ACSIJ Advances in Computer Science: an International Journal, Vol. 2, Issue 4, No.5 , September 2013
ISSN : 2322-5157
www.ACSIJ.org

Comparative study of data fusion algorithms in P300 Based BCI

Mohammed J. Alhaddad[1], Mahmoud Kamel[2] and Noura Al-Otaibi[3]

[1] **Information Technology Department, King Abdulaziz University**
Jeddah, 21589**, P O Box 80200, Saudi Arabia**
malhaddad@kau.edu.sa

[2] **Information System Department, King Abdulaziz University**
Jeddah, 21589**, P O Box 80200, Saudi Arabia**
miali@kau.edu.sa

[3] **Computer Department, King Abdulaziz University**
Jeddah, 21589**, P O Box 80200, Saudi Arabia**
nmalotaibi@kau.edu.sa

Abstract

Brain-Computer interfaces (BCI) research aims at developing systems that help those disabled people communicating through the use of computers and their brain waves. The BCI researchers put most of their effort on developing new algorithms to improve the speed and accuracy of the prediction mechanisms in BCI applications. For that reason, this study is examine the four combination methods that used for aggregate information form several trials. These methods include Summing Scores, Ensemble Average, Bayesian theory and Dempster Shafer. The main purpose of this study is to improve the speed of prediction mechanism with keep a good classification accuracy. This study was applied on able and disable subjects. Our study result show that the performance of four methods is comparable on able subjects. But the Dempster shafer theory appears best in performance for disabled subjects.

Keywords: *P300,BCI, aggregation ,Dempster Shafer, score, Bayesian theory and ensemble average.*

1. Introduction

Brain Computer Interfaces (BCI) represents a new communication option for peoples who suffering from neuromuscular impairment that prevents them from using conventional augmented communication methods [1]. The BCI system is consisted of different modules in order to transfer brain signal into command: the first module is acquisition tools that record the brain signal. Then preprocessing module that contains algorithm to remove a noise and artifact from signal, then a classification module that decide which action is taken.

There are different kinds of brain activity that can use in a BCI context such as visual evoked potentials (VEP), slow cortical potentials (SCP), mu and beta rhythms and P300 evoked potentials [2]. Here, we concern on P300 wave. P300 is a peaking signal pattern which occurs after the presentation of a rare audio/visual event (see Fig1). . It observed nearly 300 ms after the stimulus onset which gives the name to the signal pattern [3]. Furthermore P300 considered as component of Event related potential depends on rare event. Consequently P300 is possible to use in BCI systems to determine user intentions [4].

There are many applications based on P300 wave. One of most popular applications is speller system for communication and control was first introduce by Farewell & Donchin in [5]. Another applications can be seen in control wheelchair [6], control Internet Browsing [7], Smart home controller [8], Lie detection [9], and implicit emotional tagging of multimedia content [10]. One of these applications is six choice paradigm that introduced in [11]. This application based on oddball paradigm that each image was random intensification. In order to select a target image the subject focused on target elements and counted a number of times that was intensified. The p300 is elicited after intensification the target image [11].

One of the major problems in P300 application is a difficulty to find a P300 response from single trial i.e. (single intensification of each elements). The reason is that measured EEG signals are highly affected by noise. Accordingly is impossible to distinguish the target responses from the non-target ones within a single trial. Hence several trials are perform for the same target element in order to decrease the error in prediction [3]. However there is a tradeoff between the time consuming for predicted a correct element and accuracy of prediction. If number of trials increase the accuracy of prediction is improved but this needs more time of prediction. In consequence the challenge in this application is reduce time of prediction target element with keep a good accuracy. So, this leads to search for methods to aggregate information from several trials with fast converge.

In many studies classification done by averaged several number of trials. This method called ensemble average. It provides good performance, but it is not practical in real application because it reduces system speed [12]. For this purpose, in this paper we compare four methods of aggregation information from different trials in one of p300 paradigm described in [11]. These methods are ensemble averages, scores summation, probability theory and Dempster Shafer theory.

These aggregation methods applied in dataset used in [11]. The dataset acquire form six choice paradigm and p300 response classified by Bayesian linear discriminant analysis (BLDA).

The paper organized as follows: in next section, review of previous studies of aggregation algorithms is present. In third section, the description of data set and the offline

ACSIJ Advances in Computer Science: an International Journal, Vol. 2, Issue 4, No.5 , September 2013
ISSN : 2322-5157
www.ACSIJ.org

analysis is provided. In next section, the results are discussed. In the last section, the conclusion is presented.

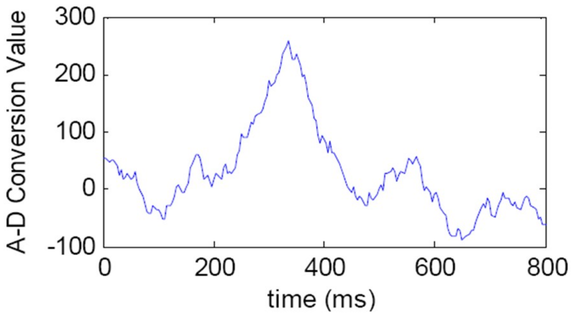

Fig 1:A typical P300 signal. A rising pattern occurs nearly 300ms after the presentation of the target stimulus

2. LITERATURE REVIEW

In BCI context, the most popular methods for aggregated information form several trials are scores summation and ensemble average. In this paper, we compare these methods with other aggregation methods such as Dempster Shafer and Bayesian theory.

In the ensemble average, the classification is done by averaging a number of trials. This approach is a popular approach for aggregated evidence from several trials in BCI context. Ensemble average is found in literature such as Farwall and Dunchin (1988)[5], ERDOĞAN (2009) [3], Xu et al (2004) [13] and Krusienski et al (2008) [14].

In scores summation, multiple trials are combined by summing classifier output for each element in each trial. One of the previous studies used the summing scores is Hoffmann's et al study (2008). They studied six images paradigm, and each image is flashed in a random manner twenty times. The single flash of each image is extracted by some preprocessing techniques and classified by a classifier. The score for each image is summed over trials, and image corresponded to a maximum score is selected as a target [11]. Another study described by Salvaris and Sepulveda (2009) applied on p300 speller system with used discrete-wavelet transform (DWT) as preprocessing technique and an ensemble of Fisher's linear to provide an accumulative score for each element. Afterwards multiple trials are combined by summing the accumulative scores for each element, and the element of the maximum score is selected as a target [15].Summing score techniques also applied in another studied such as Yazdani et al study (2009) [10].

Dempster Shafer Theory (DST) used to combine evidence from different trials by Dempster's rules. Each evidence represented as a formula which is known as a mass function. One of the studied used DST described by Bi et al (2004), they used DST to combine classifiers result for text categorization [16]. other studies described by Zhang et al (2007) used DST to tackle the problem of classification of imperfect data [17]. Boston (2007) developed a signal detection model based on DST that supports classification of a waveform as the signal-present, signal-absent, or uncertain. A performance of DS detector was compared to Fuzzy detector and to Bayesian detector [18]. Only one study used Dempster-Shafer Theory in BCI application Yazdani et al (2009)[19]. Yazdani et al study the performance of using Dempster Shafer theory based KNN classifier in BCI application. In this study, DST used to combine evidence coming from the k nearest neighbors of test examples. In this paper we used DST to combine the evidence from different trials in p300 based BCI application. Unlike Yazdani study, we used DST to combine classifier result to solve the problem of uncertainty of information coming from single trial. Uncertainty coming from low signal to noise ratio of the P300 ERP. So, trials repeated several times to enable to detect P300 signal. Then EEG segment is transferred into scores by some machines learning algorithms which indicates if p300 presents or not. These scores are combined from several trials by DT rules. Yazdani used DST during a classification procedure and help of assigned correct label to incoming data.

In Bayesian theory, the probability of each evidence is combined by using the probability theory. In this approach, the output of standard classifier is converted into a probability and then combine by Bayes rules. Hoffman in his PhD thesis only one used probability approach for aggregated classification result from a number of trials in BCI application. He applied this method in six images paradigm, and each image is flashed several times in random order. A class probability for each single trial (each stimulus) is computed by predictive distribution given by BLDA algorithm. One advantage of this study is that the number of stimuli can be dynamically adapted to the performance of the user and the noise level in the signals. In brief, the adaptive manner can be seen by comparing the maximum of stimulus probabilities with a certain threshold. If the maximum is larger than the threshold, the command associated with large probability is executed, otherwise an extra block of stimuli is presented. Then combine classifier outputs of two presented blocks and compared a maximum result with the threshold [20].

3. METHODOLOGY

3.1 Data set

Aggregation techniques were applied on data set in [11]. The data was recorded from 32 channels as 10-20 system of electrode placement as in [11] and with sampling rate 2048 Hz. The paradigm used to collect data consists of six images each was flashed in random sequences. Each subject data consists of four sessions. Each session consists of six runs. Each runs consisted of 22.5 blocks in average. Each block consists of six image intensification. Each image is intensified for 100 ms, and interstimulus interval was 400 ms. In this study; 20 blocks were used.

3.2 Preprocessing

Preprocessing operations were similar to that performed in [11]. These operation stated as follow:

- Electrode selection. Eight electrode were selected for processing. These electrodes are (Pz,Cz,Fz,Oz,P3,P4,P7,P8).
- Referencing. Average data from electrodes T7 and T8 were subtracted from other electrodes.

- Filtering. A six order forward-backward Butterworth band bass filter was used to filter selected data. cutoff frequencies was set to 1 and 12 Hz.
- Downsampling. a data from selected electrode were downsampled to 32 Hz.
- Single trial extraction. single trial extracted at stimuli onset and ended after 1second from stimuli onset.
- Windsorizing. was used to reduce the outliers effect in signal amplitude. The amplitudes value of sample from each selected electrodes were lying above 90th and below 10th were set to those value 90th and 10th respectively.
- Normalization. Samples from each electrode were normalized to interval [-1,1].
- Feature extraction. The features vector were constructed by concatenate the samples from selected electrodes. The dimensionality of feature vectors was $N_e \times N_s$ where N_e a number of selected electrodes and N_s number of sample in each trials.

3.3 Classification

The classification algorithm used in this paper is similar to classifier used in [9] which is BLDA. The classifier was trained in three sessions and tested on one left session. In, summing score, Dempster Shafer and Bayesian theory the single trial was classified. This leads to twenty blocks of classifier output. Each block consists of a six classifier output (represents a score for six images). One of those images was selected as a target. The mechanism of selected target image is different depended on which aggregation methods used. In the ensemble average method, the average over the blocks was classified. Then the output of classifier is only one averaged block with six images.

3.4 Aggregation methods

Aggregation methods used to combine a classifier output from different trials. Because the EEG high affected with noise, and system may cannot take a decision from single trial, multiple trials are necessary applied. Consequently, after classification single trial, we need to aggregate classifier results from several trials. The description of those methods as follows:

3.4.1 Ensemble average

Ensemble average means compute the average of signal wave, i.e. the signal with same stimulus is averaged. Average signal is used in training data to compute the classifier vector. The ensemble's data from each runs was concatenated on each session. Then the classifier vector was computed by BLDA. In testing phase, the same procedure was done. The average of data is computed, and then average data was classified. A result of classifier is one averaged block. The image with the maximum score is selected as a target. Fig 2 shows the procedure of the ensemble average approach.

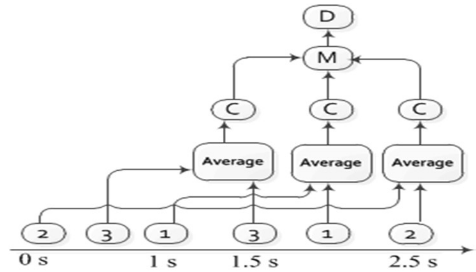

Fig 2: In the ensemble average approach the signal wave with same stimulus is averaged. As an example, there are three different stimuli presented in random order with ISI 500 ms. A number of ensembles are two, so each two similar consecutive stimulus is averaged. In testing phase only two blocks are averaged. Subsequently, average stimulus is classified by classifier. Afterwards, a maximum classifier output is computed by M and command associated with the maximum score is executed (D).

3.4.2 Scores summation

In this method the single trial is classified. Then the result of classifier is aggregated from different trials by summing the classifier output over a blocks of each image. Then the image with maximum classifier output was selected as a target. Fig 3 show a description of score summation approach.

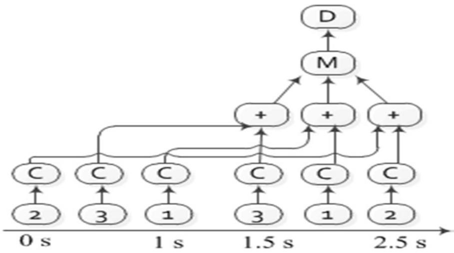

Fig 3: score summation approach means the command is executed after a presentation of a fixed number of blocks. As an example, suppose three different stimuli are flashing in random order with ISI 500 ms. The EEG correspond to each stimulus is classified by classifier. Then classifier results are summed over the blocks (i.e. a classifier result of first block is summed with the classifier result of second block and so on). After that maximum summed output is computed (represent by M letter). Finally, a decision is taken (represent with letter D) by selected a stimulus with the maximum score as a target.

3.4.3 Dempster Shafer

Dempster Shafer (DS) also known theory of evidence was introduced by G.Shafer in 1976 [21]. DST is one of the main tools for reasoning about data obtained from multiple sources, subject to uncertain information [22]. The goal of DST decreases the uncertainty by accumulated the evidence from several sources [22]. It can be considered as a generalization to the Bayesian theory that deals with probability mass functions [23]. In the following, a brief review of terminology and notations of evidence theory are given.

Let $\Theta = \{\theta_1, \ldots \theta_N\}$ be a set of N finite hypothesis. This set is referred as frame of discriminant. The power set of Θ 2^{Θ} (is the collection of all subset of Θ).

A basic probability assignment BPA also called a mass function is a function m: $2^{\Theta} \rightarrow [1,0]$ and which satisfied a following constraints [24]:

$$m(\emptyset) = 0 \quad \text{and} \quad \sum_{A \in 2^{\Theta}} m(A) = 1 \qquad (1)$$

A non-zero subset of Θ is called a focal elements. The body of evidence is set of all focal elements and the union of all the focal elements is called a kernel of mass function m [24]. Another functions defined by Shafer is belief and plausibility. It is considered two measures over the subsets of Θ as follows:

$$\mathrm{Bel}(A) = \sum_{B|B \subseteq A} m(B) \qquad (2)$$

$$\mathrm{Pl}(A) = \sum_{A \cap B \neq \emptyset} m(B) \qquad (3)$$

The belief function Bel(A) measures the total amount of probability that must be distributed among the elements of A. It constitutes a lower limit function on the probability of A[24]. While Pl(A) denotes the extent to which we fail to disbelieve A [24]. The relation between these function:

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\overline{A}) \qquad (4)$$

where $\overline{A}$ is the classical complement of A.

The precise probability of an event (in the classical sense) lies within the lower and upper bounds of Bel and Pl, respectively [25].

$$\mathrm{Bel}(A)=P(A)=\mathrm{Pl}(A) \qquad (5)$$

The evidence theory provides a good aggregations tools. A assume m_1 and m_2 is two masses functions formed from information obtained from different sources in same frame of discriminant [24]. Those masses can be combined by Dempster Shafer rules as follow:

$$m_{12}(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1-k} \qquad (6)$$

where $k = \sum_{B \cap C = \emptyset} m_1(B) . m_2(C)$ (7)
k is measure the degree of the conflict between m_1 and m_2. It determined by summing the products of the bpa's of all sets where the intersection is null. This rule is commutative, associative [25].1-k is a normalization factor, if k=0 the conflict is an absence between two sources, whereas k=1 implies a complete contradiction between m_1 and m_2 [24]. The m_{12} represent combination between m_1 and m_2 that carries a joint information from two sources [24].
The DS can generalize for combination more than two sources. The mass function result from combined J independent sources of information S_J defined as follows:

$$m(C) = m_1(S_1) \oplus m_2(S_2) \oplus \quad \ldots \oplus m_J(S_J)$$

$$= \frac{\sum_{\cap_{i=1}^{J} S_j = C} \prod_{i=1}^{J} m_i(S_J)}{\sum_{\cap_{i=1}^{J} S_j = \emptyset} \prod_{i=1}^{J} m_i(S_J)} \qquad (8)$$

where $S_1, \ldots . S_J$ are focal elements.

Here, in p300 application, we propose the use DS rules to combine the classifier result from several trials. After a single trials extracted by preprocessing operations and classified by BLDA, the classifier output is converted into masses function. Then masses function of images are combined by DS rules according to equation (5). The image with the maximum result of combination is selected as a target (see fig4). In this situation, the frame of discriminant includes two elements: target (T) (P300 presence) and non-target (N) (P300 not presence). The BPA defined for elements T and N in each block, $m_n(T)$ and $m_n(N)$ where n number of blocks. BPA satisfied:

$$m_n(T)+m_n(N)=1 \quad \text{and} \quad m_n(\emptyset) = 0$$

Suppose for example *n=2*, the combination rules from block 1 and block 2 for target element done by used equation (6). Since $T \cap N = \emptyset$, the numerator contain only one term.

$$m_{12}(T) = \frac{m_1(T). m_2(T)}{1-k}$$

where k contain the sum of all non-empty intersection such that:

$$k = m_1(T). m_2(N) + m_1(N). m_2(T)$$

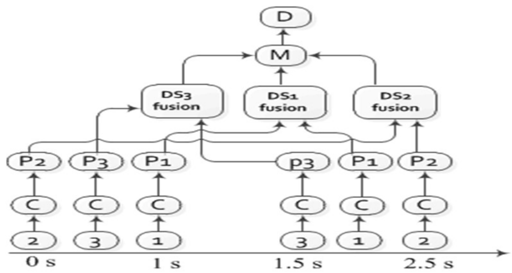

Fig 4: DS used for aggregate a classifier result from several blocks. Suppose for an example, there are two blocks, each with three stimuli and ISI 500 ms. After intensification of each element in block the corresponding EEG signal were classified (C). Then the classifier result was converted into probability by leave one out approach (P). After that a probability of each element from block 1 is combined with corresponding elements in block 2 by DS combination rule as equation (6). Finally, the elements correspond to maximum result of combination was selected as a target.

Here we provide the numerical example to clarify the concept of DS combination rule. This example combines the information from two blocks in p300 based BCI paradigm. Suppose in block1 classifier decide element is target with probability 0.9 and non-target with probability 0.1 (denoted $m_1(T)$and $m_1(N)$ respectively). A second block consider element as target with probability 0.8 and non-target with probability 0.2 (denoted $m_2(T)$and $m_2(N)$ respectively). The combination rule of mass functions from two blocks summarized in *Table 1.*

Table 1: Dempster Rule combination of block1 and block2

			Block 2	
Block 1	Element class		T	N
		m1 \ m2	0.8	0.2
	T	0.9	m1(T).m2(T)=0.72	m1(T).m2(N)=0.18
	N	0.1	m1(N).m2(T)=0.08	m1(N).m2(N)=0.02

The joint information $m_{12}(T)$ combined as:

$$m_{12}(T) = \frac{0.72}{1-0.26} = 0.973$$

$$m_{12}(N) = \frac{0.02}{1-0.26} = 0.027$$

The bpa for target element is 0.97 which correspond to bel(T)=0.97. Consequently we can consider the element is a target because it belief function is maximum and nearest to 1.

3.4.4 Bayesian theory

Bayesian approach relyed on the probability distribution/density functions to express data uncertainty[23]. Before discussion of using Bayesian theory for fusion, we provide a review for basic notation of probability theory.
A general form of Bayes theory is

$$P(w_i|x) = \frac{P(x|w_i)P(w_i)}{P(x)} \quad (9)$$

Where x is random variable, $w_i \in \{w_1, \dots w_n\}$ is finite state of n categories. $P(x|w_i)$ Is a conditional probability density function for x given class w_i.Also this term called likelihood, $P(w_i)$ is prior probability of class w_i. $P(w_i|x)$ is posterior probability. $P(x)$ is evidence factor (unconditional measurement probability density) that can express in terms of the conditional probability distributions as:

$$P(x) = \sum_{i=1}^{n} P(x|w_i)P(w_i) \quad (10)$$

if a different measurements x used with assumption each representation is a conditional independent, we can write a joint probability distribution as follow:

$$P(x_1, \dots, x_R|w_i) = \prod_{k=1}^{R} P(x_k|w_i) \quad (11)$$

where $P(x_k|w_i)$ is the measurement process model of the kth representation. We can rewrite equation (4.25) to include different measurements as:

$$P(x_1, \dots, x_R) = \sum_{i=1}^{n} P(x_1, \dots, x_R|w_i)P(w_i) \quad (12)$$

By substitute (12) and (11) into (9):

$$P(w_i|x_1, \dots, x_R) = \frac{P(w_i)\prod_{k=1}^{R} P(x_k|w_i)}{\sum_{i=1}^{n} P(w_i)\prod_{k=1}^{R} P(x_k|w_i)} \quad (13)$$

Equation (13) is used for fusion data from multiple information sources[22][26].

Here, in p300 application a Bayesian combination is used to aggregate classifier output from several trials. First, a result of classifiers is converted into class probability. A class probability is computed by used a Leave-One-Out Approach as mention in [20]. After probability is computed for each classifier output, the probability for a sequence of class label is computed as follows:

$$p(\hat{w} = w_1 \dots w_T|x_1 \dots x_T) = \frac{\prod_{i=1}^{T} P(\hat{w}=w_t|x_t)}{\sum_{I=\pounds}\prod_{i=1}^{T} P(\hat{w}=l_t|x_t)} \quad \text{for}$$

$w \in \pounds$ (14)

where I=$l_1, \dots, l_T$ is a sequence of labels of length T, ℓ is set of all possible class label sequence of length T [20]. After that the image with maximum probability is selected as a target. Fig 5 show the Bayesian approach for combined information from several trials.

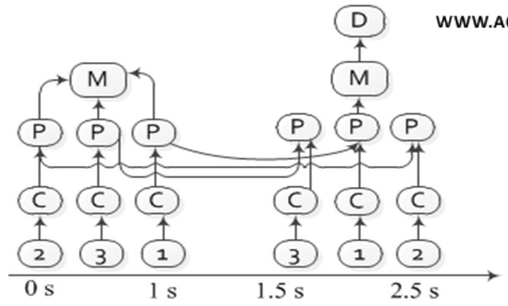

Fig 5: Bayesian theory used for aggregate data from two blocks. Each block contains three different stimulus with ISI 500 ms.Each stimulus is classified by classifier then class probability is computed. After that, we compute a probability of sequence by Bayesian formula. A stimuli with maximum probability is selected as target element.

4. Conclusion

A classification accuracy and corresponding bit rate that achieved by BLDA are averaged over sessions and over 3 subjects. Fig 6 and Fig 7 present average accuracy and bit rate for each aggregation method (ensemble average, score summation, Dempster Shafer and probability theory) for disable and able subjects respectively. Table 1 and Table 2 show the average classification accuracy of three disable and able subjects respectively.

The ensemble average appears a worst method for disable subject because classification accuracy not exceeds 97% in average after 20 blocks. In the same way, it reach to 98.6% after 20 blocks for able subject.

The summing score technique achieved 100% accuracy in average after 12 blocks (i.e. 28.8 s) for disable subject. However, it reach to 98.6% for able subject. One advantage of summing score is simplicity.

A probability theory provides a good result for able subject. it achieved 100% accuracy after 12 blocks (i.e. after 28.8s) and 98.6 for able one same as score summation technique.

A Dempster Shafer theory achieved 100% accuracy after 11 blocks for disable subject. On the other hand, it reached to 98.6% after 20 blocks for able one. Clearly, DS theory faster converges more than other methods see Fig8. Even though, DS achieves the reasonable results, but it has one drawback that is computational complexity. It takes a long time to combine the result from several blocks. The complexity proportional increased based on the number of blocks. Because it needs intersections, where N is number of blocks. On the other hand, Fig 9 represent the average classification accuracy for three able subject. From this figure, it clearly all four methods achieve a comparable results.

Table 1: Average classification accuracy for three disable subjects.

	Ensemble Average	Summing scores	Bayesian Theory	Dempster Shafer
Block 1	0.402778	0.402778	0.208333	0.402778
Block 2	0.680556	0.652778	0.513889	0.666667
Block 3	0.75	0.763889	0.680556	0.75
Block 4	0.763889	0.861111	0.805556	0.861111
Block 5	0.888889	0.930556	0.902778	0.930556
Block 6	0.958333	0.944444	0.916667	0.916667
Block 7	0.944444	0.972222	0.958333	0.972222
Block 8	0.916667	0.972222	0.972222	0.972222
Block 9	0.944444	0.986111	0.972222	0.972222
Block 10	0.930556	0.986111	0.986111	0.986111
Block 11	0.958333	0.986111	0.986111	1
Block 12	0.902778	1	1	1
Block 13	0.916667	1	1	1
Block 14	0.930556	1	1	1
Block 15	0.916667	1	1	1
Block 16	0.9583334	1	1	1
Block 17	0.958333	1	1	1
Block 18	0.972222	1	1	1
Block 19	0.958333	1	1	1
Block 20	0.972222	1	1	1

Table 2: Average classification accuracy for three able subjects.

	Ensemble Average	Summing scores	Bayesian Theory	Dempster Shafer
Block 1	0.763889	0.763889	0.208333	0.402778
Block 2	0.819444	0.833333	0.833333	0.833333
Block 3	0.902778	0.888889	0.888889	0.888889
Block 4	0.930556	0.958333	0.958333	0.958333
Block 5	0.930556	0.930556	0.944444	0.944444
Block 6	0.916667	0.930556	0.930556	0.930556
Block 7	0.944444	0.972222	0.958333	0.958333
Block 8	0.958333	0.972222	0.972222	0.972222
Block 9	0.972222	0.972222	0.972222	0.972222
Block 10	0.972222	0.972222	0.972222	0.972222
Block 11	0.972222	0.972222	0.972222	0.972222
Block 12	0.986111	0.986111	0.986111	0.986111
Block 13	0.972222	0.986111	0.986111	0.986111
Block 14	0.972222	0.986111	0.986111	0.986111
Block 15	0.986111	0.986111	0.986111	0.986111
Block 16	0.986111	0.986111	0.986111	0.986111
Block 17	0.986111	0.986111	0.986111	0.986111
Block 18	0.986111	0.986111	0.986111	0.986111
Block 19	0.986111	0.986111	0.986111	0.986111
Block 20	0.986111	0.986111	0.986111	0.986111

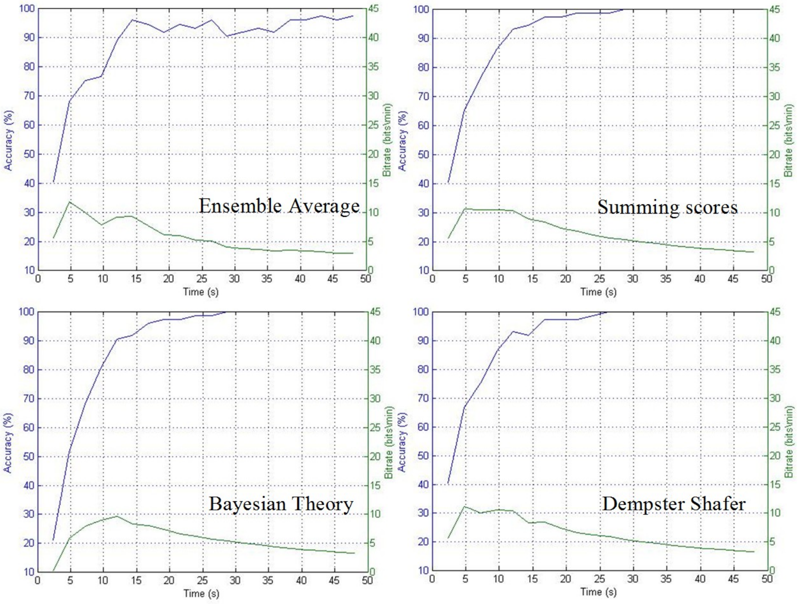

Fig 6: classification accuracy and bit rate plotted Vs. time obtained by BLDA for four combination methods. A classification accuracy averaged over four sessions for disabled subject.

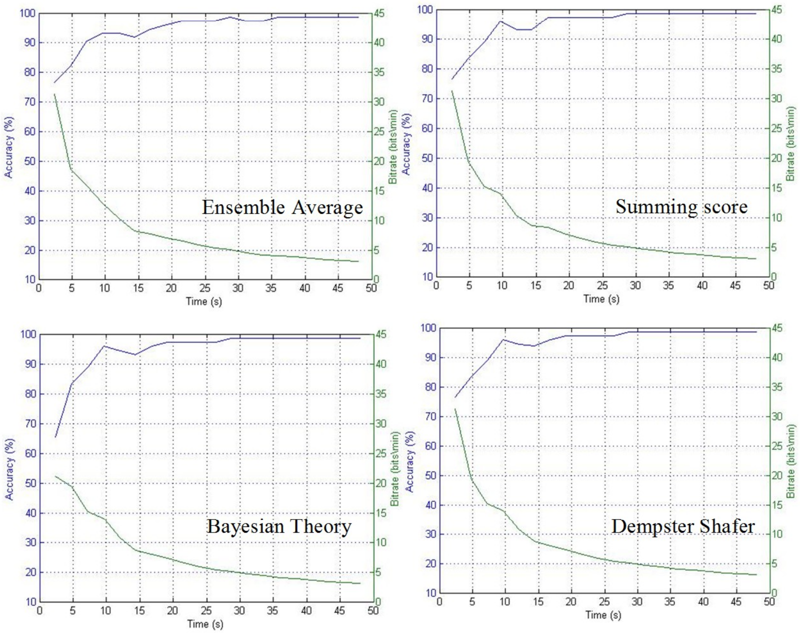

Fig 7: classification accuracy and bit rate plotted Vs. time obtained by BLDA for four combination methods. A classification accuracy averaged over four sessions for abled subject.

Acknowledgements

Many thanks go to all the subjects who volunteered to participate in the experiments described in this research . We would like to thank our team for their efforts in the BCI project. This research was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, The authors, therefore, acknowledge with thanks DSR technical and financial support.

5. Appendix

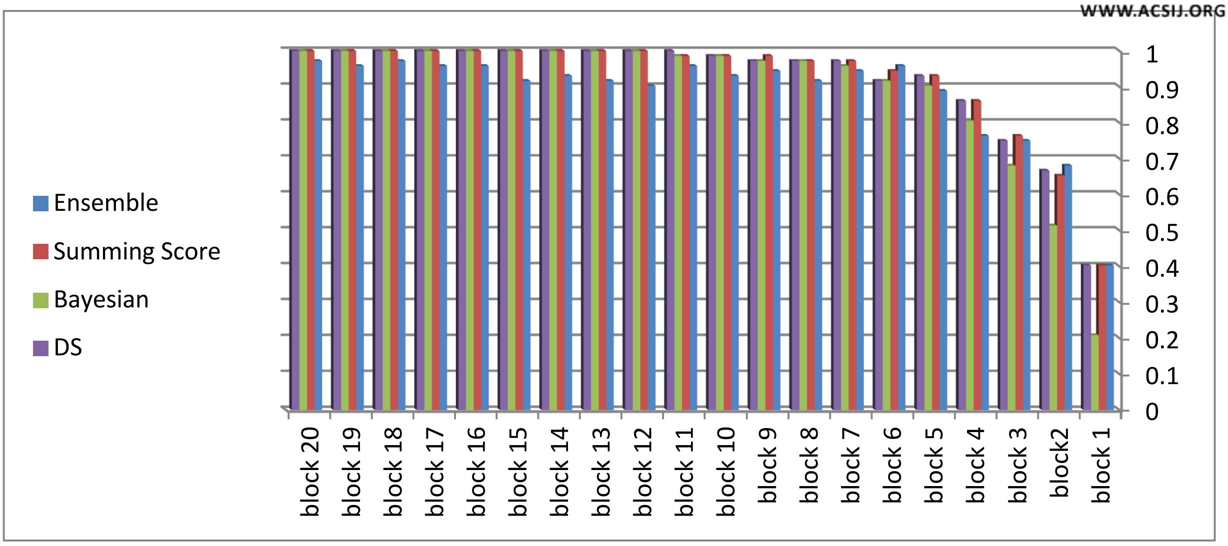

Fig 8: comparisons of average accuracy for four methods of combination over a number of blocks for disable subject.

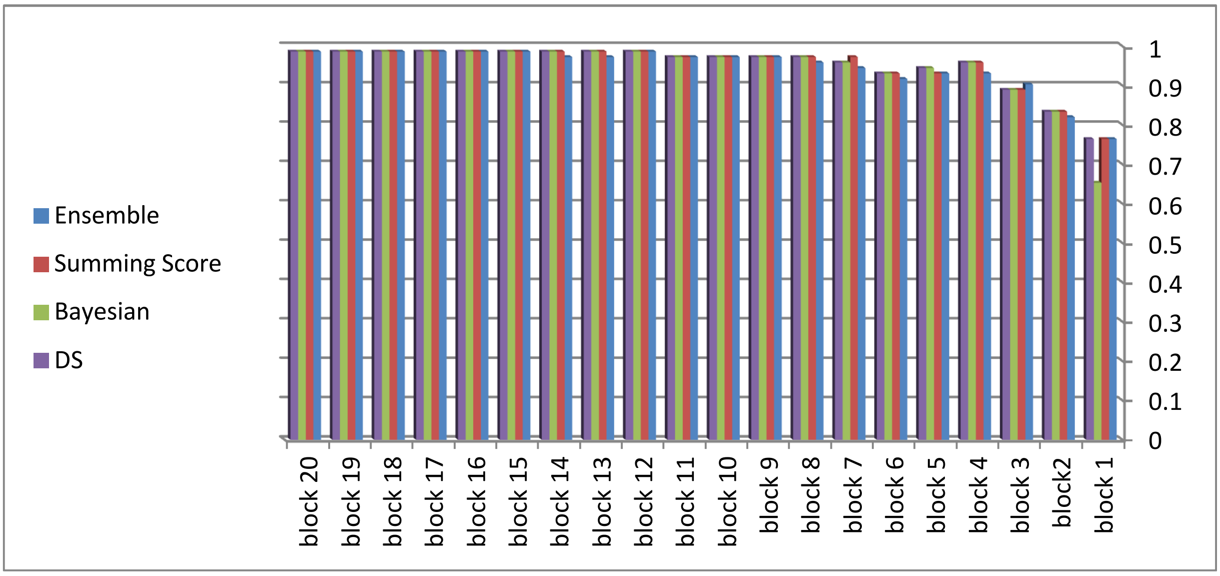

Fig 9: comparisons of average accuracy of four methods of combination over a number of blocks for able subject.

References

[1] F. Beverina, G. Palmas, S. Silvoni, F. Piccione, and S. Giove, “User adaptive BCIs : SSVEP and P300 based interfaces,” PsychNology Journal, vol. 1, no. 4, pp. 331 – 354, 2003.
[2] B. dal Seno, “Toward An Integrated P300 And ErrP Based Brain Computer Interface,” Politecinco di Milano, 2009.
[3] H. Erdogan, “A Design and Implementation of P300 Based Brain Computer Interface,” Middle East Technical University, 2009.
[4] L. Citi, R. Poli, and C. Cinel, “Exploiting P300 amplitude variations can improve classification accuracy in Donchin’s BCI speller,” 2009 4th International IEEE/EMBS Conference on Neural Engineering, pp. 478–481, Apr. 2009.
[5] E. Donchin and L. . Farwell, “Talking off the top of your head: toward a mental prothesis utilizing event-related brain potentials,” Electroencephalogr and clinical Neurophysiology, vol. 70, pp. 510–523, 1988.
[6] J. Minguez, I. Iturrate, J. Antelis, and A. Kubler, “Non-Invasive Brain-Actuated Wheelchair Based on a P300 Neurophysiological Protocol and Automated Navigation,” IEEE Transactions on Robotics, vol. 25, no. 3, pp. 614–627, 2009.
[7] E. Mugler, M. Bensch, S. Halder, W. Rosenstiel, N. Birbaumer, and A. Kübler, “Control of an Internet Browser Using the P300 Event- Related Potential,” International Journal of Bioelectromagnetism, vol. 10, no. 1, pp. 56 – 63, 2008.
[8] C. Guger, “Using a brain / computer interface for smart-home control,” PerAdaMagazine, pp. 1–2, 2009.
[9] K. Abhari, R. Fazel-rezai, and A. Meghdadi, “LIE DETECTION USING BRAIN P300 SIGNAL : PRELIMINARY RESULTS,” in in the 30th Canadian Medical and Biological Engineering Conference, 2007, pp. 2–5.
[10] A. Yazdani, J.-S. Lee, and T. Ebrahimi, “Implicit emotional tagging of multimedia using EEG signals and brain computer interface,” Proceedings of the first SIGMM workshop on Social media - WSM ’09, p. 81, 2009.
[11] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, “An efficient P300-based brain-computer interface for disabled subjects.,” Journal of neuroscience methods, vol. 167, no. 1, pp. 115–25, Jan. 2008.
[12] Y. Güçlütürk, U. Güçlü, and A. Samraj, “An Online Single Trial Analysis of the P300 Event Related Potential for the Disabled,” in 2010 IEEE 26-th convention of Electrical and Electronics Engineers in Israel, 2010, pp. 338–341.
[13] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F. Yang, “BCI Competition 2003--Data set IIb: enhancing P300 wave detection using ICA-based subspace projections for BCI applications.,” IEEE transactions on bio-medical engineering, vol. 51, no. 6, pp. 1067–72, Jun. 2004.
[14] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, “Toward Enhanced P300 Speller Performance,” PMC, vol. 167, no. 1, pp. 15–21, 2009.
[15] M. Salvaris and F. Sepulveda, “Wavelets and ensemble of FLDs for P300 classification,” 2009 4th International IEEE/EMBS Conference on Neural Engineering, pp. 339–342, Apr. 2009.
[16] Y. Bi, D. Bell, H. Wang, G. Guo, and K. Greer, “Combining Multiple Classifiers Using Dempster ’ s Rule of Combination for Text Categorization,” pp. 127–138, 2004.
[17] J. Zhang, L. Weerakkody, and D. Roelant, “A DEMPSTER-SHAFER BELIEF THEORY BASED ALGORITHM FOR CLASSIFICATION OF IMPERFECT,” in IADIS European Conference Data Ming, 2007, pp. 11–16.
[18] J. R. Boston, “A signal detection system based on Dempster-Shafer theory and comparison to fuzzy detection,” IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), vol. 30, no. 1, pp. 45–51, 2000.
[19] A. Yazdani, T. Ebrahimi, and U. Hoffmann, “Classification of EEG Signals Using Dempster Shafer Theory,” in Proceedings of the 4th International IEEE EMBS Conference on Neural Engineering, 2009, pp. 327–330.
[20] U. Hoffmann, “Bayesian Machine Learning Applied in a Brain-Computer Interface for Disabled Users,” 2007.
[21] Glenn Shafer, A Mathematical Theory of Evidence. Princeton University, 1976.
[22] R. H. Enders, A. K. Brodzik, and M. R. Pellegrini, “Algebra of Dempster-Shafer evidence accumulation,” in Proc. SPIE 6968, Signal Processing, Sensor Fusion, and Target Recognition XVII, 2008.
[23] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, “Multisensor data fusion: A review of the state-of-the-art,” Information Fusion, vol. 14, no. 1, pp. 28–44, Jan. 2013.
[24] D. Zeng and J. Xu, “Data Fusion for Traffic Incident Detection Using D-S Evidence Theory with Probabilistic SVMs,” Journal of computers, vol. 3, no. 10, pp. 36–43, 2008.
[25] K. Sentz and S. Ferson, “Combination of Evidence in Dempster-Shafer Theory,” 2002.
[26] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, “On Combining Classifiers,” IEEE Transaction on pattern analysis and machaine intelligence, vol. 20, no. 3, pp. 226–239, 1998.

Author

Prof. Mohammed J. Alhaddad received his Master from Essex University in 2001, and he obtained PhD from the school of Computer Science and Electronic Engineering, University of Essex in 2006, UK. He became Vice Dean of Admission & Registration of Northern Borders University, and Chairman of Information Technology Department at King Abdul-Aziz University. His research interests are: network Security, Artificial Intelligence, Robots, Brain Computer Interface BCI, and Radio Frequency Identification RFID, Data Mining, Semantic Query Optimization, co-operative query answering, distributed databases and Deductive databases.

Ms. Noura Al-Otaibi: Currently Lecturer at King AbdulAziz University, Saudi Arabia. In 2008 have a B. SC. in computer science with first honor degree from Faculty of Computing and information Technology, King AbdulAziz University, Jeddah, Saudi Arabia. My research interests is Brain computer Interface ,Bioinformatics,Cloud Computing,Data Mining and DataBase.