



ELSEVIER

Applied Soft Computing 2 (2002) 48–60

Applied Soft
Computing

www.elsevier.com/locate/asoc

Using genetic algorithm to identify the discriminatory subset of multi-channel spectral bands for visual response

Ramaswamy Palaniappan, Raveendran Paramesran*

Department of Electrical and Telecommunication, Faculty of Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia

Abstract

In this paper, we propose a technique that uses genetic algorithm (GA) with Fuzzy ARTMAP (FA) classifier to identify the discriminatory subset of the feature set for classification of alcoholics and non-alcoholics using brain rhythm extracted during visual stimulus. In the experimental study, the feature set consists of seven spectral power ratios extracted from 61 visual evoked potential (VEP) channels. The seven spectral bands of VEP signals in the range of 2–50 Hz are extracted using constant gain and uniform bandwidth infinite impulse response (IIR) band-pass filters. Spectral power in these bands are obtained using Parseval's time–frequency energy equivalence theorem. The spectral power ratio for each band is obtained by dividing the spectral power of the band with the total spectral power of the channel. Classification experiments using FA and multilayer perceptron-backpropagation (MLP-BP) classifiers are carried out to confirm that the identified spectral power ratios and channels using the proposed technique are discriminatory. The classification results show that the difference of VEP signals between alcoholics and non-alcoholics can be observed using two spectral power ratios in gamma band (37–50 Hz) extracted from seven channels. This fact indicates that gamma band spectral power could be used to show evidence on the lasting effects of long-term use of alcohol on visual response though the studied alcoholics have been abstinent for a minimum period of 1 month.

© 2002 Elsevier Science B.V. All rights reserved.

Index terms: Alcoholics; Feature reduction; Fuzzy ARTMAP; Gamma band; Genetic algorithm; IIR digital filter; Multilayer perceptron-backpropagation; Visual evoked potential

1. Introduction

In any classification application, there is a possibility that some of the extracted features might be redundant. Therefore, it is imperative to devise a method to identify the discriminatory features. By doing so, the computation time is reduced and the implementation of the design becomes simpler. This is especially true with multi-channel brain rhythm data like electro-encephalogram (EEG) and evoked

potential, which are generally extracted from many electrodes located all over the human scalp [2,6,8–11].

In this paper, we propose a method of using genetic algorithm (GA) and Fuzzy ARTMAP (FA) to identify the discriminatory features for classification. Multi-category artificial data simulation is used to show that the proposed method can be used for identifying discriminatory features in any classification applications. Since the proposed method does not require any form of prior statistical transformation, additional computation time is saved. In the method, the fitness function of the GA populations is evaluated using a FA classifier. FA is specifically chosen due to its low training time. This is important to applications

* Corresponding author. Tel.: +60-3-7967-5253;
fax: +60-3-7967-5316.
E-mail address: psar/ravee@fk.um.edu.my (R. Paramesran).

involving GA since GA incurs the cost of evaluating all the populations in each generation.

GA has been used to select features for EEG classification of a brain computer interface [4]. This method used two classifiers, a k-nearest neighbour classifier to evaluate the GA population fitness and LVQ3 algorithm to classify the different mental thought processes represented by EEG. Other than GA, principal component analysis (PCA) has been proposed as a method to select relevant electrodes for EEG classification of hand movements [10]. However, PCA maximises signal representation with minimum features. This might not necessarily maximise classification performance which is, however, the advantage of using GA.

The proposed method is applied to identify the discriminatory visual evoked potential (VEP) spectral power ratios and channels for classification of alcoholics and non-alcoholics. VEP are electrical signals generated by the nervous system in response to visual stimulus. In the past, they have proven to be very useful for diagnosing many different types of illness [6,8,9]. The VEP signals are extracted during the presentation of visual stimulus from 64 electrodes placed on subject's scalp where the subjects are either alcoholics or non-alcoholics (inclusive of those who drink occasionally). The stimulus paradigm is based on studies by Zhang et al. [14,15], where in [14], the authors studied the effects of long-term use of alcohol on visual short-term memory. Other studies have shown a link between gamma band spectral power and visual short-term memory for normal subjects [13]. Unlike studies in [13,14] which focused on visual short-term memory, our study focuses on using visual response to identify the discriminating spectral bands and channels for alcoholics and non-alcoholics.

Constant gain infinite impulse response (IIR) band-pass filters are designed to extract VEP signals in the range of 2–50 Hz comprising of seven spectral bands, namely, delta & theta, alpha, beta1, beta2, gamma1, gamma2 and gamma3. Spectral power in these bands are obtained using Parseval's time–frequency energy equivalence theorem. Spectral power ratio in each band is computed by dividing the spectral power in each band over the total spectral power in the channel.

The classification performance of FA is tested using all the available and the reduced feature set. This is to compare the performance of the reduced feature set with the performance of using all the features.

Similar experiments are carried out using multilayer perceptron-backpropagation (MLP-BP) classifier to show that the reduced feature set is unbiased.

The rest of the paper is organised as follows. Section 2 explains the principles of GA and FA, which are the two elements in the proposed method. Section 3 discusses the proposed method applied to artificial data models. In Section 4, the application of the proposed method to VEP data is discussed. This section covers three subsections. Section 4.1 details the procedures involved in recording the VEP data followed by a summary on the development of the digital filter design to extract VEP power spectral ratio in the seven bands. The complete details of the digital filter design are given in the Appendix A. Section 4.3 discusses the use of the proposed method to identify the discriminatory subset of the VEP feature set in the classification of alcoholics and non-alcoholics. The results of 10-fold cross validation FA and MLP-BP classification using the complete feature and the reduced feature sets are given in Section 5. Section 6 concludes the study.

2. Principles of genetic algorithm and Fuzzy ARTMAP

In this section, the principles of the two elements, GA and FA, used in the proposed method are discussed. The principles of GA are explained, followed by a discussion on FA.

2.1. Genetic algorithm

GA is a family of computational models inspired by evolution and is based on genetic processes of biological organisms. They are adaptive methods, which may be used to solve search and optimisation problems. Over many generations, natural populations evolve according to the principles of natural selection and "survival of the fittest" [3,5]. GA requires fitness or objective function, which provide a measure of performance of the population individuals. The evaluation function must be relatively fast since GA incurs the cost of evaluating the population of potential solutions. This is why FA classification is used to evaluate the fitness function and not other types of neural network like the MLP-BP. To illustrate this fact, if MLP-BP takes 1 h to do an evaluation, then it

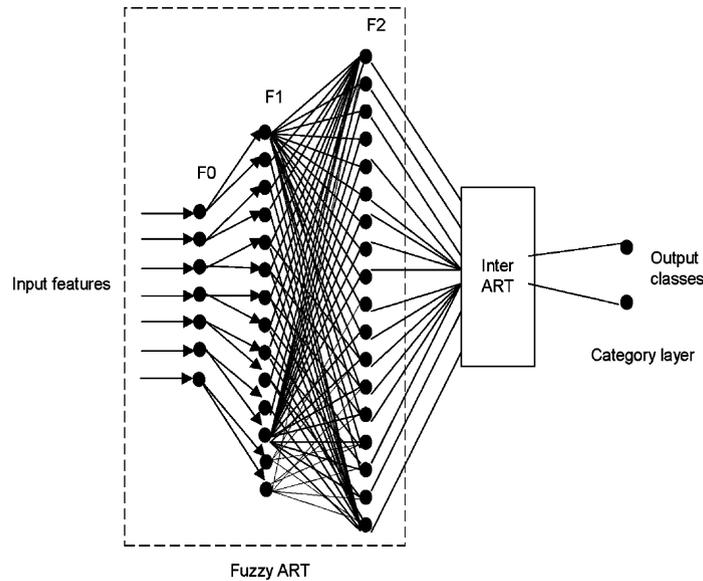


Fig. 1. Fuzzy ARTMAP structure used in this paper.

takes nearly 3 years to complete 25 000 evaluations, which could be just 500 generations of a population size of 50 strings! But FA takes only a fraction of a second to complete an evaluation, so its fusion with GA can produce similar results within 7 h.

2.2. Fuzzy ARTMAP

FA is a type of neural network that performs incremental supervised learning [1]. In this paper, a simplified version of Fuzzy ARTMAP is used [7]. It consists of Fuzzy ART module linked to the category layer through an Inter ART module. During supervised learning, Fuzzy ART receives a stream of input features representing the pattern and the output classes in the category layer are represented by a binary string with a value of 1 for the particular target class and values of 0 for all the rest of the classes. Inter ART module works by increasing the vigilance parameter, ρ of Fuzzy ART by a minimal amount to correct a predictive error at the category layer. Parameter ρ calibrates the minimum confidence that Fuzzy ART must have in an input vector in order for Fuzzy ART to accept that category, rather than search for a better one through an automatically controlled process of hypothesis testing. Lower values of ρ enable

larger categories to form and lead to a broader generalisation and higher code compression. Fig. 1 shows the network structure of simplified FA as used here.

The algorithm for FA training is given later.

Step 1: Fuzzy ART initialisation.

- Input layer F_0 nodes represent a current input vector, $I = (I_1, \dots, I_M)$, with each component I_i in the interval $[0, 1]$, $i = 1, \dots, M$.
- F_1 layer consists of a one-to-one connection with F_0 layer.
- Proliferation of categories is avoided if the inputs are normalised using the method of complement coding. Therefore, the complement coded input I to the field F_1 is the $2M$ -dimensional vector $I = (a, a^c)$ where $a_i^c = 1 - a_i$.
- For each F_2 category node j ($j = 1, \dots, N$), there is a weight vector associated with layer of F_1 nodes, $w_j = (w_{j,1}, \dots, w_{j,2M})$ of adaptive weights.
- Initially we have $w_{j,1}(0) = \dots = w_{j,2M}(0) = 1$, which means that each category is uncommitted.

Step 2: Cluster selection.

- For each input I and F_2 category node j , the choice function T is defined by $T(I) = |I \wedge$

$w_j|/(\alpha + |w_j|)$, where the Fuzzy AND operator \wedge is defined by $(p \wedge q)_i = \min(p_i, q_i)$ and the norm $|\cdot|$ is defined by $|p| = \sum_{i=1}^M |p_i|$ for any M -dimensional vectors p and q .

- A category choice is indexed by J , where $T_J = \max\{T_j: j = 1, \dots, N\}$.
- If more than one T_j is maximal, the category with a smaller index is chosen.
- Resonance occurs if the match function, $|I \wedge w_J|/|I|$ of the chosen category meets the vigilance criterion: $|I \wedge w_J|/|I| \geq \rho$.
- With resonance, weights are updated (step 3).
- Mismatch reset occurs if $|I \wedge w_J|/|I| < \rho$, then the value of the choice function T_J is set to 0 and a new index J is chosen. The search process continues until the chosen J satisfies resonance.

Step 3: Weight update.

- Once the search is completed, the weight vector w_J is updated according to the equation $w_J^{(new)} = (I \wedge w_J^{(old)})$, where fast learning is used.

Step 4: Inter ART mapping.

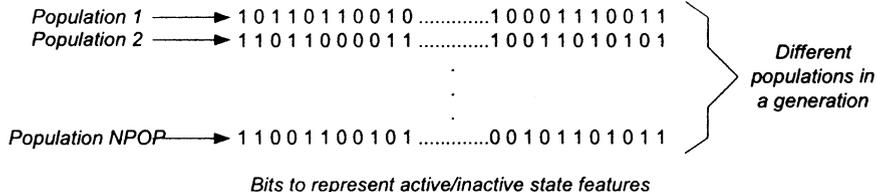
- Create mappings between the two Fuzzy ART modules to correctly learn to predict the classification patterns.

- For all the input patterns presented, it creates a dynamic weight link that consists of a many-to-one or one-to-one mapping between the output layer F_2 of Fuzzy ART and category layer.
- Every time a one to many mapping from Fuzzy ART to category layer is triggered, an error correcting mechanism called match tracking occurs which will increase the vigilance parameter of Fuzzy ART, ρ to a value slightly higher than $|I \wedge w_J|/|I|$ where J is the index of the active F_2 node. This is to avoid any confusion in mapping, and hence, predictions.
- When match tracking occurs, Fuzzy ART search leads either to another category that correctly predicts the target or to an uncommitted new category and the dynamic weight link between the Fuzzy ART modules are updated.
- After this, ρ is set back to the earlier (baseline) vigilance parameter value.

Steps 1–4 are repeated until all the training patterns have been presented. End of algorithm.

The testing stage works similar to the training stage except that there will be no match tracking. This is because the input presented to Fuzzy ART will output a category in layer F_2 , which will be used by the

Table 1
Algorithm for the proposed method

Steps	Property	Description
1	Initial GA populations	Generate NPOP initial populations with NBITS of binary numbers (0 or 1) Each bit represents each of the feature Bit 0 denotes deactivation, bit 1 denotes activation of the feature 
2	Population fitness calculation	Fuzzy ARTMAP training and validation (testing) using training and validation (testing) data $\text{Fitness}_{\text{population}} = (\text{Pattern}_{\text{correct}}/\text{Pattern}_{\text{total}}) + (\text{NBIT}_0/\text{NBITS})$
3	Next generation GA populations are generated	GA operators: reproduction, crossover, mutation are applied based on population fitness from step 2
4	Iteration	Repeat steps 2 and 3 until maximum iteration or population convergence

$\text{Pattern}_{\text{correct}}$ represents FA correctly classified patterns; $\text{Pattern}_{\text{total}}$ represents total patterns in the validation data; NBIT_0 represents number of inactive features; NBITS is the total number of features.

Inter ART module to trigger the corresponding category layer node that refers to the predicted class.

3. The proposed method applied to artificial data models

In this section, the proposed method that combines GA and FA to identify the discriminatory features for artificial data models is explained. This is to show that the proposed method can be used for any application that requires identification of discriminating features from a full features set.

Table 1 shows the algorithm for the proposed method. Initially, a number of populations (NPOP) are generated as random binary strings (a sequence of 1s and 0s) with number of bits (NBITS) used to represent the active/inactive state of the feature. A value of 1 denotes the activation of the feature (i.e. the feature is used) and a value of 0 denotes deactivation of the feature (i.e. the feature is not used). If 10 features represent each pattern, then 10 bits are required in each population.

Using these populations, active features from the training data are fed into FA to be trained. Since GA requires FA classification performance as a measure of fitness of the population, the performance of this population needs to be validated. The same active features from the validation (testing) data are now used to evaluate the FA performance. This process of training and validation is repeated for all the populations. The fitness function is as given in step 2 in Table 1. The fitness function consists of two ratios: correctly classified patterns divided by the total number of patterns and the number of inactive features divided by the total number of features. Both the ratios are set to equal value. This fitness function serves to maximise classification while equally minimising the number of features.

GA uses the population fitness values in the current generation to generate the populations in the next generation using reproduction, crossover and mutation operators [3,5]. Tournament selection is applied during reproduction from a small pool of populations chosen randomly among the total populations. A two-point crossover is used since they are able to wrap around at the end of the string and, therefore, better than a single point crossover. The crossover probability is set at 0.5 while the mutation is set at a

Table 2
First artificial data model

Steps	Class A	Class B
1	Select ε (integer randomly selected from 0 to 10) Parameter ζ is any value higher than 1.0	
2	For inputs 1–10 (except ε) For input ε Parameter δ is Gaussian noise with mean, $\mu = 0$ and variance, $\sigma^2 = 1.0$	δ $\delta + \zeta$ $\delta - \zeta$
3	Repeat step 2 for a total of 100 patterns	

lower probability of 0.01 to reduce excessive random perturbations. This entire cycle is then iterated for a certain number of maximum generations or until population convergence is reached where more than 80% of the populations are similar. FA is run with vigilance parameter value of 0.0. This is to minimise FA size thereby improving the training speed [1].

The details of the artificial data models are shown in Tables 2 and 3. In the first model, a total of 100 patterns with 10 inputs for a two-class problem are generated with only one discriminatory input, ε ; while in the second model, two inputs ε_1 and ε_2 are discriminatory. In this experiment, though the ratios in the fitness function are varied, the same discriminatory feature(s) are identified. Setting a higher FA classification ratio allows GA to initially select populations that give better classification performance although requiring more features, while setting a higher inactive

Table 3
Second artificial data model

Steps	Class A	Class B
1	Select ε_1 and ε_2 (different integers randomly selected from 0 to 10) Select ζ (any value higher than 1.0)	
2	For inputs 1–10 (except ε_1 and ε_2) For input ε_1 For input ε_2 δ is Gaussian noise with mean, $\mu = 0$ and variance, $\sigma^2 = 1.0$	δ $\delta + \zeta$ $\delta - \zeta$ $\delta + \zeta$
3	Repeat step 2 for a total of 100 patterns	

Table 4

GA parameters for artificial data simulation

Coding of genes	Binary coding
Population size (NPOP)	10
No. of genes (NBITS)	10 bits
Reproduction	Tournament selection (pool of three population)
Crossover	Two-point crossover
Crossover rate	0.5
Mutation	Random mutation
Mutation rate	0.01
Convergence	Maximum 100 generations or population convergence
Population convergence	If 80% of populations are similar

feature ratio allows GA to select populations that have less features but give lower classification performance. But towards the end, these populations are the same provided any one of the ratios is not set to lower than 0.1. This is because the ratio that is close to zero loses its influence on the fitness function. The proposed method using these two data models is run for 20 trials where the parameters of GA are as listed in Table 4. In all the trials, the proposed method identifies the correct input(s) for both the models in less than 50 generations.

4. The proposed method applied to visual evoked potential data

In this section, the proposed method is applied to identify the discriminatory spectral power ratios and channels of VEP signals between alcoholics and non-alcoholics. The recording of VEP data is discussed, followed by signal processing technique used to extract the spectral power ratios. Next, the experimental study is discussed.

4.1. Visual evoked potential data

In this study, VEP data is recorded from 20 subjects from which 10 are alcoholics and 10 are non-alcoholics. The alcoholics have been abstinent for a minimum period of 1 month and are also off all medications for the same period of time. The non-alcoholics are carefully matched for age, socio-economic status

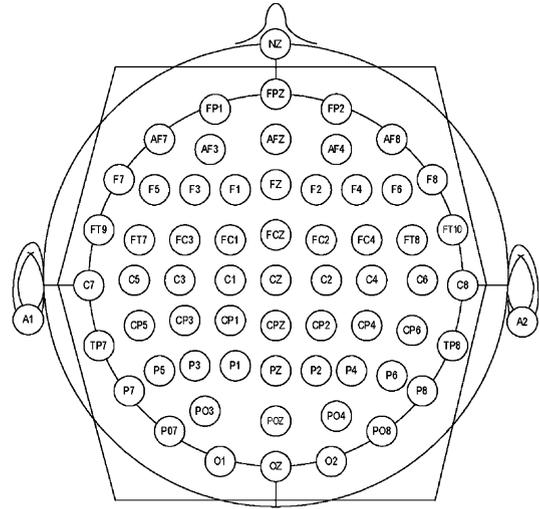


Fig. 2. Sixty-four channel electrode system (channels used outlined by the hexagon).

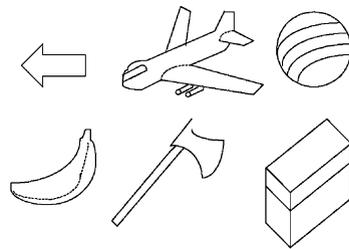


Fig. 3. Some objects from Snodgrass and Vanderwart picture set.

and are not alcohol or substance abusers. Measurements are taken for one second from 64¹ electrodes placed on the subject’s scalp, which are sampled at 256 Hz. The electrode positions (as shown in Fig. 2) are located at standard sites using extension of Standard Electrode Position Nomenclature, American Encephalographic Association. The VEP data is recorded from subjects while being exposed to a single stimulus, which are pictures of objects chosen from Snodgrass and Vanderwart picture set [12]. These pictures are common black and white line drawings like airplane, banana, ball, etc. executed according to a set of rules that provide consistency of pictorial representation. Fig. 3 shows some of these

¹ Data from 61 channels are used to extract features; the other 3 are reference channels.

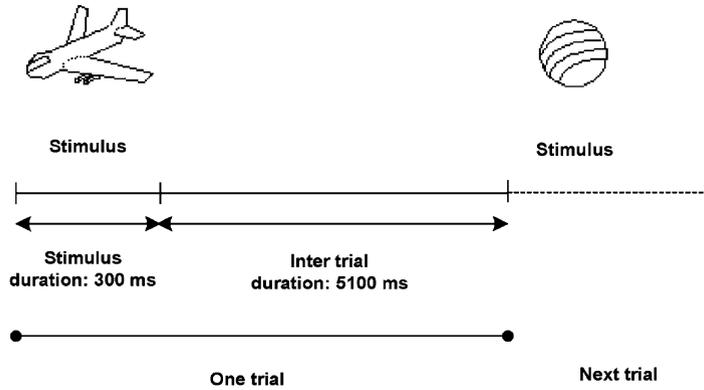


Fig. 4. Presentation of Snodgrass and Vanderwart stimulus.

pictures and Fig. 4 illustrates the presentation of these pictures. Further details of the data collection process can be obtained from [14,15].

For this study, VEP signals with eye blink artifact contamination are removed in the pre-processing stage using a computer program written to detect VEP signals with magnitudes above $100 \mu\text{V}$. These VEP signals detected with eye blinks are then discarded from the experimental study and additional trials are included as replacements. The threshold value of $100 \mu\text{V}$ is used since blinking produces $100\text{--}200 \mu\text{V}$ potential lasting 250 ms [8]. Each subject gives 30 artifact free trials, therefore, giving a total of 600 VEP patterns for the experimental study. The 30 VEP patterns for each subject are then split into four parts randomly. Training and validation (i.e. testing) data sets for GA and FA to select the discriminatory spectral power ratios and channels have five VEP patterns each (for each subject) giving a total of 200 VEP patterns for both these data sets. Training and testing for FA and MLP-BP classifications have 10 VEP patterns each (for each subject) giving a total of 400 VEP patterns.

4.2. Spectral power ratio extraction using IIR digital filters

A series of constant gain and uniform bandwidth infinite impulse response filters are designed to extract VEP signals in spectral range of 2–50 Hz comprising of seven spectral bands. Information about these bands is as listed in Table 5.

The filter is designed using the theory of placing conjugate poles and zeros inside the z -plane unit circle. The transfer function for this filter is

$$H(z) = \frac{z^2 - zr \cos \phi}{z^2 - 2zr \cos \phi + r^2} \quad (1)$$

which expressed recursively in time domain is

$$y(n) = (2r \cos \phi)y[n-1] - r^2y[n-2] + x[n] - (r \cos \phi)x[n-1] \quad (2)$$

A detailed description of the filter design is given in the Appendix A. Using (2), BPF is designed for the different bands listed in Table 5 by changing the value of ϕ for different centre frequencies. A value of $r = 0.85$ is chosen since it gives approximately 3 dB pass-band width of 6 Hz. The spectral power ratio in each band is obtained from the filtered output, $y(n)$

Table 5
Information about spectral bands

Steps	Band name	Centre frequency (Hz)	Approximate 3 dB pass-band (Hz)
1	Delta & theta	5	2–8
2	Alpha	12	9–15
3	Beta1	19	16–22
4	Beta2	26	23–29
5	Gamma1	33	30–36
6	Gamma2	40	37–43
7	Gamma3	47	44–50

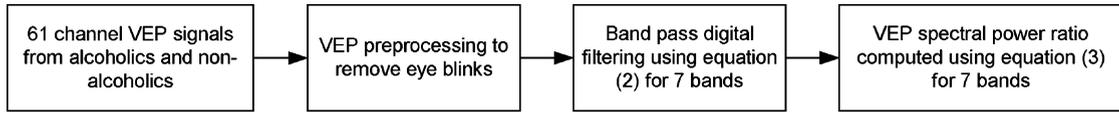


Fig. 5. VEP feature extraction.

by applying Parseval’s time–frequency energy equivalence theorem using

$$\text{spectral power ratio (}i\text{th band)} = \frac{\sum_{n=1}^N [y_i(n)]^2}{\sum_{i=0}^7 \left[\sum_{n=1}^N [y_i(n)]^2 \right]} \quad (3)$$

where N is the total number of data. The seven spectral power ratios from each of the 61 channels are used as features representing the particular VEP pattern. Fig. 5 illustrates the VEP feature extraction.

4.3. Experimental study

The proposed method is applied to identify the discriminatory VEP spectral power ratios and channels for classification of alcoholics and non-alcoholics. The first stage identifies the discriminatory spectral power ratios while the second stage identifies the discriminatory channels. A part of the VEP dataset consisting of 100 alcoholic and 100 non-alcoholic VEP patterns is used here. The feature set consists of seven spectral bands from 61 channels giving a total of 427 features. Half of the data set is used in training while

the remaining data is used in validation (testing). The patterns for training and validation are chosen randomly. Table 6 gives the GA parameters used.

In the first stage, the method is used to select the discriminatory spectral power ratio from the seven spectral power ratios. Therefore, NBITS consist of seven bits to represent the active/inactive state of the spectral power ratios. The method is run three times to ensure that the final GA population is independent of the starting point in search space. In all the three trials, GA gave the same final population that maximises the fitness function given in Table 1, i.e. two spectral power ratios: gamma2 and gamma3 bands. Referring to Table 5, these two bands have centre frequencies of 40 and 47 Hz. Combining both the spectral bands, it can be seen that the 3 dB spectral range of 37–50 Hz contributes towards discriminating alcoholics and non-alcoholics.

In the second stage, only the spectral power ratios of gamma2 and gamma3 bands from each channel are used. Here, the proposed method is used to identify the discriminatory channels. NBITS now consist of

Table 6
GA parameters for identifying discriminatory spectral power ratios and channels for classifying alcoholics and non-alcoholics

Coding of genes	Binary coding
Population size (NPOP)	50
No. of genes (NBITS)	7 bits (first stage), 61 bits (second stage)
Reproduction	Tournament selection (pool of 10 population)
Crossover	Two-point crossover
Crossover rate	0.5
Mutation	Random mutation
Mutation rate	0.01
Convergence	Maximum 500 generations or population convergence
Population convergence	If 80% of populations are similar

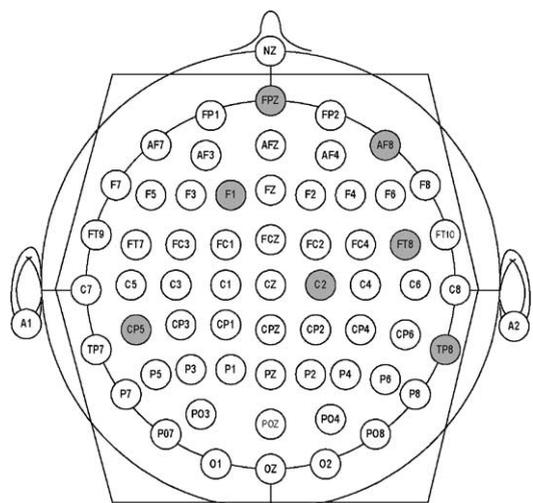


Fig. 6. Channels selected by the proposed method (shaded).

61 bits representing the active/inactive state of each of the channels. Running the proposed method three times identify the same seven channels with locations: CP5, AF8, FT8, FPZ, F1, TP8 and C2. These channel locations are shown shadowed in Fig. 6.

5. Classification results

In this section, the classification results of FA and MLP-BP using the complete and discriminatory VEP

spectral power ratios and channels are discussed. To maintain a certain level of confidence as to the results, a 10-fold cross validation strategy is adopted. A total of 400 VEP patterns are used in this experimental study, where half of the patterns are from alcoholics and the remaining half from non-alcoholics. This VEP dataset is different from the VEP dataset used earlier by GA in Section 4.3 to ensure unbiasedness in the FA and MLP-BP classification. The data set is divided into 10 equal parts, with equal number of patterns from alcoholics and non-alcoholics. Out of

Table 7
Results of FA classification and the time taken for classification of 200 patterns

Vigilance parameter	FA classification (%)	Classification time (s)
(a) Case A: using seven spectral power ratios from 61 channels		
0.0	80.1	2.5
0.1	81.2	2.7
0.2	81.3	2.7
0.3	81.7	2.7
0.4	82.3	3.8
0.5	84.3	5.8
0.6	84.3	10.5
0.7	87.5	17.8
0.8	89.6	30.2
0.9	92.0	56.5
Average	84.4	13.5
(b) Case B: using gamma2 and gamma3 spectral power ratios from 61 channels		
0.0	81.3	0.5
0.1	81.3	0.6
0.2	81.3	0.6
0.3	81.8	0.6
0.4	82.7	1.0
0.5	83.9	1.5
0.6	84.0	2.7
0.7	86.1	4.2
0.8	89.1	7.0
0.9	93.4	13.1
Average	84.5	3.2
(c) Case C: using gamma2 and gamma3 spectral power ratios from seven selected channels		
0.0	78.9	0.1
0.1	78.9	0.1
0.2	78.9	0.1
0.3	78.9	0.1
0.4	79.3	0.1
0.5	79.3	0.2
0.6	83.6	0.3
0.7	84.9	0.4
0.8	86.2	0.6
0.9	89.2	1.2
Average	81.8	0.3

the 10, 5 parts are used in training (200 VEP patterns) while the remaining 5 parts are used in testing (200 VEP patterns). The selection of the parts for training and testing are chosen randomly. FA and MLP-BP classification experiments are repeated 10 times using different parts of the data for training and testing. The average results of these 10 experiments are tabulated.

5.1. FA classification

Table 7(a)–(c) give the results of FA classification and the time taken for classification of 200 patterns. The results are given for varying vigilance parameter values of 0–0.9 in steps of 0.1. All the results noted are averaged values for 10 experiments using the 10-fold cross validation strategy mentioned earlier. Table 7(a) gives the results of using the seven spectral power ratios extracted from 61 channels (case A), while Table 7(b) gives the results using the two discriminatory spectral power ratios (gamma2 and gamma3) extracted from the 61 channels (case B). Table 7(c) gives the results using two spectral power

ratios extracted from seven channels (case C). These results are obtained from simulations carried out on a Pentium III 800 MHz PC with 128 MB RAM and software written in C language.

From Table 7(a)–(c), it can be seen that the average performance of 84.4% is obtained for case A and 84.5% is obtained for case B while 81.8% is obtained for case C. But the average time taken for case C is 0.3 s, which is much faster than 3.2 s for case B and 13.5 s for case A. This is because the input features for case C is 14, while for case B is 122 and for case A is 427. The results also show that gamma band in 3 dB range of 37–50 Hz contribute the most towards classification accuracy.

5.2. MLP-BP classification

Here, MLP-BP classification using the complete and discriminatory VEP spectral power ratios and channels are discussed. This experiment with MLP-BP classifier is run to show that the discriminatory VEP spectral power ratios and channels are

Table 8
Results of MLP-BP classification and the time taken for classification of 200 patterns

Hidden units	MLP-BP classification (%)	Classification time (s)
(a) Case A: using seven spectral power ratios extracted from 61 channels		
20	93.9	1.65
40	93.3	3.24
60	93.6	4.29
80	93.2	6.37
100	92.2	7.58
Average	93.2	4.63
(b) Case B: using gamma2 and gamma3 spectral power ratios from 61 channels		
20	96.0	0.55
40	96.2	1.04
60	97.2	1.38
80	95.2	1.70
100	94.7	2.14
Average	95.9	1.36
(c) Case C: using gamma2 and gamma3 spectral power ratios from seven selected channels		
20	93.5	0.07
40	94.4	0.13
60	95.5	0.18
80	93.9	0.22
100	94.1	0.25
Average	94.3	0.17

unbiased irrespective of the classifier used. The VEP patterns used here are the same as those used for FA classification in Section 5.1. Table 8(a)–(c) give the results of MLP-BP classification and the time taken for classification of 200 patterns. The MLP-BP network is trained with varying hidden units from 20 to 100 in steps of 20. All the results noted are averaged values for 10 experiments using the 10-fold cross validation strategy mentioned earlier. Table 8(a) gives the results using the seven spectral power ratios extracted from 61 channels (case A), while Table 8(b) gives the results using the two discriminatory spectral power ratios (gamma2 and gamma3) extracted from the 61 channels (case B). Table 8(c) gives the results using two spectral power ratios extracted from seven channels (case C).

It can be seen that the best averaged performance of 95.9% is obtained for case B, while for case C, the classification performance is 94.3. The lowest classification performance is obtained for A (93.2%). However, the classification time is lowest for case C, followed by cases B and A, which is the same as FA classification. Comparing the performances of MLP-BP and FA, it can be concluded that the MLP-BP gives better performance than the FA. Classification time (i.e. during testing) using MLP-BP is less than the FA due to the simpler architecture of the MLP-BP, which involves only a single forward propagation. However, the training of the MLP-BP takes a longer time requiring many more iterations than the FA before achieving convergence.

6. Conclusion

In this paper, we have proposed a method using GA combined with a FA classifier to identify the discriminatory subset of the feature set for classification of alcoholics and non-alcoholics using brain rhythm extracted during visual stimulus. Constant gain and uniform bandwidth IIR filters have been utilised in extracting VEP spectral power ratios of seven bands in the range of 2–50 Hz from 61 channels. Experimental results show that the proposed method identifies the discriminatory spectral power ratios and channels that maintain the classification performance. The number of discriminatory spectral bands is reduced from 7 to 2 and the number of channels is reduced from 61 to 7. The discriminatory feature set also improves

computational speed and simplifies the design. The MLP-BP classifier performed better than FA, which shows that the discriminatory feature set is unbiased irrespective of the classifier used.

The discriminatory spectral bands are from the gamma band range of 37–50 Hz. This fact indicates that gamma band spectral power could be used to show evidence on the lasting effects of long-term use of alcohol on visual response though the studied alcoholics have been abstinent for a minimum period of 1 month. The seven channels selected for the discriminatory feature set also denote that the areas of difference between alcoholics and non-alcoholics are localised. Studies by others have shown differences of visual short-term memory for alcoholics and non-alcoholics are in the frontal, central and temporal regions [14]. In our study using visual response, some of the discriminatory channels are located in these regions, too.

Acknowledgements

We acknowledge the assistance of Prof. Henri Begleiter at the Neurodynamics Laboratory, State University of New York Health Centre, Brooklyn, USA

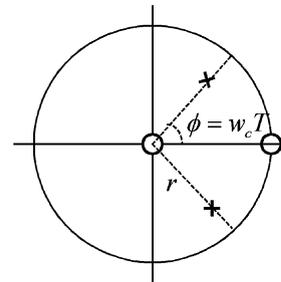


Fig. 7. Pole-zero plot for a band-pass filter.

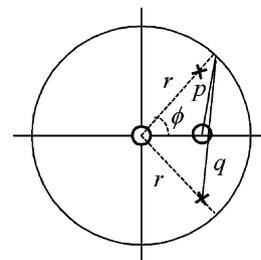


Fig. 8. Pole-zero plot for a constant gain band-pass filter.

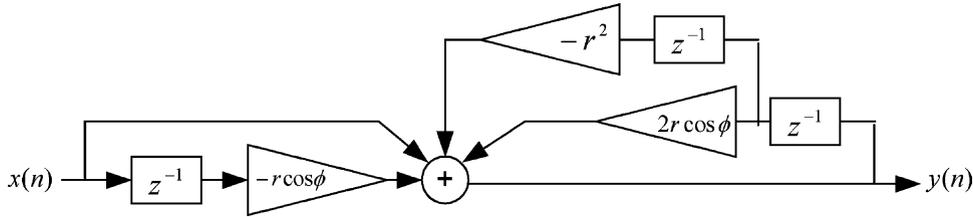


Fig. 9. Hardware implementation of constant gain BPF.

who recorded the raw VEP data and Mr. Paul Conlon of Sasco Hill Research, USA for sending the data to us.

Appendix A.

A general band-pass filter (BPF) with centre frequency, w_c can be designed using the z -domain transfer function:

$$H(z) = \frac{z(z - 1)}{(z - r\phi)(z + r\phi)} \tag{A.1}$$

where $\phi = w_c T$. The pole-zero plot is shown in Fig. 7. The value of r must be less than 1 to ensure stability in computer applications due to bit limitations. However, the gain of this filter changes with the centre frequency. In order to have a constant gain BPF, a design approach based on the method proposed in [9] is used.

Consider the pole-zero plot as shown in Fig. 8. The zero is aligned at the centre of the line joining the two conjugate poles. The gain for this filter can be given as

$$G = \frac{p}{(1 - r)q} \tag{A.2}$$

Since r is generally selected to be close to 1, $p \approx 0.5q$ and the gain can be approximated as

$$G = \frac{1}{2(1 - r)} \tag{A.3}$$

which is independent of the centre frequency, therefore, giving constant gain for a fixed value of r .

The transfer function for this filter is

$$H(z) = \frac{z(z - r \cos \phi)}{(z - r(\cos \phi + j \sin \phi)) \times (z - r(\cos \phi - j \sin \phi))} \tag{A.4}$$

or

$$H(z) = \frac{z^2 - zr \cos \phi}{z^2 - 2zr \cos \phi + r^2} \tag{A.5}$$

which expressed recursively in time domain is

$$y(n) = (2r \cos \phi)y[n - 1] - r^2y[n - 2] + x[n] - (r \cos \phi)x[n - 1] \tag{A.6}$$

Using (A.6), BPF is designed for the different bands listed in Table 5 by changing the value of ϕ for different centre frequencies. The hardware implementation is as shown in Fig. 9. For designing BPF for different spectral bands, only the two multipliers with the function ϕ need to be changed while the rest of the hardware remains fixed. Note that the constant gain need not be divided in (A.6). This is the advantage of using this constant gain filter as compared to using the filter given by (A.1).

References

- [1] G.A. Carpenter, S. Grossberg, J.H. Reynolds, A Fuzzy ARTMAP nonparametric probability estimator for non-stationary pattern recognition problems, IEEE Trans. Neural Networks 6 (6) (1995) 1330–1336.
- [2] D.G. Childers, I.S. Fischler, T.L. Boaz, N.W. Perry, A.A. Arroyo, Multichannel, single trial event related potential classification, IEEE Trans. Biomed. Eng. 33 (12) (1986) 1069–1075.
- [3] D.E. Goldberg, Genetic Algorithm in Search, Optimization and Machine Learning, Addison-Wesley, Reading, MA, 1989.
- [4] D. Flotzinger, M. Pregoner, G. Pfurttscheller, Feature selection with distinction sensitive learning vector quantisation and genetic algorithm, in: Proceedings of IEEE International Conference on World Congress on Computational Intelligence, Vol. 6, 1994, pp. 3448–3458.
- [5] R.L. Haight, S.E. Haupt, Practical Genetic Algorithm, Wiley, New York, 1998.

- [6] J.R. Hughes, EEG in Clinical Practice, Butterworths, Heinemann, 1994.
- [7] T. Kasuba, Simplified Fuzzy ARTMAP, *AI Expert* 8 (11) (1993) 19–25.
- [8] A. Kriss, Recording Technique, in: A.M. Halliday (Ed.), *Evoked Potentials in Clinical Testing*, Churchill Livingstone, Livingstone, 1993.
- [9] K.E. Misulis, Spehlmann's Evoked Potential Primer: Visual, Auditory and Somatosensory Evoked Potentials in Clinical Diagnosis, Butterworths, Heinemann, 1994.
- [10] T. Muller, T. Ball, R. Kristeva-Feige, T. Mergner, J. Timmer, Selecting relevant electrode positions for classification tasks based on the electro-encephalogram, *Med. Biol. Eng. Comput.* 38 (2000) 62–67.
- [11] R. Palaniappan, P. Raveendran, Single Trial VEP Extraction Using Digital Filter, in: *Proceedings of IEEE 11th Workshop on Statistical Signal Processing*, Singapore, 6–8 August 2001, pp. 249–252.
- [12] J.G. Snodgrass, M. Vanderwart, A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity, *J. Exp. Psychol. Hum. Learning Memory* 6 (2) (1980) 174–215.
- [13] C. Tallon-Baudry, O. Bertrand, F. Peronnet, J. Pernier, Induced λ -band activity during the delay of a visual short-term memory task in humans, *J. Neurosci.* 18 (11) (1998) 4244–4254.
- [14] X.L. Zhang, H. Begleiter, B. Porjesz, A. Litke, Electrophysical evidence of memory impairment in alcoholic patients, *Biol. Psychiatr.* 42 (1997) 1157–1171.
- [15] X.L. Zhang, H. Begleiter, B. Porjesz, W. Wang, A. Litke, Event related potentials during object recognition tasks, *Brain Res. Bull.* 38 (6) (1995) 531–538.