SCIS & ISIS 2010, Dec. 8-12, 2010, Okayama Convention Center, Okayama, Japan

Combination of Reliability-based Automatic Repeat reQuest with Error Potential-based Error Correction for Improving P300 Speller Performance

Hiromu Takahashi*, Tomohiro Yoshikawa†, and Takeshi Furuhashi‡
Nagoya University
Furo-cho, Chikusa-ku, Nagoya 4648603, Japan
Email: *takahashi@cmplx.cse.nagoya-u.ac.jp, {†yoshikawa, ‡furuhashi}@cse.nagoya-u.ac.jp

***Abstract*—The P300 speller is one of the BCI applications, which allows users to select letters just by thoughts. However, due to the low signal-to-noise ratio of the P300, signal averaging is often performed, which improves the spelling accuracy but degrades the spelling speed. The authors have proposed *reliability-based automatic repeat request* (RB-ARQ) to ease this problem. RB-ARQ could be enhanced when combined with the error correction based on the error-related potentials (ErrPs) that occur on erroneous feedbacks. Thus, this study aims to reveal the characteristics of the ErrPs in the P300 speller paradigm, and to combine RB-ARQ with the ErrP-based error correction to improve the performance further. The results show that the performance of the P300 speller could be improved by 40 % on average.**

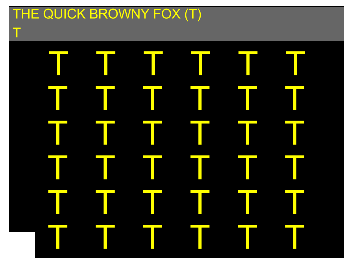


Fig. 1. The user interface used in the experiment, presenting a result "T" in yellow in every element in the 6×6 matrix so that the user need not move their gaze point, which could prevent undesirable EOG artifacts and shorten the duration needed for the ErrP detection.

I. Introduction

Brain-Computer Interfaces (BCIs) are promising technologies to restore control and communication to severely paralyzed people such as those with amyotrophic lateral sclerosis (ALS), and they are appealing to healthy people as well. The P300 speller is one of the BCI applications, which allows users to select letters just by thoughts [1]. However, due to the low signal-to-noise ratio of the P300, signal averaging is often performed [1], [2], which improves the spelling accuracy but degrades the spelling speed. The authors have proposed *reliability-based automatic repeat request* (RB-ARQ) to ease this problem [3], i.e., the spelling can be faster while preserving the accuracy. Meanwhile, it has been reported that the error-related potentials (ErrPs) occurring on erroneous feedbacks could be used for error correction in BCIs [4], [5], [6]. This suggests that RB-ARQ could be enhanced when combined with the ErrP-based error correction. Ferrez and Millán [5] categorized ErrPs found in choice reaction tasks into three groups: response ErrP, feedback ErrP, and observation ErrP. In the first two cases it is the subject who makes an error, and in the last case it is another person. By contrast, in BCI systems it is not very clear who is responsible for errors in recognizing the user's intent, it could be the system, the user, or even both. In fact, the ErrPs observed in their cursor control BCI experiment, where the subject delivered commands manually rather than mentally, were different from the reported ErrPs, and they were named "interaction ErrP." It was also reported that the classification results were fed back to the subject with the accuracy of either 80% or 50%; as a result, the ErrP amplitude with the accuracy of 80% was larger than that with the accuracy of 50%. This implies that a greater accuracy would lead to more accurate ErrP detections. Dal Seno et al. [6] conducted an on-line error correction based on the ErrPs in the P300 speller, which was the first attempt according to them, and the ErrP classification accuracy was around 60%. Nonetheless, the characteristics of the ErrPs in the P300 speller have not been revealed, and a pause of 1 s before the result presentation is necessary in their method, which would worsen the spelling speed. Therefore, the present study has three purposes: firstly to reveal the characteristics of the ErrPs in the P300 speller and compare them with the reported ErrPs, secondly to reduce the extra duration needed for the ErrP detection, lastly to combine RB-ARQ with the ErrP-based error correction to further improve the performance.

II. Experiment

A. P300 speller

The P300 speller is available in the BCI2000 [7]. The current study employed the 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: target letters were presented at the top of the interface, and the letter in parenthesis denoted the target in the current trial (see Fig. 1). The subject was 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 flashed. Each sequence consisted of 12 flashes, i.e., 6 rows and 6 columns; and 5 sequences were performed to spell a letter. Note that the sequences were relatively few so as to obtain sufficient ErrPs data. Generally, the P300 speller selects the letter that most likely has the P300 in its corresponding electroencephalograms (EEGs). In our experiment, the selected letter was shown for 1 s immediately after the stimuli finished (as in Fig. 1), to elicit the ErrPs when the selected letter was wrong. However, only in the first two experiments involving subject A and B, a fixation point appeared at the center of the monitor after the stimuli and it was replaced by the result 1 s later. Each subject participated eleven runs: the first run without displaying the result and the others with it. In this experiment, the results did not depend on their EEGs, but they were the same as the target with a given accuracy: 80 % in the randomly chosen half and 50 % in the rest half, which the user was not informed of. Note that the white square in the bottom-left appeared synchronously with the result presentation so that a photo-sensor attached to the monitor could detect the exact time when the result became visible. However, this was used only for the result presentations, and those of the row/column flashes were determined using the *state* information recorded with EEGs [8], which was generated roughly 30 ms before the actual flash due to a display lag.

B. Data collection, processing, and classification

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).

In the off-line waveform analysis, their EEGs were down-sampled to 100 Hz and filtered with a pass-band of 1 Hz to 10 Hz since both the P300 and the ErrPs are relatively slow potential changes. The EEGs after the result presentations were extracted with a time-window of 500 ms, and a threshold rejection of $\pm 50 \mu V$ was applied not to include those contaminated by artifacts such as eye movements. In the off-line classification analysis, no rejection was applied, i.e., $20 \times 10 = 200$ ErrP samples were all used for each subject. The data were further down-sampled to 20 Hz, and a time-window from 200 ms to 500 ms was used as the feature vector, i.e., the vector had $5 \times 7 = 35$ dimensions. The linear discriminant analysis (LDA) was employed, and the classifier was trained separately for each subject. The ten-fold cross validation [9] has been performed to estimate the classification accuracy because the samples were very limited, and the average accuracy of 100 runs of the cross validations was used for the later performance evaluation. The EEGs corresponding to the P300 were similarly collected except that a time-window of 650 ms was used instead. The datasets of the first two runs were used to train a classifier, 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 studies. Hence, the number of sequences n was 15 and that of spelled letters was 60.

C. Reliablity-based automatic repeat request

Suppose there are L possible letters in the P300 speller and $\boldsymbol{x}_i$ is the EEG data recorded in the ith sequence. $P(l|\boldsymbol{x}_i)$ denotes the posterior probability for letter $l \in L$ given $\boldsymbol{x}_i$. After ith sequence (often 15th sequence), the letter that has the maximum posterior probability or the *reliability* λ_i (1) is selected.

$$\lambda_i = \max_l P(l|\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_i) \tag{1}$$

Reliability-based automatic repeat request [3] employs the reliability as a criterion to determine the number of sequences in each trial. The stimulus presentations halt when a sufficient reliability is acquired, i.e., when $\lambda_i \geq \lambda$ is satisfied, where $\lambda \in [0 \quad 1]$ is an arbitrary threshold. Note that a larger threshold leads to a better accuracy and a longer trial duration. The number of sequences for each trial n_i could be different; thus, the average of them $\bar{n}$ is used for the later performance evaluation.

D. Performance evaluation

The information transfer rate (ITR) is one of the popular performance measures in BCI research [10]. Dal seno et al., however, argue that the ITR does not represent the actual performance in practice, and they have proposed a new measure called *Utility* [11], defined as follows:

$$U = \frac{\log_2(N-1)(2p-1)}{d} \tag{2}$$

if $p > 0.5$ otherwise $U = 0$, where N is the number of possible selections, e.g., $N = 36$ in the current experiment, p is the possibility of correct letter selections or simply the accuracy, and d is the trial duration, defined as $d = 3 + 0.175 \times 12 \times \bar{n} + 2$ [s] for subject A and B, $d = 3 + 0.175 \times 12 \times \bar{n} + 1$ [s] for the rest in the present study. 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. Suppose that e and c is the true-positive rate and the true-negative rate respectively, i.e., they are the possibilities that the EEG data after a wrong and correct result presentation is correctly recognized as ErrP and non-ErrP, respectively. Also suppose that if ErrPs detected, the selected letter is cancelled and the user need to spell it again. Then, the accuracy and the duration of spelling a letter when using the ErrP-based error correction become $\frac{pc}{pc+(1-p)(1-e)}$ and $\frac{d}{pc+(1-p)(1-e)}$, respectively.

III. Results

Fig. 2 describes the grand average waveforms over all subjects at Cz after the result presentations. It compares those for error, correct, and error-minus-correct when the accuracy was 80 %; and the grand averages between different accuracies,

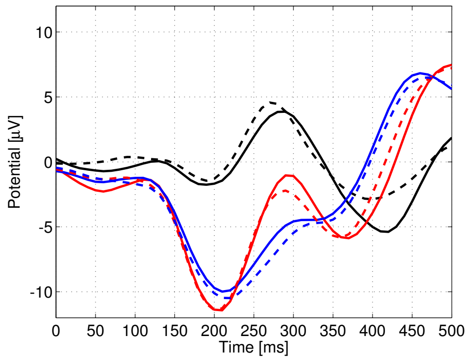


Fig. 2. The solid red, blue, and black waves represent those for error, correct, and error-minus-correct, respectively, when the accuracy was 80 %. The dashed waves are those when the accuracy was 50%

TABLE I
TRUE-POSITIVE AND TRUE-NEGATIVE RATES IN ERRP DETECTION [%]

p	e/c	A	B	C	D	E	Avg.
50	e	83.6	78.1	74.1	80.7	81.2	79.6
	c	78.0	71.2	73.5	80.5	73.8	75.4
80	e	68.6	77.3	72.0	85.0	65.3	73.6
	c	88.7	88.4	86.1	93.4	81.1	87.5

TABLE II
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

i.e., 80 % and 50 %. Table I shows the average accuracies of the cross validation results. Table II shows the accuracy of spelling letters using all the 15 sequences. Remember that these accuracies did not affect the result presentations. Fig. 3 describes the performance curves of subject B and C, whose accuracies were the best and the worst in Table II, respectively. When RB-ARQ applied, the average number of sequences $\bar{n}$ is conditional on the threshold λ; thus, the thresholds were determined so that $\bar{n}$ became $1, 2, \ldots, 15 \pm 0.1$, and each series had 15 data points. Note that based on an assumption that the ErrP detection rates depend on the accuracy, they were calculated using the following linear interpolation formulas made from Table I:

$$e = -0.200 \times p + 0.896, \qquad (3)$$

$$c = 0.403 \times p + 0.552. \qquad (4)$$

Also note that the ITR and the Utility were compared with the average number of sequences because the duration of spelling a letter went to infinity when $U = 0$. Fig. 4(a) and 4(b) show the ITR and the Utility gain of each method from Averaging. The gain was defined as the proportion of the maximal ITR/Utility obtained by each method to that obtained by Averaging, assuming that in practice the system operator would select the best performance in each method.

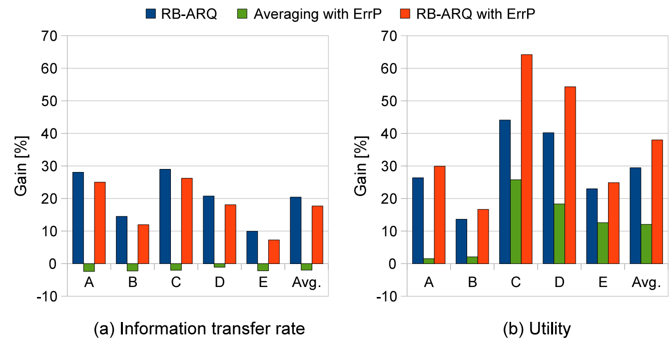


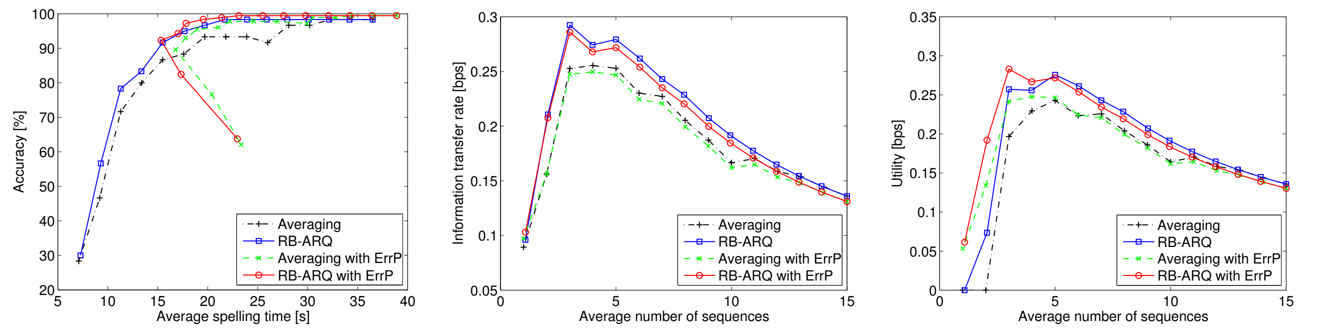
Fig. 4. Performance gain from Averaging.

IV. DISCUSSIONS

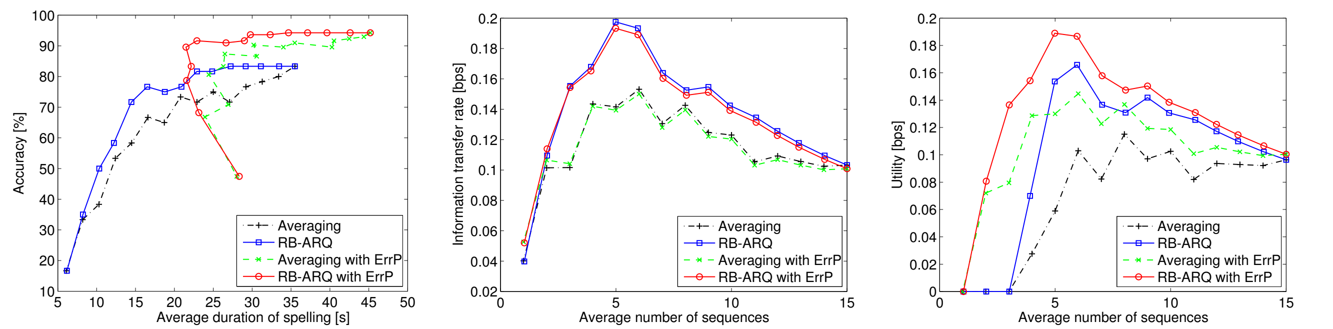
Fig. 2 shows that the error-minus-correct EEG had a weak negative peak, a positive peak, and a negative peak at roughly 190, 290, and 420 ms, respectively. These characteristics are fairly consistent with those described in [5]; thus, the ErrPs observed in the current study should be categorized into "interaction" ErrP rather than the other three. Fig. 2 shows that the 420-ms component is larger at $p = 80$ [%] than at $p = 50$ [%], which is also comparable to the existing result, whereas the difference of the 290-ms component is not very evident. The present results tell that the ErrPs in the P300 speller task are similar to those in the cursor control task, and that whether the subject is informed of how the presented results are generated does not seemingly affect the ErrPs (remember that in the cursor control task, the subject should had known that the results were independent of their commands [5]).

Table I tells that both the true-positive and the true-negative rates do not necessarily improve when the accuracy p improves. However, considering that the current results were obtained in a rather arbitrary parameter setting, e.g., the time-window, a further analysis should be carried out to draw a conclusion. Table I also tells that there does not seem to be a difference in the rates between those who saw the results with the pause of 1 s and those without it. This suggests that the pause of 1 s can be eliminated when using our user interface, which allows the user to see the result without moving their eyes whichever letter he or she is watching.

Table II tells the attained accuracies are satisfactory enough to proceed to a further analysis. Fig. 3(a) and 3(d) show that the accuracy improved as the trial duration increased in both series of Averaging and RB-ARQ, and that the ErrP-based error correction boosted the accuracy, but aggravated the trial duration. Fig. 3(b) and 3(e) illustrate that RB-ARQ improved the information transfer rate while the ErrP-based error correction did not contribute to the improvements. Contrastingly, Fig. 3(c) and 3(f) illustrate that the difference between those with and without the error correction observably widened, especially when sequences were few, i.e., when the accuracy was poor. These results are consistent with Fig. 4: Fig. 4(a) shows it was not the ErrP-based error correction



(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. 3. 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.

but RB-ARQ which augmented the information transfer rate; however, Fig. 4(b) shows that the error correction improved the Utility, particularly for subject C, who had the worst accuracy in Table II; and that RB-ARQ combined with the error correction increased the performance by about 40 % on average. Nonetheless, these results are conditional on both the ErrP classification rates estimated by the cross-validations and the classifier for the P300 trained using the datasets recorded in the same day; thus, further analyses and experiments are necessary.

V. Conclusion and Future Works

In this study, five subjects performed the P300 speller task with result feedbacks. The results show that ErrPs observed in this experiment were similar to those observed in the cursor control task. The classification results tell that the duration needed for ErrP detection could be shorten, and that the reliability-based automatic repeat request were enhanced its performance by 40 % on average when combined with the ErrP-based error correction. However, these results are conditional; therefore, further analyses and experiments are necessary.

Acknowledgements

This work was supported by the Grant-in-Aid for JSPS Fellows No. 22-8417 and for Scientific Research (C) No. 22500200 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

[1] 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.

[2] D. Krusienski, E. Sellers, D. McFarland, *et al.*, "Toward enhanced p300 speller performance," *J. Neuroscience Methods*, vol. 167, no. 1, pp. 15–21, 2008.

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

[4] G. Schalk, J. R. Wolpaw, D. J. McFarland, *et al.*, "EEG-based communication: presence of an error potential," *Clin. Neurophysiol.*, vol. 111, no. 12, pp. 2138–2144, 2000.

[5] P. W. Ferrez and J. del R. Millan, "Error-Related EEG potentials generated during simulated Brain-Computer interaction," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 3, pp. 923–929, 2008.

[6] B. Dal Seno, M. Matteucci, and L. Mainardi, "On-line detection of p300 and error potentials in a BCI speller," *Computational Intelligence and Neuroscience*, vol. 2010, Article ID 307254, 5 pages, 2010.

[7] G. Schalk, D. McFarland, T. Hinterberger, *et al.*, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1034–1043, 2004.

[8] G. Schalk and J. Mellinger, *A Practical Guide to Brain-Computer Interfacing with BCI2000*. Springer, 2010.

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

[10] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, *et al.*, "Brain-computer interface technology: a review of the first international meeting," *IEEE Trans. Rehab. Eng.*, vol. 8, no. 2, pp. 164–173, 2000.

[11] 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.