FUZZ-IEEE 2009, Korea, August 20-24, 2009

Application of Reliability-Based Automatic Repeat Request to Multi-Class Classification for Brain-Computer Interfaces

Hiromu Takahashi, Tomohiro Yoshikawa and Takeshi Furuhashi

***Abstract*— A Brain-Computer Interface (BCI) is a system which could enable patients like those with Amyotrophic Lateral Sclerosis to control some equipment and to communicate with other people, and has been anticipated to be achieved. One of the problems in BCI research is a trade-off between transmission speed and accuracy. In the field of data transmission, on the other hand, Reliability-Based Hybrid Automatic Repeat reQuest (RB-HARQ), one of the error control methods, has been developed to achieve both of the performances. The authors, therefore, have considered BCIs as communications between users and computers, and applied Reliability-Based ARQ, customized RB-HARQ, to BCIs. It has been shown that the proposed method is superior to other error control methods in two-class classification. In this paper, the proposed method is extended to deal with multi-class classification of EEG data, and is shown to be effective in multi-class problems.**

I. INTRODUCTION

Recently, a lot of research on Brain-Computer Interfaces (BCIs) which record brain activities, discriminate the thoughts, and then enable patients like those with Amyotrophic Lateral Sclerosis (ALS) to control some equipment or to communicate with others has been reported. The authors also have been studying on a BCI based on Electroencephalogram (EEG), which is considered as one of the most reasonable measurements since it is non-invasive and costs less [1]. In fact, EEG-based BCIs have been researched well; for example, Thought Translations Device (TTD) [2] by Birbaumer et al.; and Graz-BCI [3] by Pfurtscheller et al., which employs the band power from 8 to 13 Hz (alpha band) as a feature of EEG, and applies Linear Discriminant Analysis (LDA) [4] to it. The accuracy, however, is not high since biological signals such as EEGs contain much noise, partly due to users' physical or mental conditions. On the other hand, it also has been suggested that the longer EEG data used for one discrimination is, the more accuracy could be achieved [3], [5], [6]. It could be possible to say that high accuracy can be gained in exchange for speed, that is the length of data, in those methods; here seems a trade-off between accuracy and speed. The purpose of this study, therefore, is to develop a BCI which accomplishes both of them simultaneously.

In the field of data transmission, there are some error control methods; for instance, Forward Error Correction (FEC), which allows the receiver to detect and correct errors; Automatic Repeat reQuest (ARQ), which asks the transmitter to repeat code words; and Hybrid ARQ (HARQ), which is a combination of ARQ and FEC. In the past several years, Reliability-Based Hybrid ARQ (RB-HARQ) has been proposed [7]. This method is a variation of HARQ, in which the requests are based on reliability of each bit in code words. It also has been reported that RB-HARQ can provide performance close to the channel capacity.

BCIs can also be considered as communications between users and computers [8]. The standard averaging can be seen as an error control in BCIs, where a transmitter requests certain times, and a received series is combined and decoded. Millán et al. propose a classification method, which classifies EEG data as "unknown" when its expected accuracy of classification is not high enough [9]. This paper calls their method and the standard averaging as Basic RB-ARQ and Constant ARQ, respectively. Similar to these, the authors have proposed Reliability-Based ARQ (RB-ARQ), customized RB-HARQ [10], and shown that the proposed method is superior to either Constant ARQ or Basic RB-ARQ theoretically and empirically in two-class classification.

In this paper, RB-ARQ is extended to deal with multi-class classification, and applied to one of the EEG data in the BCI Competition[1], which is held to examine signal processing and classification methods for BCIs, and provides EEG data to the public.

II. RELIABILITY-BASED AUTOMATIC REPEAT REQUEST

Suppose a user has one thought out of two in his mind (e.g. imagination of left or right hand movement), and p-dimensional feature vector is obtained from EEG data. Correspondingly, let $u \in \{0, 1\}$ be the thought label, $\boldsymbol{x}^t \in \mathbb{R}^p$ be the feature vector at time t, and X^T be the set of $\boldsymbol{x}^t$ at time T ($X^T = \{x^t | t = 1, 2, \ldots, T\}$). Then, the log-likelihood ratio (LLR) given X^T, λ^T, can be obtained as follows:

$$\lambda^T = \ln \frac{\pi_0 p(X^T|0)}{\pi_1 p(X^T|1)}, \tag{1}$$

where π_k and $p(X^T|k)$ denote the priori probability of k, and the the conditional density function of X^T given k, respectively. Suppose that each $\boldsymbol{x}$ is independent and identically distributed (i.i.d.), and $f_k(\boldsymbol{x})$ is the conditional density function of $\boldsymbol{x}$ given k, (1) can be re-written as follows:

$$\lambda^T = \ln \frac{\pi_0 \prod_t f_0(\boldsymbol{x}^t)}{\pi_1 \prod_t f_1(\boldsymbol{x}^t)}, \tag{2}$$

$$= \ln \frac{\pi_0}{\pi_1} + \sum_t \ln \frac{f_0(\boldsymbol{x}^t)}{f_1(\boldsymbol{x}^t)}. \tag{3}$$

Department of Computational Science and Engineering, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 4648603, Japan takahashi@cmplx.cse.nagoya-u.ac.jp

[1] http://ida.first.fhg.de/projects/bci/competitions/

978-1-4244-3597-5/09/$25.00 ©2009 IEEE

The absolute value of LLR, $|\lambda^T|$, means the reliability; the larger it is, the higher the probability of correct classification is. Hence, in RB-ARQ, if the reliability is smaller than a given threshold, λ, the system requests the same thought; otherwise the thought label, u, can be estimated as follows:

$$\begin{cases} \text{repeat request} & \text{if } |\lambda^T| < \lambda, \\ \text{estimate } \hat{u} = & \begin{cases} 0 & \text{else if } \lambda^T \geq \lambda, \\ 1 & \text{otherwise,} \end{cases} \end{cases} \quad (4)$$

where a "hat" denotes an estimation.

RB-ARQ was designed for two-class classification as RB-HARQ deals with it (i.e. 0 or 1). On the other hand, there could be BCIs based on multi-class classification in order to increase the amount of information per classification. Therefore, RB-ARQ needs to be extended for such BCIs.

Let K be a set of thoughts, and $\boldsymbol{x}^t \in \mathbb{R}^p$ be the feature vector at time t. Then, the posteriori probability, $P(k|X^T)$, for $k \in K$ given X^T, or the likelihood ratio can be obtained as follows:

$$P(k|X^T) = \frac{\pi_k p(X^T|k)}{\sum_{l \in K} \pi_l p(X^T|l)}, \quad (5)$$

$$= \frac{\pi_k \prod_t f_k(\boldsymbol{x}^t)}{\sum_{l \in K} \pi_l \prod_t f_l(\boldsymbol{x}^t)}. \quad (6)$$

According to the Bayes decision theory, the label is estimated as the k whose likelihood ratio is maximum as

$$\hat{u} = \arg\max_k P(k|X^T). \quad (7)$$

Similar to the absolute value of LLR, the maximum likelihood ratio (MLR) also means the reliability; hence, only if the MLR is larger than a given threshold τ in (8), the label is estimated; otherwise the system requests the same thought.

$$\max_k P(k|X^T) \geq \tau \quad (8)$$

III. Experiments

A. Comparison of Methods by Simulation

The proposed method was compared with two error control methods: Constant ARQ, requesting n times constantly regardless of the reliability; and Basic RB-ARQ, taking the likelihood ratio obtained from the following $P(k|X^T)$,

$$P(k|X^T) = \frac{\pi_k f_k(\boldsymbol{x}^T)}{\sum_{l \in K} \pi_l f_l(\boldsymbol{x}^T)}. \quad (9)$$

Note that Basic RB-ARQ does not take the past samples X^{T-1} into account.

In this experiment, $f_k(\boldsymbol{x})$ obeyed 2-dimensional Gaussian distributions with the following parameters,

$$\boldsymbol{\mu}_k = r\left(\cos(2\pi k/|K|), \sin(2\pi k/|K|)\right), \quad (10)$$

$$\Sigma_k = I, \quad (11)$$

where I denotes identity matrix; the number of classes was three (i.e. $|K|$=3); and it was assumed that each sample $\boldsymbol{x}$ took one second to measure. In the cases of both $r = 1$ and $r = 1.5$, the three methods were applied to LDA to classify samples distributed normally.

B. Application to actual EEG data

In order to validate the effectiveness of the proposed method, the three methods were applied to one of the EEG data in BCI Competition, which was data set V in BCI Competition III [9].

1) Data Description: The data set contains EEG data recorded from three subjects, when they were doing the following tasks:

- imagination of left-hand movement
- imagination of right-hand movement
- generation of words beginning with certain letters

And it includes four sessions: the first three as training data, and the last as test data. They are provided in two styles: raw EEG signals, and precomputed features; and the latter one was employed in this experiment. Every 62.5 msec (i.e. 16 Hz), the power spectral density (PSD) of the last second was estimated with a resolution of 2 Hz from 8-30 Hz for 8 channels (C3, Cz, C4, CP1, CP2, P3, Pz, and P4 [11]); therefore, the preprocessed data lied on 96 dimensional space. The number of the samples in one session was around 3,500.

2) Applications of error control methods: Considering a rule in the competition, which required us to make outputs of classifications every 0.5 sec (i.e. 2 Hz), the outputs were made longer than that. To make sure this, consecutive 8 samples in the provided data were averaged; thus, the number of the averaged samples per session was around 440. Taking account of the non-stationarity of EEG, the data set of the 3rd session was solely used as training data for each subject; and that of the 4th session was used to test the classifier. LDA was employed as a classifier; and RB-ARQ, Basic RB-ARQ, and Constant ARQ were applied as error control methods.

3) Application of RB-ARQ to tuned classifier: Since the proposed method, RB-ARQ, is an error control method, it can be combined with any classifier which assumes data obeys a certain probability distribution. Therefore, if the proposed method is applied to a "tuned" classifier, the combined system could accomplish higher performance. Fortunately, the website of the BCI competition also provides the results and descriptions of each method; for the data set V, there were twenty methods ranked by their performances, and the 5th of them was selected here[2]. According to the description[3], the channels were chosen as Table I, and LDA was utilized as a classifier. Unlike the cases of III-B.2, samples were not averaged; thus, to provide outputs longer than every 0.5 sec, either of the following conditions needed to be satisfied in addition to (8),

$$T \in \{n \in \mathbb{Z} | n \geq 8\}, \quad (12)$$

$$T \in \{n \in \mathbb{Z} | n \geq 8, n \bmod 8 = 0\}. \quad (13)$$

[2]The first three of them utilized classifiers which does not assume probability distributions such as Support Vector Machine (SVM), and the details of the 4th one were not clearly described, though it used LDA.

[3]`http://ida.first.fhg.de/projects/bci/competition_iii/results/martigny/IreneSturm_desc.txt`, (accessed Jan. 24th, 2009)

TABLE I

SELECTED CHANNELS

S1	C3, Cz, C4, CP1, and CP2
S2	C3, Cz, C4, and CP1
S3	C3, Cz, C4, CP1, CP2, and Pz

TABLE II

EXCERPT FROM THE RESULTS FOR THE DATA SET V

Rank	S1	S2	S3	AVG
1	79.60%	70.31%	56.02%	68.65%
⋮		⋮		⋮
5	78.08%	63.83%	52.75%	64.91%
*	68.72%	61.52%	41.74%	57.33%

Note that the time required for one output can be 0.5 sec, 0.5625 sec, etc. in the case of (12); and it can be only 0.5 sec, 1.0 sec, etc. in the case of (13).

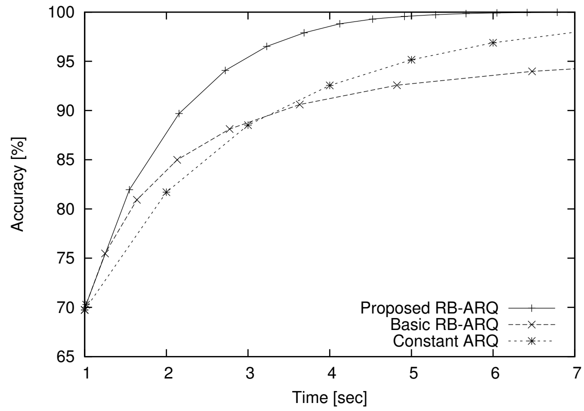
IV. RESULTS AND DISCUSSION

Figure 1 describes the relationships between accuracy (14) and transmission time (i.e. speed) at (a) $r = 1$ and (b) $r = 1.5$; each plot means the result from different threshold (RB-ARQ and Basic RB-ARQ), or the number of requests (Constant ARQ).

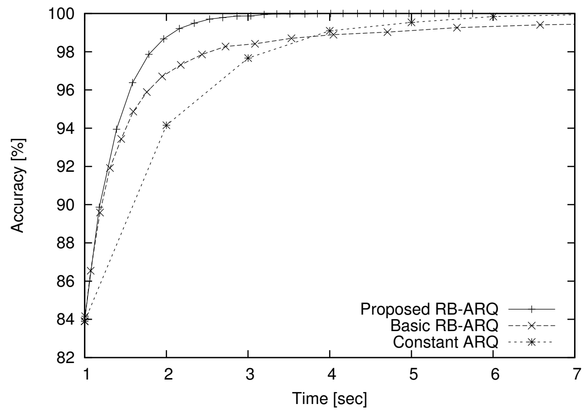
$$\text{Accuracy} = \frac{\text{number of correct classifications}}{\text{number of classifications}} \times 100 \quad (14)$$

It shows that the longer the transmission time is, the better the accuracy is; and the accuracy of (b) is better than that of (a), because in the case of (b) the difference between the mean of each class is larger than that of (a). It also clearly shows that RB-ARQ is roughly superior to the others in terms of both performances, though the accuracy of RB-ARQ at the time less than 1.5 sec is almost equal to that of Basic RB-ARQ in the case of (b).

Table II shows the excerpt from the results for the data set V on the website, including the one (*) implemented according to the description. Figure 2 shows the relationships between accuracy and transmission time for each subject when each error control method was applied. It tells that combining the tuned classifier with error control methods surely improved accuracy with an increase in transmission time. Although, according to Table II, the accuracies of the method implemented by the authors was worse than those of the 5th one, which would be due to lack of the description (e.g. it was not clear which data set was used for training), it is important that the proposed method can be combined with any classifier as long as it is based on probability distributions. Figure 2 (a) and (b) also tell that the performance of RB-ARQ was almost equal to that of Basic RB-ARQ, and both of them were clearly superior to that of Constant ARQ; alternatively, the superiority of RB-ARQ to other methods can be seen in Fig. 2 (c) only by close look. This would be because in the cases of (a) and



(a) $r = 1$



(b) $r = 1.5$

Fig. 1. Comparison of Methods (Theoretical Value)

(b), the distances among the means of each class relative to the covariances were relatively larger than that of (c), implied by the fact that the accuracies of (a) and (b) were better than that of (c).

Figure 3 shows the relationships between experimental values and thresholds τ for each subject in the case of RB-ARQ. Theoretically, an expected accuracy is greater than a selected threshold; however, the experimental value was less than the corresponding threshold especially in (c). The reason would be that data were not normally distributed, that they were not stationary, or both; indeed, non-stationarity of EEG has been known empirically for years, and quantified by Shenoy et al. [12]. The possible solutions for this problem would be use of an adaptive learning, or an extraction of Gaussian or stationary features.

Figure 4 shows the results of the application of RB-ARQ to the tuned classifier. The series labeled "16 Hz" obtained from the case using (12), and the one labeled "2 Hz" from the case using (13). It tells that the series "16 Hz" have achieved better performance compared with the one "2 Hz".

This would be because of the time efficiency of the method using (12); for instance, suppose that the condition (8) was not satisfied at $T = 8$ (i.e. 0.5 sec), it could be satisfied at $T = 9$ (0.5625 msec) in the case of (12), while it could not be satisfied until at $T = 16$ (1.0 sec) in the case of (13). This means that it is better not to average samples in use of RB-ARQ; moreover, it implies that it could achieve even better performance using raw EEG signals, which were recorded at 512 Hz sampling rate.

In this paper, the performance of speed was evaluated by the length of data used for one classification. More precisely, however, the delay when a user changes his/her thought also needs to be taken into consideration. To investigate the delay, it is necessary to conduct an experiment with a new scheme, where a user is requested to either change or keep his/her thought according to the error control methods.

V. Conclusions

The authors have proposed Reliability-Based Automatic Repeat reQuest (RB-ARQ), which is an error control method, suitable for Brain-Computer Interfaces (BCIs). In this paper, RB-ARQ was extended to deal with multi-class classification in BCIs; and it was shown that the proposed method, RB-ARQ, was more effective than other error control methods even for multi-class problems. However, due to non-Gaussianity or non-stationarity, the experimental performance was worse than the theoretically expected one. In future works, an adaptive learning, and an extraction of Gaussian or stationary features will be examined in order to solve this problem. Also, it was shown in the experimental results that the higher the temporal resolution of data used for classification was, the better the performance was; thus, it implies that it could achieve even better performance using not down-sampled and precomputed data but raw EEG signals. To investigate the precise performance of the proposed method, it is necessary to conduct an experiment with a new scheme that uses the proposed method.

Acknowledgement

We would like to thank all of the organizers for holding the BCI Competition, and IDIAP Research Institute for providing the EEG data.

References

[1] Y. Tateoka, T. Yoshikawa, T. Furuhashi, and K. Tanaka, "A basic study on electroencephalogram-based control," in *Proc. Joint 3rd Int. Conf. on Soft Computing and Intelligent Systems and 7th Int. Symp. on advanced Intelligent Systems*, 2006, pp. 1959–1962.

[2] N. Birbaumer, A. Kubler, 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. Rehab. Eng.*, vol. 8, no. 2, pp. 190–193, 2000.

[3] R. Scherer, G. Muller, 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] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2001.

[5] G. Santhanam, S. I. Ryu, B. M. Yu, A. Afshar, and K. V. Shenoy, "A high-performance brain-computer interface," *nature*, vol. 442, pp. 195–198, 2006.

[6] 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.

[7] J. M. Shea, "Reliability-based hybrid arq," *IEE Electronics Letters*, vol. 38, no. 13, pp. 644–645, 2002.

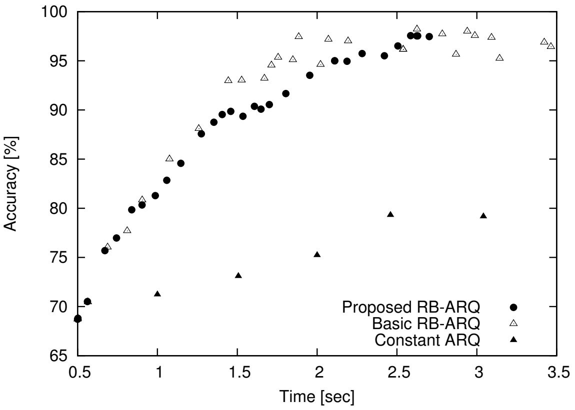
[8] G. Dornhege, "Increasing information transfer rates for Brain-Computer interfacing," Ph.D. dissertation, University of Potsdam, 2006.

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

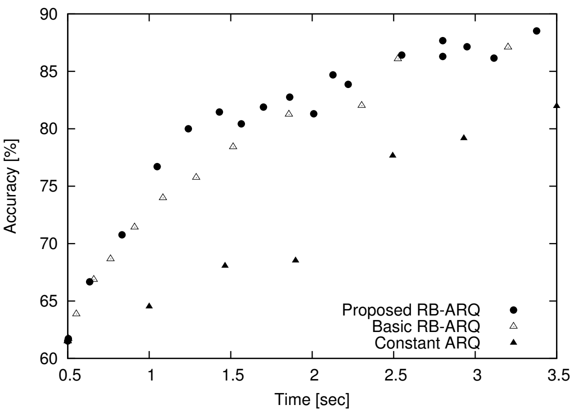
[10] H. Takahashi, T. Yoshikawa, and T. Furuhashi, "A study on application of reliability based automatic repeat request to brain computer interfaces," in *Proc. 15th Int. Conf. Neural Information Processing*. Springer-Varlag, 2009, pp. 1013–1020.

[11] H. H. Jasper, "The ten twenty electrode system of the international federation," *Electroencephalography and Clinical Neurophysiology*, vol. 10, pp. 371–375, 1958.

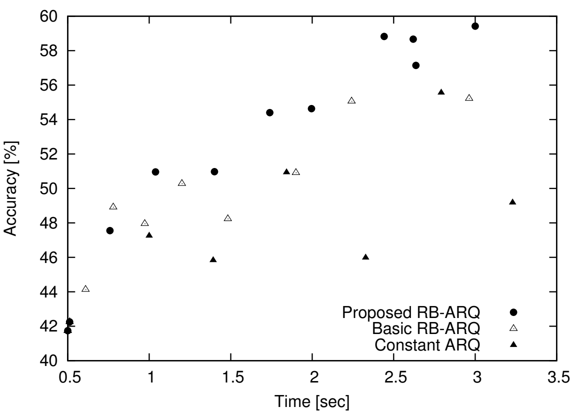
[12] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K. R. Müller, "Towards adaptive classification for BCI," *Journal of Neural Engineering*, vol. 3, pp. R13–R23, 2006.



(a) Subject 1

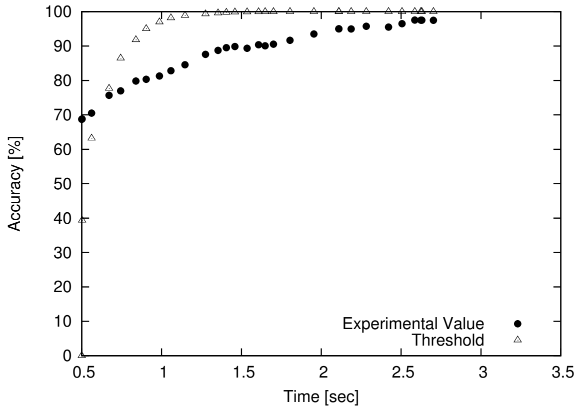


(b) Subject 2

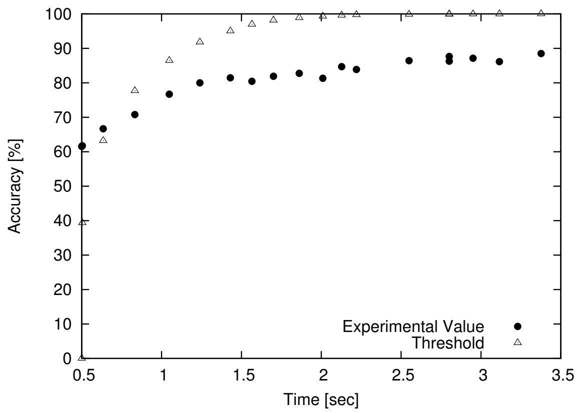


(c) Subject 3

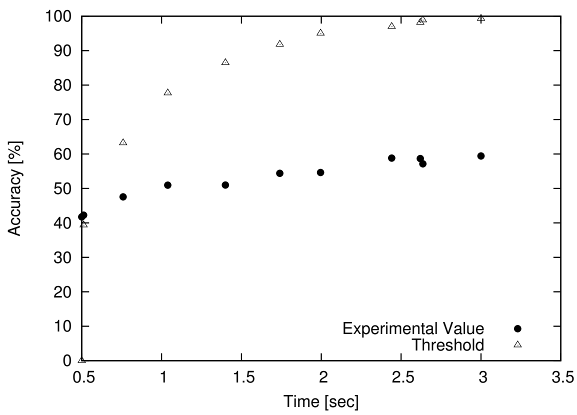
Fig. 2. Comparison of Methods (Experimental Value)



(a) Subject 1

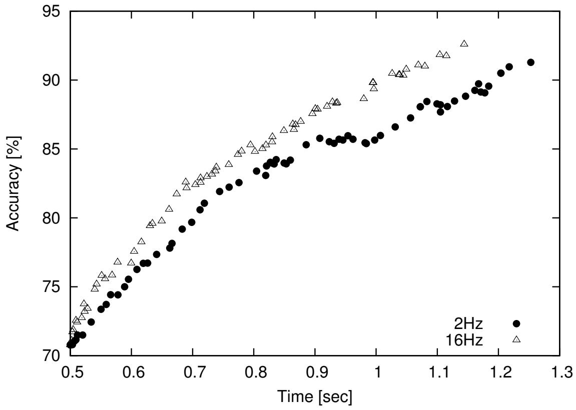


(b) Subject 2

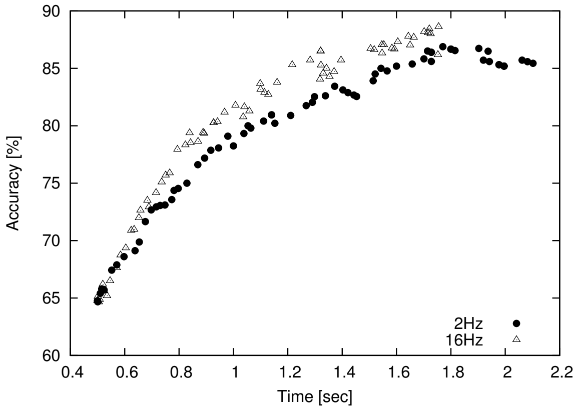


(c) Subject 3

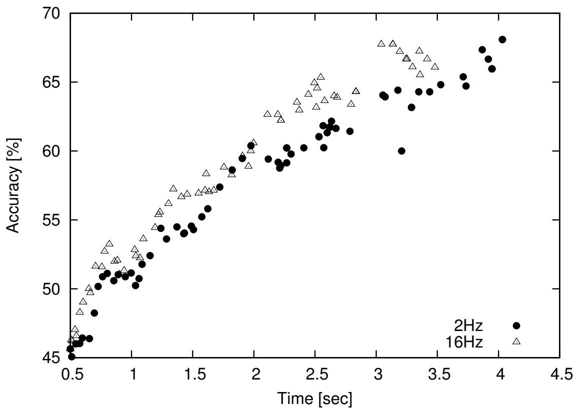
Fig. 3. Experimental Value and Threshold (Proposed RB-ARQ)



(a) Subject 1



(b) Subject 2



(c) Subject 3

Fig. 4. Difference in time resonance