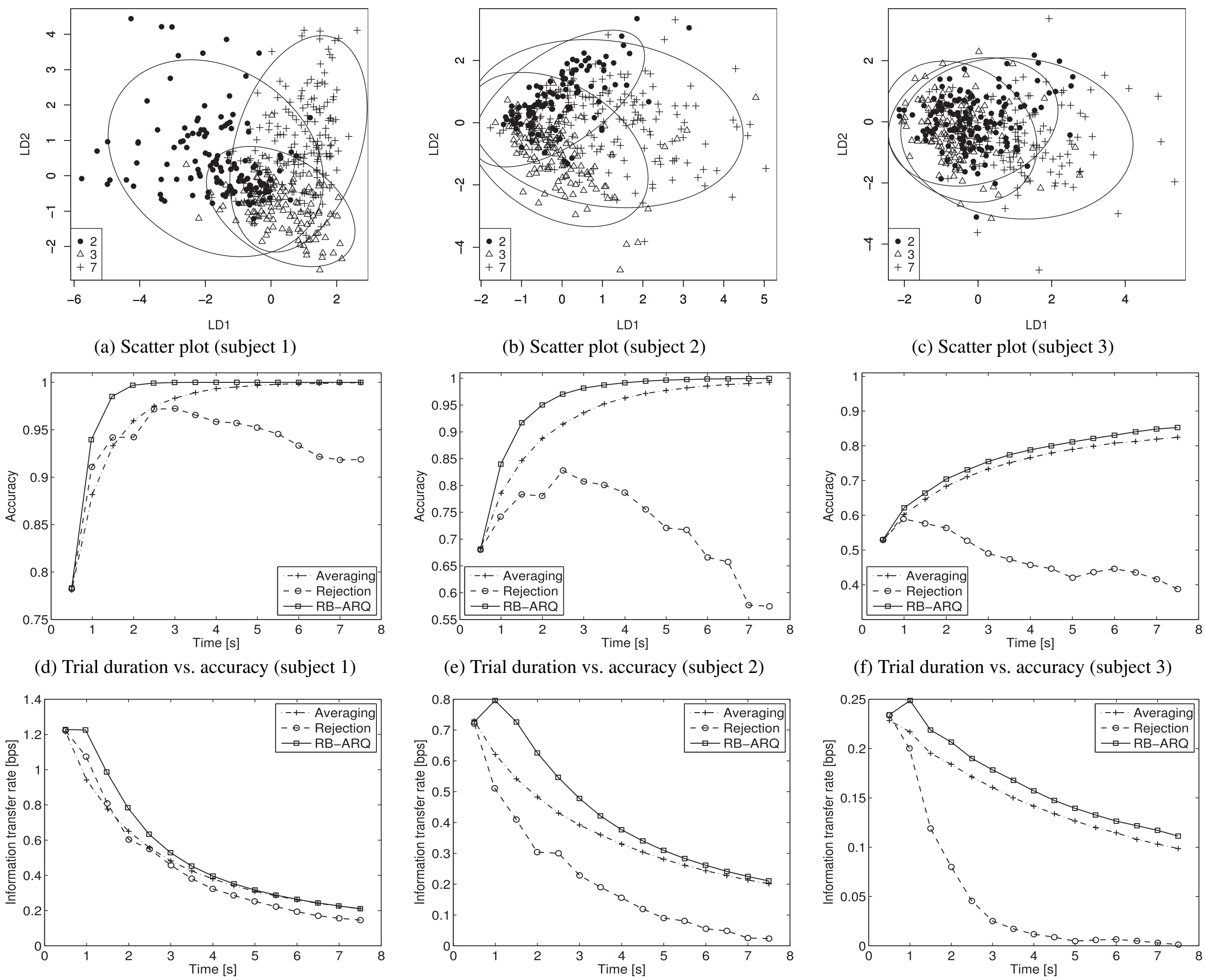
Subject	Class	W
1	2	0.91**
	3	0.98
	7	0.96**
2	2	0.95**
	3	0.77**
	7	0.99
3	2	0.99
	3	0.94**
	7	0.98*

** $p < 0.01$, * $p < 0.05$

3.3 Experiment III: P300 Speller

3.3.1 P300 Speller

The P300 speller is one of the BCI applications, which allows users to select letters just by thoughts [24], and is available in the BCI2000 [25]. The current study employed the ordinary 6×6 matrix interface composed of 26 alphabets, 9 numbers, and the whitespace. Each run consisted of spelling twenty letters "THE QUICK BROWNY FOX" in the *copy* mode. The subjects were given 3 s for moving their gaze to the target letter before the following stimulus presentations. Each row and column was successively and randomly intensified for 100 ms with an interval of 75 ms, the subject was asked to count how many times the row and the column containing the target letter flashed. The P300 response is elicited when the attended target letter flashes; accordingly, the target can be selected. Each sequence consisted of 12 flashes, i.e., 6 rows and 6 columns; and 5 sequences were performed to spell a letter. Immediately after the stimuli finished, the selected letter was shown for 1 s to elicit the ErrPs when the selected letter was wrong. However, only in the first two experiments involving subject A and B, a fixation



(a) Scatter plot (subject 1) (b) Scatter plot (subject 2) (c) Scatter plot (subject 3)

(d) Trial duration vs. accuracy (subject 1) (e) Trial duration vs. accuracy (subject 2) (f) Trial duration vs. accuracy (subject 3)

(g) Trial duration vs. ITR (subject 1) (h) Trial duration vs. ITR (subject 2) (i) Trial duration vs. ITR (subject 3)

Fig. 5 The first row: scatter plots in the canonical variate space [18]. Each ellipse represents 95 percent confidence region for each distribution. The second and the third rows: comparison of methods in terms of the accuracy, and of the ITR, respectively. Each column represents subject 1, 2, and 3, respectively.

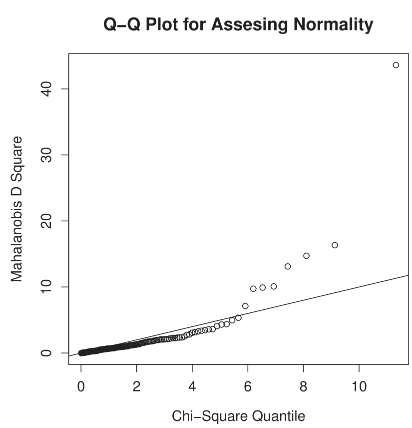


Fig. 6 Q-Q plot of subject 2's class 3.

point appeared at the center of the monitor after the stimuli and it was replaced by the result 1 s later.

3.3.2 Data Collection and Preprocessing

Five male volunteer subjects: A, B, C, D, and E in their early 20's with no prior experience in the P300 speller task participated in this experiment. Each subject was sitting on an armchair in a darkroom, facing a 20-inch LCD monitor. Their EEGs were recorded from Fz, C3, Cz, C4, and Pz referenced to the linked-ears with the sampling rate of 1000 Hz using a Polymate AP216 (DIGITEX LAB. CO., LTD, Tokyo, Japan). Their EEGs were down-sampled to 20 Hz and filtered with a pass-band of 1 Hz to 10 Hz since the P300 is a relatively slow potential change. Fourteen time points after the stimulus presentations were extracted from all five electrodes, and then they were concatenated into a feature vector; thus, the feature vector had $5 \times 14 = 70$ dimensions. Similar to the experiment II, the feature vector was projected onto 1 canonical variate by LDA [18]. The datasets of the first two runs were used to estimate two normal distributions: the P300 and the non-P300, which was different for each subject. The rest nine were used for evaluation: successive three runs were assumed to be a single run so that the results could be comparable to the conventional 15 sequences per letter paradigm. Hence, the number of sequences was 15 and that of spelled letters was 60.

3.3.3 Application of Proposed Method

Let $r \in \{1, 2, \ldots, 6\}$ and $c \in \{1, 2, \ldots, 6\}$ be the row and the column number, respectively, so that the letter in the rth row and the cth column can be expressed as (r, c). Suppose that $\boldsymbol{y}_R^r$ is an observed EEG, corresponding to rth row, and that $\mathcal{Y}_R = \{\boldsymbol{y}_R^r\}$ is a set of observed EEGs, only one of which contains the P300 response. Let $z_R^r \in \{0, 1\}$ denotes whether $\boldsymbol{y}_R^r$ contains the P300 ($z_R^r = 1$) or not ($z_R^r = 0$). The posterior probability that the P300 appears in the rth row given $\mathcal{Y}_R$ is

$$\begin{aligned} &P(\text{P300 in the } r\text{th row}|\mathcal{Y}_R) \\ &= P(z_R^r = 1|\mathcal{Y}_R) \prod_{r' \neq r} P(z_R^{r'} = 0|\mathcal{Y}_R), \\ &= P(z_R^r = 1|\boldsymbol{y}_R^r) \prod_{r' \neq r} P(z_R^{r'} = 0|\boldsymbol{y}_R^{r'}), \end{aligned} \tag{13}$$

assuming that only $\boldsymbol{y}_R^r$ depends on z_R^r. Taking that the P300 appears only once into consideration,

$$\begin{aligned} &P(\text{P300 in the } r\text{th row}|\mathcal{Y}_R) \\ &= \frac{P(z_R^r = 1|\boldsymbol{y}_R^r) \prod_{r' \neq r} P(z_R^{r'} = 0|\boldsymbol{y}_R^{r'})}{\sum_r \left\{ P(z_R^r = 1|\boldsymbol{y}_R^r) \prod_{r' \neq r} P(z_R^{r'} = 0|\boldsymbol{y}_R^{r'}) \right\}} \end{aligned} \tag{14}$$

Similarly, $\mathcal{Y}_C$ can be defined for columns. Let $\mathcal{Y}_{iR}^N$ and $\mathcal{Y}_{iC}^N$ be the $\mathcal{Y}_R$ and $\mathcal{Y}_C$ in the Nth sequence of the ith letter, respectively. A letter is selected so that the following posterior probability is maximized,

$$\begin{aligned} (\hat{r}, \hat{c}) &= \operatorname*{argmax}_{(r,c)} P(\text{P300 in the } r\text{th row}|\mathcal{Y}_{iR}^N) \\ &\quad \times P(\text{P300 in the } c\text{th column}|\mathcal{Y}_{iC}^N), \\ &= \operatorname*{argmax}_{(r,c)} P\left(\text{P300 in } (r, c) \,\middle|\, \mathcal{Y}_i^N\right), \end{aligned} \tag{15}$$

where $\mathcal{Y}_i^N = \mathcal{Y}_{iR}^N \cup \mathcal{Y}_{iC}^N$. The proposed method employs the posterior probability appeared in Eq. (15) for the request decision.

Similar to the experiment II, the proposed method was not applicable to the data sets as they were. Therefore, the proposed method was modified so that the number of sequences was limited to 15 (see Algorithm 3). Note that according to the experimental setting, the trial duration was defined as $d = \frac{1}{N_{\text{trial}}} \sum_i^{N_{\text{trial}}} (3 + T_i + 2) = \frac{1}{N_{\text{trial}}} \sum_i^{N_{\text{trial}}} (3 + 0.175 \times 12 \times N_i + 2)$ for subject A and B, $d = \frac{1}{N_{\text{trial}}} \sum_i^{N_{\text{trial}}} (3 + 0.175 \times 12 \times N_i + 1)$ for the rest subjects.

Algorithm 3 The modified algorithm of the proposed method

$N \leftarrow 0$
$\mathcal{X}_i^N \leftarrow \{\}$
while $N < 15$ **do**
  $N \leftarrow N + 1$
  get $\boldsymbol{x}_i^n$
  $\mathcal{X}_i^N \leftarrow \mathcal{X}_i^{N-1} \cup \{\boldsymbol{x}_i^N\}$
  if (4) holds **then**
    break
  end if
end while
estimate the attended letter $(\hat{r}, \hat{c})$
$N_i \leftarrow N$
return $\hat{u}_i$ and N_i

Table 3 Accuracy of spelling letters using all 15 sequences [%].

A	B	C	D	E	Avg.
95.0	98.3	83.3	91.7	91.7	92.0

Table 4 Results of Shapiro-Wilk normality test for experiment III.

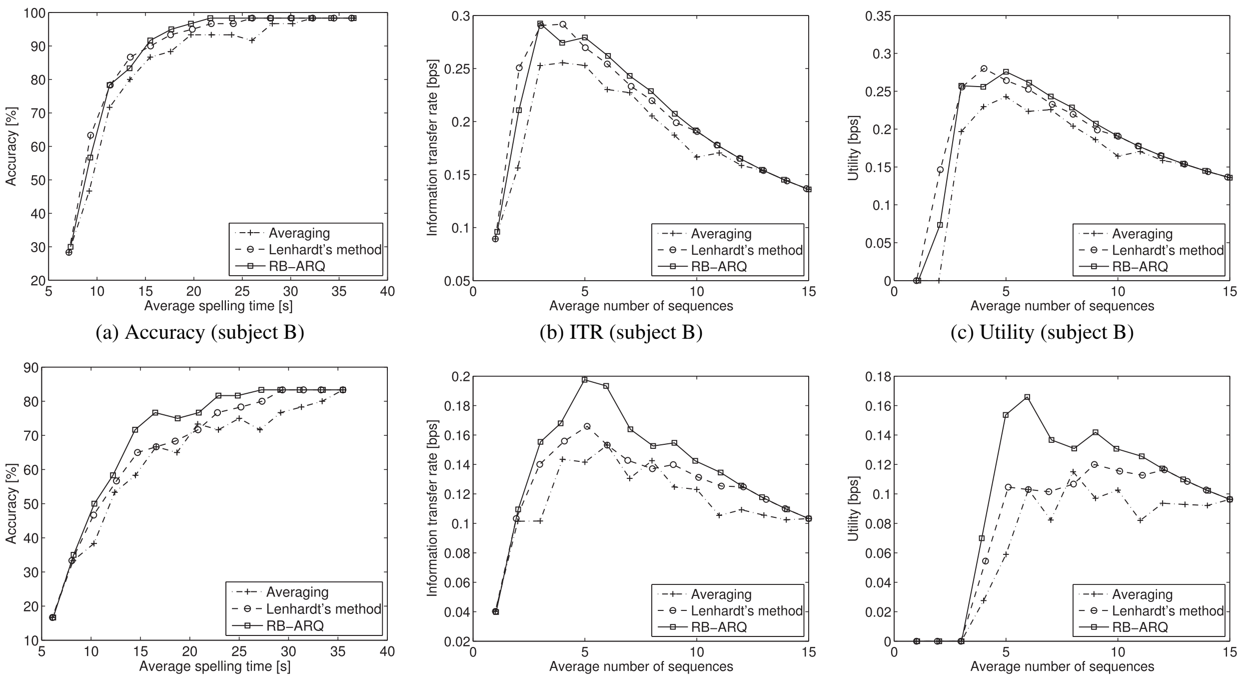
Subject	Class	W
A	P300	0.9980*
	Non-P300	0.9926**
B	P300	0.9991
	Non-P300	0.9996
C	P300	0.9985
	Non-P300	0.9939**
D	P300	0.9982*
	Non-P300	0.9997
E	P300	0.9994
	Non-P300	0.9996

** $p < 0.01$, * $p < 0.05$

3.3.4 Result and Discussion

Table 3 shows the accuracy of spelling letters using all the 15 sequences. Figure 7 describes the performance curves of subject B and C, whose accuracies were the best and the worst in Table 3, respectively. Figure 8 shows the percent gain of ITR and utility with respect to the Averaging. Table 4 also shows the results of the Shapiro-Wilk's normality test for the data sets in the experiment III.

Table 3 tells the accuracy was about 90 percents on average. This result can compare with the result of the data set II of the third BCI competition, where several competitors achieved the accuracy of over 90 percents [31]. Consequently, the following results could hold for data sets recorded in the conventional 15 sequences per letter paradigm. As shown in Fig. 7, the accuracy improved as the trial duration lengthened like the experiment I. The ITR and the utility were maximized when the number of sequences was 3 to 5 and 5 to 6, respectively, by RB-ARQ. According to Fig. 8, the maximum ITR and the maximum utility achieved by RB-ARQ were larger by about 20 and 30 percents on average, respectively, than those by the averaging. Moreover, RB-ARQ achieved more than double gains compared to the Lenhardt'd method. Table 4 tells that some



(a) Accuracy (subject B) (b) ITR (subject B) (c) Utility (subject B)

(d) Accuracy (subject C) (e) ITR (subject C) (f) Utility (subject C)

Fig. 7 Performance curves of the accuracy versus the duration of spelling a letter [(a), (d)], the ITR versus the number of sequences [(b), (e)], and the Utility versus the number of sequences [(c), (f)]. The first and the second row represent those for the subject B and the subject C, respectively.

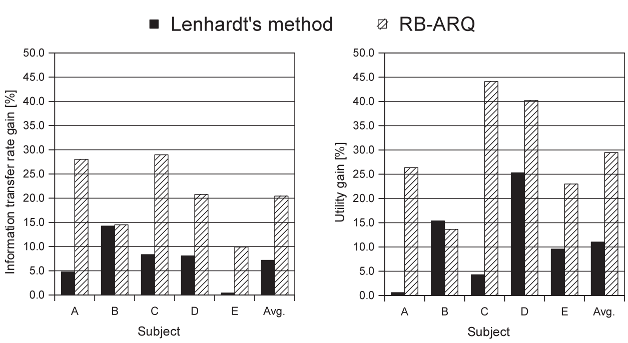


Fig. 8 Percentage gain of information transfer rate (left) and utility (right), obtained from the experiment III.

sample sets were not from normal distributions from the statistical viewpoint. However, the W statistic of all sample sets was almost equal to one, meaning that they could be practically regarded as the samples from the normal distributions. This implies that RB-ARQ works more effectively in the P300 speller task than in the mental tasks. It should be noted that the posterior probability can be estimated also when the support vector machine (SVM) is used as a classifier; however, the current result shows no statistical difference in performance between the uses of posterior probabilities given by the estimated Gaussian distributions and SVM [21]. This is partially because the posterior probabilities are poorly estimated by SVM [32]; thus, a novel estimation method is necessary to make RB-ARQ more useful also in mental tasks.

4. Conclusions

This paper presented reliability-based automatic repeat request (RB-ARQ), an error control method for BCIs, where a user keeps thinking or keeps being exposed to stimuli until the reliability, or the maximum posterior probability becomes larger than a given threshold. In order to evaluate the effectiveness of the proposed method, three experiments were conducted with different settings: a simulation using normally distributed samples, mental imagery tasks, and the P300 speller task. The experimental results showed that RB-ARQ was more effective than conventional methods, i.e., better accuracy when trial duration was the same, and shorter trial duration when the accuracy was the same. This resulted in a greater information transfer rate and a greater utility, which was a more practical performance measure in the P300 speller task. The results also showed that such users who achieved a poor accuracy for some reason could benefit the most from RB-ARQ, which could make BCIs more universal. In addition, since the posterior probabilities were calculated using estimated normal distributions, normality tests were carried out, whose results showed that samples were not necessarily from normal distributions from the statistical viewpoint. However, RB-ARQ worked successfully, partially because they could be practically regarded as normal samples and the sample mean approached the normal distribution as the sample size increased according to the central limit theorem. A further performance improvement could be possible by applying some outlier removal or other estimation methods for the posterior probabilities. Finally, on-line experiments are necessary to investigate the usefulness of the proposed method in the real environment.

Acknowledgement

We would like to thank all of the organizers for holding the BCI Competition, and IDIAP Research Institute for providing the EEG data. This work was supported by the Grant-in-Aid for JSPS Fellows No. 22-8417 and the Grant-in-Aid for Scientific Research (C) No. 22500200 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan, "Brain-computer interfaces for communication and control," Clinical Neurophysiology, vol.113, no.6, pp.767–791, 2002.

[2] N. Birbaumer, A. Kübler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann, and H. Flor, "The thought translation device (TTD) for completely paralyzed patients," IEEE Trans. Rehabil. Eng., vol.8, no.2, pp.190–193, 2000.

[3] R. Scherer, G.R. Müller, C. Neuper, B. Graimann, and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: Improvement of the spelling rate," IEEE Trans. Biomed. Eng., vol.51, no.6, pp.979–984, 2004.

[4] K. Tanaka, K. Matsunaga, and H.O. Wang, "Electroencephalogram-based control of an electric wheelchair," IEEE Trans. Robot., vol.21, no.4, pp.762–766, 2005.

[5] C. Vidaurre, A. Schögl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "A fully on-line adaptive BCI," IEEE Trans. Biomed. Eng., vol.53, no.6, pp.1214–1219, 2006.

[6] A. Buttfield, P.W. Ferrez, and J.R. Millán, "Towards a robust BCI: Error potentials and online learning," IEEE Trans. Neural Syst. Rehab. Eng., vol.14, no.2, pp.164–168, 2006.

[7] T. Hinterberger, N. Neumann, M. Pham, A. Kübler, A. Grether, N. Hofmayer, B. Wilhelm, H. Flor, and N. Birbaumer, "A multimodal brain-based feedback and communication system," Experimental Brain Research, vol.154, no.4, pp.521–526, 2004.

[8] D.J. McFarland, W.A. Sarnacki, and J.R. Wolpaw, "Brain-computer interface (BCI) operation: Optimizing information transfer rates," Biological Psychology, vol.63, pp.237–251, 2003.

[9] H. Serby, E. Yom-Tov, and G.F. Inbar, "An improved p300-based brain-computer interface," IEEE Trans. Neural Syst. Rehab. Eng., vol.13, no.1, pp.89–98, 2005.

[10] M.G. Helander, T.K. Landauer, and P.V. Prabhu, Handbook of human-computer interaction, 2nd ed., Elsevier, 1997.

[11] J.R. Wolpaw, H. Ramoser, D.J. McFarland, and G. Pfurtscheller, "EEG-based communication: Improved accuracy by response verification," IEEE Trans. Rehabil. Eng., vol.6, no.3, pp.326–333, Sept. 1998.

[12] B. Obermaier, G.R. Müller, and G. Pfurtscheller, " "Virtual keyboard" controlled by spontaneous EEG activity," IEEE Trans. Neural Syst. Rehab. Eng., vol.11, no.4, pp.422–426, 2003.

[13] G. Schalk, J.R. Wolpaw, D.J. McFarland, and G. Pfurtscheller, "EEG-based communication: Presence of an error potential," Clinical Neurophysiology, vol.111, no.12, pp.2138–2144, 2000.

[14] J.M. Shea, "Reliability-based hybrid ARQ," IEE Electron. Lett., vol.38, no.13, pp.644–645, 2002.

[15] J.R. Millán, "On the need for on-line learning in brain-computer interfaces," Proc. Int. Joint Conf. Neural Networks, pp.2877–2882, 2004.

[16] A. Lenhardt, M. Kaper, and H.J. Ritter, "An adaptive p300-based online brain-computer interface," IEEE Trans. Neural Syst. Rehab. Eng., vol.16, no.2, pp.121–130, April 2008.

[17] J.B. Anderson and S. Mohan, Source and channel coding: An algorithmic approach, Springer, 1991.

[18] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009.

[19] H. Takahashi, T. Yoshikawa, and T. Furuhashi, "An introduction of communication model and binary erasure channel to BCI," IEICE Trans. Inf. & Syst. (Japanese Edition), vol.J92-D, no.1, pp.153–161, Jan. 2009.

[20] H. Takahashi, T. Yoshikawa, and T. Furuhashi, "Application of reliability-based automatic repeat request to multi-class classification for brain-computer interfaces," Proc. 2009 IEEE Int. Conf. on Fuzzy Systems, pp.1027–1032, 2009.

[21] H. Takahashi, T. Yoshikawa, and T. Furuhashi, "Application of support vector machines to reliability-based automatic repeat request for brain-computer interfaces," 31st Annual Int. Conf. IEEE Engineering in Medicine and Biology Society, pp.6457–6460, 2009.

[22] G.H. Klem, H.O. Lüders, H.H. Jasper, and C. Elger, "The ten-twenty electrode system of the international federation," Electroencephalogr. Clin. Neurophysiol. Suppl., vol.52, pp.3–6, 1999.

[23] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification, 2nd ed., New Technology Communications, 2001.

[24] L.A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," Electroencephalogr. Clin. Neurophysiol., vol.70, no.6, pp.510–523, 1988.

[25] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw, "BCI2000: A general-purpose brain-computer interface (BCI) system," IEEE Trans. Biomed. Eng., vol.51, no.6, pp.1034–1043, 2004.

[26] C. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J., vol.27, pp.379–423, pp.623–656, 1948.

[27] B.D. Seno, M. Matteucci, and L. Mainardi, "The utility metric: A novel method to assess the overall performance of discrete brain-computer interfaces," IEEE Trans. Neural Syst. Rehab. Eng., vol.18, no.1, pp.20–28, 2009.

[28] S.S. Shapiro and M.B. Wilk, "An analysis of variance test for normality," Biometrika, vol.52, no.3, p.591, 1965.

[29] R. Khattree and D.N. Naik, Applied multivariate statistics with SAS software, SAS Publishing, 1999.

[30] C.H. Brase and C.P. Brase, Understandable Statistics: Concepts and Methods, Cengage Learning, 2007.

[31] B. Blankertz, K.R. Müller, D.J. Krusienski, G. Schalk, J.R. Wolpaw, A. Schögl, G. Pfurtscheller, J.R. Millán, M. Schröder, and N. Birbaumer, "The BCI competition III: Validating alternative approaches to actual BCI problems," IEEE Trans. Neural Syst. Rehab. Eng., vol.14, no.2, pp.153–159, 2006.

[32] C.M. Bishop, Pattern recognition and machine learning, p.336, Springer, 2006.

Hiromu Takahashi received the B. Eng. and M. Eng. degrees both from Nagoya University, Japan, in 2007 and 2009, respectively. Currently, he is pursuing the Ph.D. degree in the graduate school of engineering, Nagoya University, Japan, and is a research fellow of the Japan Society for the Promotion of Science. His research interest includes neuroengineering and data-mining. He is a member of IEEE.

Tomohiro Yoshikawa received the Ph.D. degree from Nagoya University, Japan, in 1997. He was a visiting researcher at University of California at Berkeley in 1997, an Assistant Professor at Mie University, Japan from 1998 to 2005. He is currently an Associate Professor at Nagoya University. His research interests are in Soft Computing and its applications. He is a member of IEEE, Japan Society for Fuzzy Theory and Intelligent Informatics, and Japan Society of Kansei Engineering.

Takeshi Furuhashi received the Ph.D. degree in electrical engineering from Nagoya University, Nagoya, Japan, in 1985. He is currently a Professor of the Department of Computational Science and Engineering at Nagoya University, Nagoya, Japan. Before joining Nagoya University in 2004, he was a Professor of Computer Science at Mie University, Tsu, Japan. From 1988 to 2000, he was with Nagoya University. From 1985 to 1988, he was an engineer of Toshiba Corporation in Japan. His research interests are in soft computing and Kansei Engineering. He is a member of IEEE, Japan Society for Fuzzy Theory and Intelligent Informatics, and The Institute of Electrical Engineers of Japan, etc.