Computer Methods and Programs in Biomedicine (2004) **75**, 11–22

ELSEVIER

Computer Methods and Programs in Biomedicine

www.intl.elsevierhealth.com/journals/cmpb

Design and implementation of an SVM-based computer classification system for discriminating depressive patients from healthy controls using the P600 component of ERP signals

I. Kalatzis[a], N. Piliouras[a], E. Ventouras[a], C.C. Papageorgiou[b], A.D. Rabavilas[b], D. Cavouras[a,*]

[a] *Department of Medical Instrumentation Technology, Technological Educational Institution of Athens, Ag. Spyridonos Street, Egaleo GR-122 10, Athens, Greece*
[b] *Psychophysiology Laboratory, Department of Psychiatry, Medical School, Eginition Hospital, University of Athens, Athens, Greece*

Received 20 June 2003; received in revised form 14 September 2003; accepted 14 September 2003

KEYWORDS
Depression;
Event-related potentials
ERPs;
P600 component;
Pattern recognition

Summary A computer-based classification system has been designed capable of distinguishing patients with depression from normal controls by event-related potential (ERP) signals using the P600 component. Clinical material comprised 25 patients with depression and an equal number of gender and aged-matched healthy controls. All subjects were evaluated by a computerized version of the digit span Wechsler test. EEG activity was recorded and digitized from 15 scalp electrodes (leads). Seventeen features related to the shape of the waveform were generated and were employed in the design of an optimum support vector machine (SVM) classifier at each lead. The outcomes of those SVM classifiers were selected by a majority-vote engine (MVE), which assigned each subject to either the normal or depressive classes. MVE classification accuracy was 94% when using all leads and 92% or 82% when using only the right or left scalp leads, respectively. These findings support the hypothesis that depression is associated with dysfunction of right hemisphere mechanisms mediating the processing of information that assigns a specific response to a specific stimulus, as those mechanisms are reflected by the P600 component of ERPs. Our method may aid the further understanding of the neurophysiology underlying depression, due to its potentiality to integrate theories of depression and psychophysiology.

1. Introduction

Depression, a condition burdening both the suffering individual and the community, is associated with disturbances of cognition. Based on psychophysiological and neuropsychological data, it has been suggested that deficits in depression are a result of a lack of cognitive initiative or motivation [1], and that depressed patients' fundamental deficits were in initiative and in strategic use of information [2].

To understand the nature and course of these abnormalities in patients with depression it is

*Corresponding author. Tel.: +30-210-5385-375; fax: +30-210-5910-975.
E-mail address: cavouras@teiath.gr (D. Cavouras).

0169-2607/$ – see front matter © 2003 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2003.09.003

essential to elucidate issues related with diagnostic demands and therapeutic evaluation. Event-related potentials (ERPs), measured on the scalp, are a method of choice for this purpose, because their high temporal resolution allows for real-time and non-invasive observation of electrical activity changes in neural circuits, during the processing of information related to the presentation of stimuli (or events) [3]. ERPs consist of a series of electrical potentials, labeled peaks or components, which are parts of the waveform containing significant local maxima and minima. Long latency ERPs start after 50—70 ms from the time the stimulus was presented. Research about long latency ERP components such as N100, P200, N200, P300, N400 and P600 is gaining wide interest in psychiatry, because these components provide information on the nature, timing, and extent of brain activation underlying covert cognitive processes [4,5].

Recently, considerable attention has been given to the P600 or the slow wave (SW) component [6,7]. In particular P600 is suggested to be sensitive to ''storage and retention of information in working memory'' [7], where working memory (WM) refers to the ability to keep information actively in mind and manipulate it in ways required by a given task [8]. This component has been also associated to the mnemonic binding processes by which the different aspects (i.e. internal and external) of information are linked into a coherent representation [9], in other words, it has been conceptualized as an index of information processing that ''assigns a specific response to a specific stimulus'' [10]. In an experiment measuring the SW (in the 375—840 ms epoch), during processing of a WM task in 14 depressed patients, it has been reported that depressed patients exhibited reduced late positive wave located at Cz site, as compared to healthy controls [11]. These effects have been accounted for ''by dysfunction of the central executive control of working memory'' [11]. In another recent study [12], using the SW elicited during emotionally valenced information in a WM task, reported that depressive patients, relative to controls, showed decreased processing of positive, but not negative information. Further, it has been found that depressive individuals demonstrated more negative SW amplitudes with increasing depressive mood severity. These results have been conceptualized as indexing that depressive cognition is characterized by a deficit in the processing of positive information.

Previous studies on computer-based ERP classification systems have focused mostly on the use of components whose latency ranges up to 500 ms and the employment of neural network classifiers for discriminating between normal controls and patients with neurologic [13,14] or a feature extraction component was used for extracting Karhunen—Loeve vectors, which were inputted to a classifier using the fuzzy *c*-mean (FCM) clustering algorithm. or orrrorpsychiatric disorders [15—17]. Features were extracted from the total duration of the ERP recording [14—16], or from only a specific component, such as the P300 for discriminating normal controls from patients suffering from multiple sclerosis [13], or the P600 for discriminating normal controls from patients suffering from schizophrenia [17].

Based on the above, it would be of interest to investigate whether the P600 component contains valuable information that could be used in discriminating depressive patients from normal controls, employing appropriate pattern recognition-based test procedures.

In the present study, we have developed a computer-based P600 component pattern recognition system for discriminating patients with depression from healthy subjects. P600 signals (500—800 ms) were recorded at 15 scalp leads on each subject. Signals were analyzed by means of features extracted from the P600 signal waveform. These new features, together with the recently developed powerful support vector machine (SVM) classification algorithm [18,19], were used in the design of the classification system. Another contribution of the present work was that the classification system consisted of an SVM classifier working at each lead and the outcome of each lead (either ''depressive'' or ''control'') was collected by a majority-vote engine (MVE), which decided on the class by a majority-vote rule. Furthermore, for comparison reasons, a conventional statistical method was used, e.g. stepwise discriminant analysis, in order to investigate the ability of the latency and the amplitude of the P600 component to differentiate between the two subject groups.

2. Materials and methods

2.1. Subjects

Twenty-five patients with depression and an equal number of gender and aged-matched healthy controls were examined. The controls were recruited from hospital staff and local volunteer groups and they were free of psychiatric and physical illness. All participants had no history of any neurological or hearing problems. All participants were right-handed as assessed by the Edinburgh Inventory [20]. Written informed consent was obtained from both patients and controls.

2.2. Stimuli and ERP recording procedure

The subjects were evaluated by a computerized version of the digit span Wechsler test [21], as reported in detail in a previous work by members of our research team [22]. In brief, the subjects sat in an anatomical chair placed inside an electromagnetically shielded room. A single sound of either high (3000 Hz) or low frequency (500 Hz) was presented to the subjects, who were asked to memorize the numbers that followed. The warning stimulus lasted 100 ms. A 1 s interval followed and then the numbers to be memorized were presented. At the end of the number sequence presentation, the signal tone was repeated and subjects were asked to recall the administered numbers as quickly as possible. The numbers were recalled by the subject in the same (low frequency tone) or in the opposite order (high frequency tone) than that presented to him/her. ERPs were recorded using Ag/AgCl electrodes, during the 1 s interval between the warning stimulus and the first administered number. EEG activity was recorded from 15 scalp electrodes based on the international 10–20 system of electroencephalography [23], referred to both earlobes (leads at Fp1, Fp2, F3, F4, C3, C4, (C3 − T5)/2, (C4 − T6)/2, P3, P4, O1, O2, Pz, Cz, and Fz) (see Fig. 4). It should be noted that the positions (C3 − T5)/2 and (C4 − T6)/2 are used as electrode leads, because these positions correspond to brain areas serving verbal memory and language [24]. Recordings were digitized at a sampling rate of 500 Hz.

2.3. Feature generation

A dedicated computer software (Fig. 1) was developed in C++ and it was used to read the ERP signals and to calculate features related to the ERP signal of the P600 component (500–800 ms time interval). The following features were automatically calculated as follows (see Fig. 1), where $500\,\text{ms} \le t \le 800\,\text{ms}$:

(1) Latency (LAT, $t_{s_{\max}}$)—the ERP's latency time, i.e. the time where the maximum signal value appears:

$$t_{s_{\max}} = \{t | s(t) = s_{\max}\} \tag{1}$$

(2) Amplitude (AMP, $s_{\max}$)—the maximum signal value:

$$s_{\max} = \max\{s(t)\} \tag{2}$$

(3) Latency/amplitude ratio (LAR, $t_{s_{\max}}/s_{\max}$).
(4) Absolute amplitude (AAMP, $|s_{\max}|$).
(5) Absolute latency/amplitude ratio (ALAR, $|t_{s_{\max}}/s_{\max}|$).
(6) Positive area (PAR, A_p)—the sum of the positive signal values:

$$A_p = \sum_{t=500\,\text{ms}}^{800\,\text{ms}} 0.5(s(t) + |s(t)|) \tag{3}$$

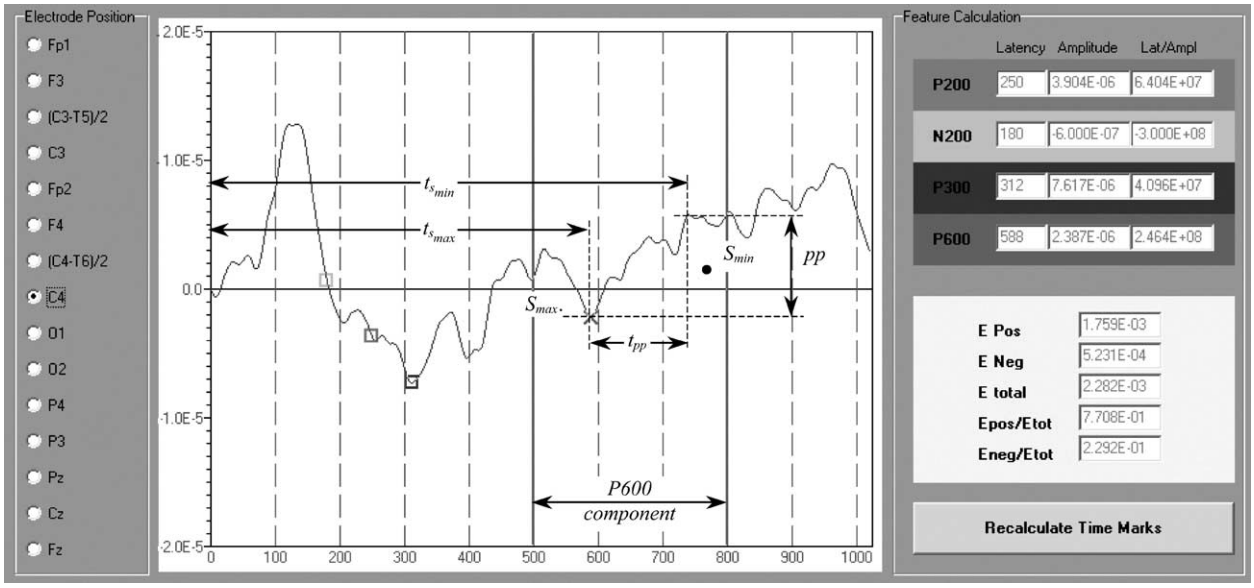

Fig. 1 Interface of the custom-made software system designed to read ERP signals and to extract specific features. The graph represents the variation of the ERP electrical potential (in volts) with time after stimulus (in milliseconds) (see Eqs. (1)–(14) for explanation of specific quantities displayed).

(7) Negative area (NAR, A_n)—the sum of the negative signal values:

$$A_n = \sum_{t=500\,ms}^{800\,ms} 0.5(s(t) - |s(t)|) \quad (4)$$

(8) Absolute negative area (ANAR, $|A_n|$).
(9) Total area (TAR, A_{pn}):

$$A_{pn} = A_p + A_n \quad (5)$$

(10) Absolute total area (ATAR, $|A_{pn}|$).
(11) Total absolute area (TAAR, $A_{p|n|}$):

$$A_{p|n|} = A_p + |A_n| \quad (6)$$

(12) Average absolute signal slope (AASS, $|\bar{\dot{s}}|$):

$$|\bar{\dot{s}}| = \frac{1}{n} \sum_{t=500\,ms}^{800\,ms-\tau} \frac{1}{\tau} |s(t+\tau) - s(t)| \quad (7)$$

where τ is the sampling interval of the signal ($\tau = 2$ ms, for the sampling rate of 500 Hz), n the number of samples of the digital signal (actual $n = (800 - 500\,ms)/2\,ms = 150$), and $s(t)$ the signal value of the t-th sample.
(13) Peak-to-peak (PP, pp):

$$pp = s_{max} - s_{min} \quad (8)$$

where s_{max} and s_{min} are the maximum and the minimum signal values, respectively:

$$s_{max} = \max\{s(t)\}, \qquad s_{min} = \min\{s(t)\} \quad (9)$$

(14) Peak-to-peak time window (PPT, t_{pp}):

$$t_{pp} = t_{s_{max}} - t_{s_{min}} \quad (10)$$

(15) Peak-to-peak slope (PPS, $\dot{s}_{pp}$):

$$\dot{s}_{pp} = \frac{pp}{t_{pp}} \quad (11)$$

(16) Zero crossings (ZC, n_{zc})—the number of times t that $s(t) = 0$, in peak-to-peak time window:

$$n_{zc} = \sum_{t=t_{s_{min}}}^{t_{s_{max}}} \delta_s \quad (12)$$

where $\delta_s = 1$ if $s(t) = 0$, 0 otherwise.
(17) Zero crossings density (ZCD, d_{zc})—zero crossings per time unit, in peak-to-peak time window:

$$d_{zc} = \frac{n_{zc}}{t_{pp}} \quad (13)$$

where n_{zc} are the zero crossings and t_{pp} is the peak-to-peak time window.

(18) Slope sign alterations (SSA, n_{sa})—the number of slope sign alterations of two adjacent points of the ERP signal:

$$n_{sa} = \sum_{t=500\,ms+\tau}^{800\,ms-\tau} 0.5 \times \left| \frac{s(t-\tau) - s(t)}{|s(t-\tau) - s(t)|} + \frac{s(t+\tau) - s(t)}{|s(t+\tau) - s(t)|} \right| \quad (14)$$

where τ is the sampling interval of the signal ($\tau = 2$ ms, for the sampling rate of 500 Hz).

All features were normalized to zero mean and unit standard deviation [25], according to relation:

$$x_i' = \frac{x_i - \mu}{\sigma} \quad (15)$$

where x_i and x_i' are the i-th feature values before and after the normalization respectively, and μ and σ are the mean value and standard deviation, respectively, of feature x over all subjects (depressives and normal controls).

2.4. Conventional statistical analysis

To investigate whether the two groups of subjects could be discriminated by conventional statistical analysis methods, a stepwise discriminant method was employed, utilizing the amplitudes (parameter AMP) and latencies (parameter LAT) of the P600 component at all 15 leads. It should be noted that the equality of the covariance matrices of the variables entered for the two groups was ascertained with Box's *M*-test.

2.5. Best feature selection

An exhaustive search, using the leave-one-out method [25] and the fast but high-performance cubic least-squares minimum-distance (C-LSMD) classifier (see relation 16 and Appendix A), which is an extended version of the LSMD [26], of all possible 2, 3, 4, 5, and 6 feature combinations was performed at each one of the 15 leads. The purpose was to determine at each lead the best feature combination having the highest classification accuracy, using a robust and fast classification technique prior to employing more sophisticated but time demanding classifiers.

Application of the leave-one-out method requires that one sample is left out, and the classifier is trained (designed) by the rest of the samples. In this way, the left-out sample is considered by the system as unknown. Repeating this procedure for all samples, we can get a realistic evaluation

of the system's capabilities. However, by applying the leave-one-out method, feature normalization (Eq. (15)) had to be recalculated (new values for μ and σ) each time a subject was left out.

Data processing was split appropriately and it was performed on two 2.4 MHz Pentium IV workstations. It took several hours of processing time to complete the exhaustive search. Combinations with higher numbers of features were also tested, employing the forward stepwise feature selection technique [25] and the C-LSMD classifier.

The discriminant function of the C-LSMD classifier for class c, employed in the feature selection procedure, is given by:

$$\begin{aligned} g_c(\boldsymbol{X}) = &\sum_{i=1}^{d} a_{ciii} x_i^3 + \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} a_{cijj} x_i^2 x_j \\ &+ \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} a_{cijj} x_i x_j^2 + \sum_{i=1}^{d-2}\sum_{j=i+1}^{d-1}\sum_{k=j+1}^{d} a_{cijk} x_i x_j x_k \\ &+ \sum_{i=1}^{d} a_{cii} x_i^2 + \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} a_{cij} x_i x_j \\ &+ \sum_{i=1}^{d} a_{ci} x_i - b_c \end{aligned} \tag{16}$$

where d is the number of features, α_c the weight coefficients, b_c the threshold parameter, and x_j the pattern vector elements.

The best-feature combinations thus determined were used to design at each lead the support vector machines classifier (see relation (17) and Appendix B), using the leave-one-out method for discriminating depressive patients from normal controls.

2.6. Classification

2.6.1. SVM classification

The discriminant equation of the SVM classifier [18,19] is a function of kernel $k(\boldsymbol{x}_i, \boldsymbol{x})$ and is given by:

$$g(\boldsymbol{x}) = \operatorname{sign}\left(\sum_{i=1}^{N_S} \alpha_i y_i k(\boldsymbol{x}_i, \boldsymbol{x}) + b\right) \tag{17}$$

where $\boldsymbol{x}_i$ are support vectors, N_S the number of support vectors, a_i the weight parameters, b the threshold parameter, and $y_i \in \{-1, +1\}$ depending on the class.

In the present study the cubic non-homogeneous polynomial function was used as kernel, given by $k(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} \cdot \boldsymbol{y} + 1)^d$ with $d = 3$, resulting in the following discriminant function for the SVM classifier:

$$g(\boldsymbol{x}) = \operatorname{sign}\left(\sum_{i=1}^{N_S} \alpha_i y_i (\boldsymbol{x}_i \cdot \boldsymbol{x} + 1)^3 + b\right) \tag{18}$$

Regarding the kernel function, other kernels were also tested, such as the Gaussian radial basis function (RBF) kernel and the linear kernel (see Appendix B).

2.6.2. Majority-vote classification

For classifying a subject as belonging to either the ''depressives'' or ''controls'' category, a classification system was developed as shown in Fig. 4. At each lead there is an SVM classifier working, designed to use the lead's particular P600 features and to assign the P600 component to one of the two classes. The outcome of each lead (either ''depressive'' or ''control'') is collected by a majority-vote engine, which decides on the class by a majority-vote rule. The overall system was evaluated by the leave-one subject-out method.

3. Results and discussion

Fig. (2) shows the grand averages of the ERP signals of the two groups of subjects. Dashed lines represent the depressive patients and solid lines the controls. Discriminant analysis, regarding P600 amplitude (parameter AMP), revealed that only two leads entered the discriminant function (C4 and $C3 - T5/2$), being able to classify correctly only 67.3%, of the originally grouped cases. It is noteworthy that this classification rate dropped slightly to 63.3% in cross-validation. The two leads belong to different components and enter the discriminant function with opposite signs. This was to be expected since in the case of C4 the absolute mean value of the depressive group was greater, while in the case of $(C3 - T5)/2$ it was less than that obtained from the normal group. Results of comparisons of the latencies (parameter LAT) did not reveal any discrimination. In order to illustrate the lack of adequate differentiation between the two subjects' groups, when the amplitude and the latency of the P600 are taken into account the scatter diagram of the P600 amplitude and latency at FP1 lead is presented (Fig. 3). There is a significant overlap between depressives and controls leading to no conclusive evidence as to probable differences between the two groups of subjects.

Table 1 displays the optimum feature combinations (highest classification accuracy with smallest number of features involved) achieved at each lead in discriminating the two groups of subjects, employing the leave-one-out method and the C-LSMD

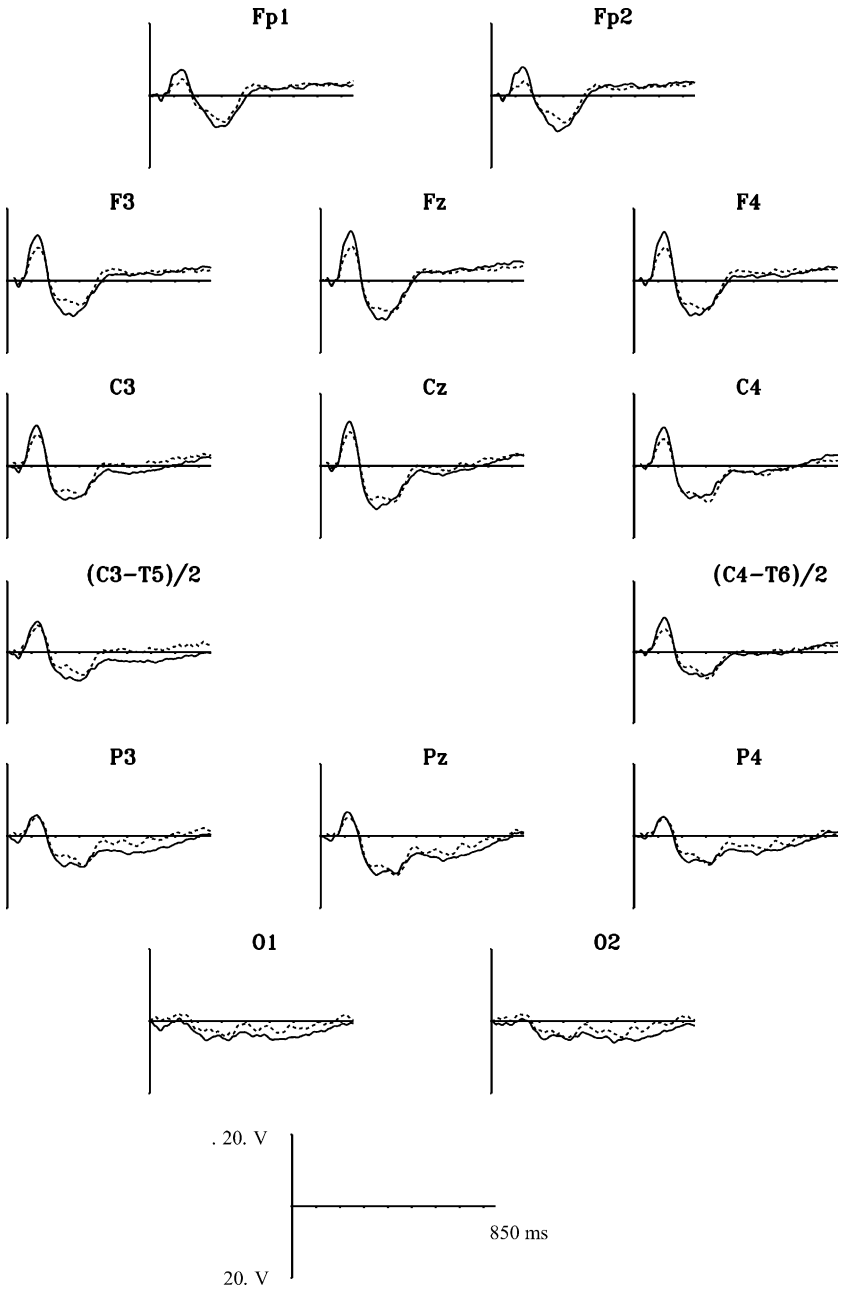

Fig. 2 Grand averages of ERP signals of depressives (dashed lines) and normal controls recorded at each lead. The electrode notation is based on the international 10—20 system of electroencephalography [23].

classifier. Best combinations comprised between two and five features. Additionally, sensitivities and specificities are shown, corresponding to the classification accuracy of the ''controls'' and ''depressives'' classes respectively. Maximum classification accuracies varied between 68% and 78% in the 15 leads, signifying the difficulty in distinguishing between the two groups by means of the P600 component. This was true even when new features were generated and they were used in the attempt to extract from the P600 signals information of high discriminatory power. It may thus be said that the information contained in the P600 signals of a single lead is not capable of distinguishing effectively between depressives and normal controls, even when powerful classification algorithms, such as the SVM, are employed. SVM classification accuracy was also tested with other SVM-kernel functions, such as the linear and the Gaussian radial basis functions. The SVM with the linear kernel function provided low discrimination accuracy while with the radial basis function gave results (obtained with $C = 500$ and standard deviation $\sigma = 0.45$, see Appendix B) of slightly lower accuracy

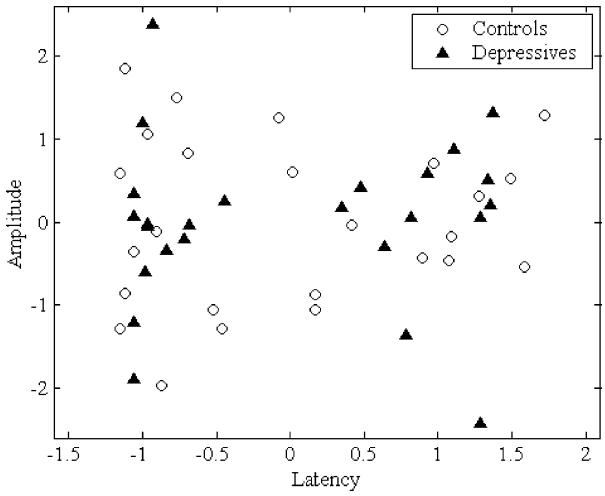

Fig. 3 Feature 'latency' against feature 'amplitude' (see Eqs. (1) and (2)) plot of P600 signals at the Fp1 lead of depressives (triangles) and normal controls (circles). Feature values are normalized to zero mean and unit standard deviation (see [25] and Eq. (15)).

to those of the third degree polynomial kernel (with $C = 500$).

However, a more careful observation of Table 1 may reveal that leads Fp2, F4, C4, (C4 − T6)/2, and O2 at the right side of the head (see Fig. 4) display a slightly higher discrimination accuracy than the corresponding leads Fp1, F3, C3, (C3 − T5)/2, and O1 at the left side. In fact, when the outcome of those leads at right were combined by the MVE, employing the leave-one subject-out method, they revealed 92% accuracy in distinguishing normal controls from depressives (Table 2). A similar trial but at the scalp's left side corresponding leads revealed lower (82%) classification accuracy (Table 3). In Tables 2 and 3, negative predictive value is the fraction of those subjects predicted by the algorithm to be normal controls that are truly normal controls, while positive predictive value represents the fraction of subjects determined by the algorithm to be depressives and are actually depressives. Right hemispheric differences between depressive patients and controls have also been noted in previous studies. In a study employing late P3 (peak latency 460 ms) ERP signals and principal component analysis, differences were found at the right parietal region between controls and depressive patients [27]. Similarly, other studies on depressed patients subjected to neuropsychological tests [28] or behavioral laterality tasks using either visuo-spatial [29] or auditory stimuli [30] have found evidence of a right posterior deficit in depression. Further support for this hypothesis comes from studies showing the right cerebral hemisphere is preferentially sensitive to the affective context of language [31] in depressives and/or depressives have difficulty processing receptive affective prosodic speech [32]. Additionally, this abnormal lateralization has been linked to the treatment outcome of the depression [33].

When the outcomes of all leads were incorporated into the MVE then high discrimination results were achieved. Table 4 gives a detailed account of the MVE result for each subject. The number of correctly classified leads and the particular misclassified leads, as well as the overall sensitivity (the fraction of depressives correctly identified by the MVE), specificity (the fraction of normal controls correctly identified by the MVE), and the overall classification accuracy are presented. According to the MVE-rule, a subject (control or depressive)

Table 1 Best feature combination at each lead for SVM classification, and the corresponding SVM accuracies

Lead	Feature combination	Sensitivity (%)	Specificity (%)	Overall (%)
Fp1	LAT, AAMP, LAR, ALAR, TAR	60	84	72
Fp2	LAT, ZC, ALAR, ZCD	76	72	74
F3	LAT, PPS, AAMP, PPT	76	68	72
F4	PAR, SSA, AAMP, ZCD	72	84	78
C3	AMP, TAAR, ZC, ATAR, PPT	72	80	76
C4	PP, LAR, TAR	80	72	76
(C3 − T5)/2	SSA, ZCD	68	68	68
(C4 − T6)/2	AMP, ZC, TAR	80	72	76
P3	AMP, ANAR, ZC, PP, SSA	68	76	72
P4	LAT, TAAR, AAMP, LAR, PPT	64	72	68
O1	LAT, ATAR, ZCD	68	84	76
O2	ZC, PP, SSA	80	76	78
Pz	LAT, ANAR, TAAR, PP, LAR	60	84	72
Cz	AMP, ZC	64	76	70
Fz	LAT, ZC, PPS, ATAR	68	72	70

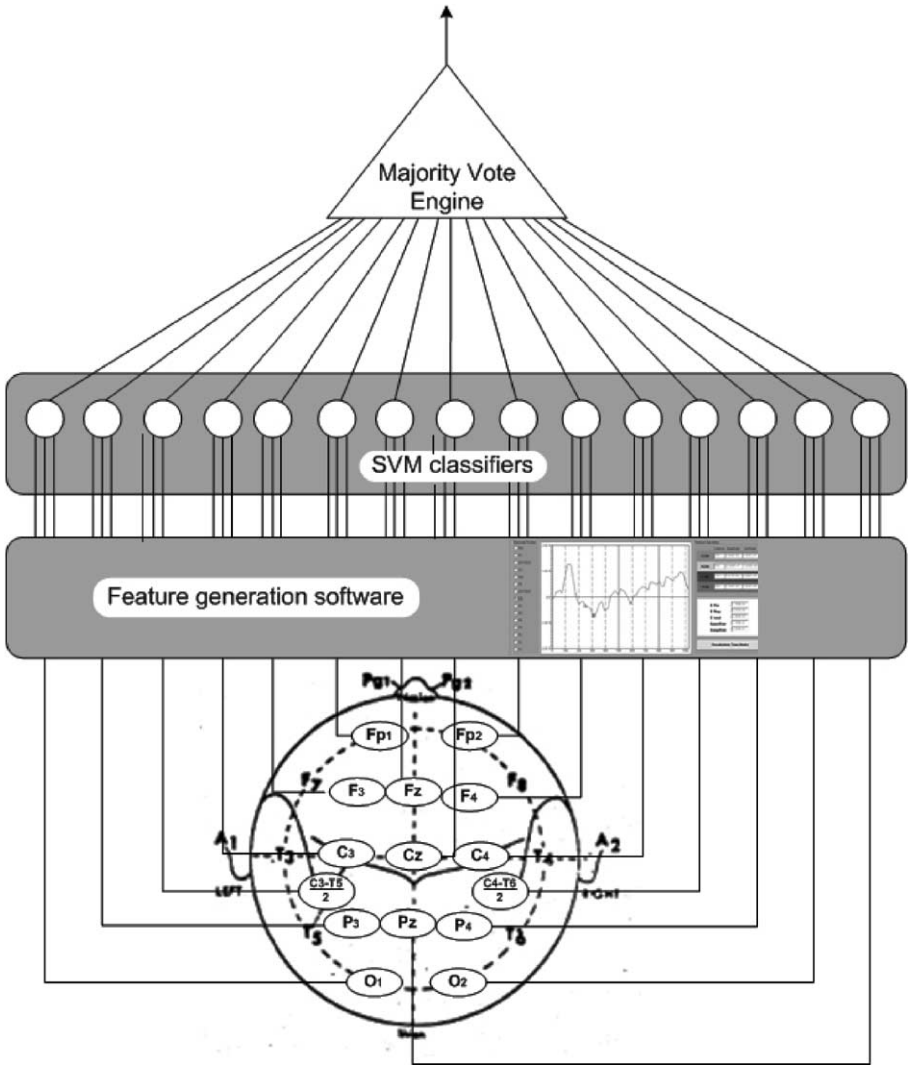

Fig. 4 Schematic diagram of scalp leads distribution and MVE classification system steps: first, the signals from all leads are inserted in the automatic feature generation software system. Then, an SVM classifier is employed at each lead to classify each subject to one of two classes (depressives and normal controls). Finally, on the basis of those lead sub-classifications, each subject is assigned to a particular class using a majority-vote rule.

Table 2 MVE classification accuracy employing right hemisphere leads Fp2, F4, C4, (C4 – T6)/2, and O2

	Control	Depressive	Accuracy (%)	
Control	22	3	88	Specificity
Depressive	1	24	96	Sensitivity
Accuracy (%)	95.7	88.9	92	Overall
	Negative predictive value	Positive predictive value		

Table 3 MVE classification accuracy employing left hemisphere leads Fp1, F3, C3, (C3 – T5)/2, and O1

	Control	Depressive	Accuracy (%)	
Controls	22	2	88	Specificity
Depressives	6	19	76	Sensitivity
Accuracy (%)	78.6	90.5	82	Overall
	Negative predictive value	Positive predictive value		

Table 4 Detailed MVE algorithm classification accuracy results

Normal controls[a]				Depressive patients[b]			
Normal control number	Number of correctly classified leads	Misclassified leads	MVE result	Patient number	Number of correctly classified leads	Misclassified leads	MVE result
1	12	P3, Pz, Fz	Normal	1	9	F3, F4, P3, P4, Pz, Fz	Depressive
2	11	C3, (C3 − T5)/2, O2, Fz	Normal	2	9	C3, (C3 − T5)/2, P4, O2, Cz, Fz	Depressive
3	12	Fp1, Fp2, Cz	Normal	3	9	Fp1, (C3 − T5)/2, P3, P4, O1, Cz	Depressive
4	14	P4	Normal	4	10	F3, C3, P3, P4, Pz	Depressive
5	11	Fp2, C4, O1, Pz	Normal	5	12	Fp2, C3, O1	Depressive
6	13	(C4 − T6)/2, P3	Normal	6	12	C3, (C3 − T5)/2, (C4 − T6)/2	Depressive
7	10	C3, (C3 − T5)/2, (C4 − T6)/2, Pz, Fz	Normal	7	11	Fp1, Fp2, C3, Pz	Depressive
8	10	Fp2, F4, C4, P4, Cz	Normal	8	10	Fp1, F4, C3, (C3 − T5)/2, O1, Pz	Depressive
9	11	F3, (C3 − T5)/2, P3, Cz	Normal	9	9	Fp1, C3, O2, Pz, Cz, Fz	Depressive
10	9	Fp1, F4, C3, O1, O2, Cz	Normal	10	7	Fp1, F3, F4, C3, C4, O1, Pz, Fz	Normal
11	6	Fp2, F3, C4, (C3 − T5)/2, (C4 − T6)/2, P4, O1, O2, Fz	Depressive	11	12	C3, Fp1, Fp2	Depressive
12	13	P3, Pz	Normal	12	11	Fp2, P3, Cz, Pz	Depressive
13	13	(C4 − T6)/2, O2	Normal	13	12	C3, O2, Fz	Depressive
14	11	Fp2, F3, P3, Fz	Normal	14	8	Fp1, Fp2, F4, C4, O1, O2, Cz	Depressive
15	13	Fp1, C4	Normal	15	9	Fp1, F4, (C3 − T5)/2, (C4 − T6)/2, P3, P4	Depressive
16	11	F4, C3, C4, (C4 − T6)/2	Normal	16	12	(C3 − T5)/2, O2, Cz	Depressive
17	12	Fp2, F3, C4	Normal	17	14	C3	Depressive
18	10	F3, C3, (C3 − T5)/2, Cz, Fz	Normal	18	13	(C4 − T6)/2, Pz	Depressive
19	10	C3, (C3 − T5)/2, (C4 − T6)/2, P4, Cz	Normal	19	12	Fp2, F3, P4	Depressive
20	15	—	Normal	20	6	Fp1, F3, F4, C4, (C3 − T5)/2, P4, O1, Pz, Fz	Normal
21	12	Fp1, O2, Cz	Normal	21	9	F4, C3, C4, P3, O1, Fz	Depressive
22	13	F3, P3	Normal	22	10	C3, C4, (C3 − T5)/2, (C4 − T6)/2, Pz	Depressive
23	8	F3, C3, (C3 − T5)/2, (C4 − T4)/2, P4, O1, O2	Normal	23	9	Fp1, F3, P3, P4, O1, Cz	Depressive
24	11	F3, F4, P4, Cz	Normal	24	12	(C4 − T6)/2, P3, Fz	Depressive
25	10	Fp2, C4, (C3 − T5)/2, P4, Fz	Normal	25	13	P4, Cz	Depressive

Overall: 94%.
[a] Specificity: 96%.
[b] Sensitivity: 92%.

is classified to the class with the majority (>7) of class-assigned leads. An overall classification accuracy was 94%, misclassifying one normal-control as depressive (96% specificity) and mistaking two depressives for normal-controls (92% sensitivity). It is noteworthy that comparison between the proposed SVM-based classification system and the conventional one shows that the proposed method is prevailing. Thus, these results may be regarded as quite promising, especially when compared to individually achieved classification accuracies at each lead. Another important finding related to the reliability of the MVE is that when its self-consistency was tested (i.e. using no leave-one subject-out), patients and normal-controls were all classified correctly. The meaning is that the MVE-system could classify accurately all subjects that were employed in its design. However, its robustness would be expected to increase if, in future, a larger number of subjects was to be employed in its design.

Appendix A. The cubic least-squares minimum-distance classifier

The least-squares minimum-distance (LSMD) is a linear classifier [26]. Extending the input feature vectors of the LSMD with polynomial terms up to third degree (as a trade-off between the second degree limited classification capabilities and the computational requirements of higher degree polynomial terms [25]), it is possible to discriminate data belonging to non-linearly separable classes. Another important feature of the LSMD is the least-squares mapping process, which transforms the training patterns from the input space to the decision space (dimensionally equal to the number of classes), where the members of each class are clustered around arbitrary pre-selected points, such that the mapping error is minimized. Prior to least-squares transformation, the input space is further augmented by ''−1'' for convenience.

Let $\boldsymbol{X}^{\mathrm{T}} = [x_1 x_2 \dots x_d]$ a pattern vector, where d is the number of features, $\boldsymbol{Z}$ the extended with quadratic and cubic terms pattern vector, and $\hat{Z}^{\mathrm{T}} = [\mathbf{Z}^{\mathrm{T}} - 1]$ the final augmented pattern. The transformation matrix $\boldsymbol{A}$ is calculated to minimize the total mean-square error, which yields $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \left[\sum_{c=1}^{K}\sum_{i=1}^{N_c+1}\frac{P_c}{N_c}V_c\hat{Z}_{ci}^{\mathrm{T}}\right]\left[\sum_{c=1}^{K}\sum_{i=1}^{N_c+1}\frac{P_c}{N_c}Z_{ci}\hat{Z}_{ci}^{\mathrm{T}}\right]^{-1} \quad (A.1)$$

where $\hat{Z}_{ci}^{\mathrm{T}}$ is the i-th augmented pattern vector of class c, $\boldsymbol{V}_c$ the decision space vector around which the patterns of class c are to be clustered (usually the $\boldsymbol{V}_c$ is the *unit vector* of class c, $\boldsymbol{V}_c^{\mathrm{T}} = [0 \dots 0\,1\,0 \dots 0]$, where the only non-zero element ''1'' is in the c-th row of $\boldsymbol{V}_c$), K the number of classes, N_c the number of patterns in class c, and P_c the a priori probability of class c (for equal probabilities, $P_c = 1/K$ for each c).

Having the transformation matrix $\boldsymbol{A}$, a new pattern $\boldsymbol{X}$ is classified to class c, for which the following discriminant function $g_c(\boldsymbol{X})$ is maximum:

$$\begin{aligned} g_c(\boldsymbol{X}) = {} & \sum_{i=1}^{d} a_{ciii}x_i^3 + \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} a_{ciij}x_i^2x_j \\ & + \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} a_{cijj}x_ix_j^2 + \sum_{i=1}^{d-2}\sum_{j=i+1}^{d-1}\sum_{k=j+1}^{d} a_{cijk}x_ix_jx_k \\ & + \sum_{i=1}^{d} a_{cii}x_i^2 + \sum_{i=1}^{d-1}\sum_{j=i+1}^{d} a_{cij}x_ix_j + \sum_{i=1}^{d} a_{ci}x_i - b_c \end{aligned} \quad (A.2)$$

where α_c are weight coefficients, b_c the threshold parameters, and $c = 1, 2, \dots, K$. The weight coefficients consist of: d coefficients of x_i terms (a_i), d coefficients of x_i^2 terms (a_{ii}), d coefficients of x_i^3 terms (a_{iii}), $d!/[2!(d-2)!]$ coefficients of x_ix_j terms, $i < j$ (a_{ij}), $d!/[2!(d-2)!]$ coefficients of x_ix_j terms, $i < j$ (a_{ij}), $d!/[2!(d-2)!]$ coefficients of $x_i^2x_j$ terms, $i < j$ (a_{iij}), $d!/[2!(d-2)!]$ coefficients of $x_i^2x_j$ terms, $i < j$ (a_{ijj}), $d!/[3!(d-3)!]$ coefficients of $x_ix_jx_k$ terms, $i < j < k$ (a_{ijk}), which, including parameter b, result in a total of $(d+3)!/(d!3!)$ coefficients or parameters.

Appendix B. The support vector machines classifier

A classifier based on support vector machines [18,19] is a general classifier that it can be applied to linearly as well to non-linearly separable data, with or without overlap between the classes.

In the most general case of overlapped and non-linearly separable data, the problem is (a) to transform the training patterns from the input space to a feature space with higher dimensionality ($\boldsymbol{x} \in \boldsymbol{R}^d \mapsto \Theta(\boldsymbol{x}) \in \boldsymbol{R}^h$), where the classes become linearly separable, and (b) to find two parallel hyperplanes with maximum distance between them and at the same time with minimum number of training points in the area between them (also called the *margin*).

The discriminant function is then given by:

$$g(\boldsymbol{x}) = \mathrm{sign}(\boldsymbol{w}\cdot\Theta(\boldsymbol{x}) + b) \quad (B.1)$$

where sign '+' corresponds to class 1, sign '−' to class 2, $\boldsymbol{x}$ is the pattern vector, $\boldsymbol{w}$ the normal vector to the hyperplanes, and b the bias or threshold which describes the distance between the decision hyperplane (which is a hyperplane in the middle of the margin) and the origin (which is equal to $b/||\boldsymbol{w}||$, [18]).

Solving the optimization problem of maximizing the margin and minimizing the number of training patterns within the margin, and by use of a *kernel function* (that satisfies Mercer's conditions [18]) in the place of the inner product $\Theta(\boldsymbol{x}_i)\cdot\Theta(\boldsymbol{x}_j)$, the discriminant function of the SVM classifier can be written as:

$$g(\boldsymbol{x}) = \text{sign}\left(\sum_{i=1}^{N_S}\alpha_i y_i k(\boldsymbol{x}_i, \boldsymbol{x}) + b\right) \tag{B.2}$$

where $\boldsymbol{x}_i \in \boldsymbol{R}^d$, $i = 1{-}N$, d the number of features, N the number of training pattern vectors belonging to two classes $y_i \in \{-1, +1\}$, α_i are coefficients that are obtained by solving the above mentioned optimization problem, and N_S is the number of pattern vectors (also called the *support vectors*) with non-zero α_i's.

Threshold b may be expressed [18] as:

$$b = \frac{1}{N_S}\sum_{j=1}^{N_S}\left(y_j - \sum_{i=1}^{N}\alpha_i y_i k(\boldsymbol{x}_i, \boldsymbol{x}_j)\right) \tag{B.3}$$

Functions that are commonly used as kernels are:

(i) The linear kernel:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i \cdot \boldsymbol{x}_j \tag{B.4a}$$

(ii) The polynomial kernel:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i \cdot \boldsymbol{x}_j + \theta)^d \tag{B.4b}$$

where d is the degree of the polynomial and θ an offset parameter.

(iii) The Gaussian radial basis kernel:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(\frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathsf{T}}(\boldsymbol{x}_i - \boldsymbol{x}_j}{2\sigma^2}\right) \tag{B.4c}$$

where σ is the standard deviation.

(iv) The sigmoidal kernel:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \tanh(\kappa(\boldsymbol{x}_i \cdot \boldsymbol{x}_j) + \theta) \tag{B.4d}$$

where κ is the gain and θ the offset.

(v) The inverse multiquadric kernel:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = ((\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathsf{T}}(\boldsymbol{x}_i - \boldsymbol{x}_j) + c^2)^{-1/2} \tag{B.4e}$$

where c is a non-negative real number.

References

[1] P.T. Hertel, S.S. Rude, Depressive deficits in memory: focusing attention improves subsequent recall, J. Exp. Psychol. Gen. 120 (1991) 301–309.

[2] W. Heller, J.B. Nitschke, Regional brain activity in emotion: a framework for understanding cognition in depression, Cognition Emotion 11 (1998) 637–661.

[3] R. Johnson Jr., Event-Related Brain Potentials and Cognition, Handbook of Neuropsychology, Elsevier, Amsterdam, 1995.

[4] N. Boutros, H. Nasrallah, R. Leighty, M. Torello, P. Tueting, S. Olson, Auditory evoked potentials, clinical vs. research applications, Psychiatr. Res. 69 (1997) 183–195.

[5] M. Fabiani, G. Gratton, M. Coles, Event-related potentials: methods, theory, and applications, in: J. Cacioppo, L. Tassinary, G. Bernston (Eds.), Handbook of Psychophysiology, Cambridge University Press, New York, NY, 2000, pp. 53–84.

[6] S. Frisch, S. Kotz, D. von Cramon, A. Friederici, Why the P600 is not just a P300: the role of the basal ganglia, Clin. Neurophysiol. 114 (2003) 336–340.

[7] D.S. Ruchkin, R. Johnson Jr., H. Canoune, W. Ritter, Event-related potentials during arithmetic and mental rotation, Electroencephalogr. Clin. Neurophysiol. 85 (1992) 396–398.

[8] A. Baddeley, R. Logie, Working memory: the multiple-component model, in: A. Miyake, P. Shah (Eds.), Models of Working Memory, Cambridge University Press, New York, NY, 1999, pp. 28–61.

[9] F. Guillem, M. Bicu, T. Pampoulova, R. Hooper, D. Bloom, M.-A. Wolf, J. Messier, R. Desautels, C. Todorov, P. Lalonde, J.B. Debruillea, The cognitive and anatomo-functional basis of reality distortion in schizophrenia: a view from memory event-related potentials, Psychiatr. Res. 117 (2003) 137–158.

[10] M. Falkenstein, J. Hohnsbein, J. Hoormann, Effects of choice complexity on different subcomponents of the late positive complex of the event-related potential, Electroencephalogr. Clin. Neurophysiol. 92 (1994) 148–160.

[11] L. Pelosi, T. Slade, L.D. Blumhardt, V.K. Sharma, Working memory dysfunction in major depression: an event-related potential study, Clin. Neurophysiol. 111 (2000) 1531–1543.

[12] P.J. Deldin, C.M. Deveney, A.S. Kim, B.R. Casas, J.L. Best, A slow wave investigation of working memory biases in mood disorders, J. Abnormal Psychol. 110 (2) (2001) 267–281.

[13] J.D. Slater, F.Y. Wu, L.S. Honig, R.E. Ramsay, R. Morgan, Neural network analysis of the P300 event-related potential in multiple sclerosis, Electroencephalogr. Clin. Neurophysiol. 90 (1990) 114–122.

[14] T.J. Dasey, E. Micheli-Tzanakou, Detection of multiple sclerosis with visual evoked potentials—an unsupervised computational intelligence system, IEEE T. Inf. Technol. Biomed. 4 (2000) 216–224.

[15] J.R. Sveinsson, J.A. Benediktsson, S.B. Stefansson, K. Davidsson, Parallel principal component neural networks for classification of event-related potentials, Med. Eng. Phys. 19 (1997) 15–20.

[16] R. Palaniappan, R. Paramesran, Using genetic algorithm to identify the discriminatory subset of multi-channel spectral bands for visual response, Appl. Soft Comput. 2 (2002) 48–60.

[17] C. Vasios, C. Papageorgiou, G.K. Matsopoulos, K.S. Nikita, N. Uzunoglu, A decision support system of

evoked potentials for the classification of patients with first-episode schizophrenia, German J. Psychiatr. 5 (2002) 78—84.
[18] V. Kecman, Learning and Soft Computing, Support Vector Machines, Neural Networks, and Fuzzy Logic Models, MIT Press, Cambridge, MA, 2001, pp. 121—191.
[19] K.-R. Muller, S. Mika, G. Ratsch, K. Tsuda, B. Schölkopf, An introduction to kernel-based learning algorithms, IEEE T. Neural Network 12 (2001) 181—202.
[20] R.C. Oldfield, The assessment and analysis of handedness: the Edinburgh Inventory, Neuropsychologia 9 (1971) 97—113.
[21] D. Wechsler, Manual for the Wechsler Adult Intelligence Scale, Psychological Corporation, New York, 1955.
[22] C. Papageorgiou, E. Ventouras, N. Uzunoglu, A. Rabavilas, C. Stefanis, Changes of P300 elicited during a working memory test in individuals with depersonalization-derealization experiences, Neuropsychobiology 46 (2002) 70—75.
[23] H. Jasper, The ten—twenty electrode system of the International Federation, Electroencephalogr. Clin. Neurophysiol. 10 (1958) 371—375.
[24] J. Binder, Functional magnetic resonance imaging: language mapping, Neurosurg. Clin. North Am. 8 (1997) 383—392.
[25] S. Theodoridis, K. Koutroumbas, System evaluation, in: Pattern Recognition, Academic Press, UK, 1998 (Chapter 10).
[26] N. Ahmed, K.R. Rao, Feature selection in pattern recognition, in: Orthogonal Transforms for Digital Signal Processing, Springer-Verlag, NY, 1975 (Chapter 10).
[27] J. Kayser, G.E. Bruder, C.E. Tenke, J.W. Stewart, F.M. Quitkin, Event-related potentials ERPs to hemifield presentations of emotional stimuli: differences between depressed patients and healthy adults in P3 amplitude and asymmetry, Int. J. Psychophysiol. 36 (2000) 211—236.
[28] E.N. Miller, T.A. Fujioka, L.J. Chapman, J.P. Chapman, Hemispheric asymmetries of function in patients with major affective disorders, J. Psychiatr. Res. 29 (1995) 173—183.
[29] W. Heller, M.A. Etienne, G.A. Miller, Patterns of perceptual asymmetry in depression and anxiety: implications for neuropsychological models of emotion and psychopathology, J. Abnormal Psychol. 104 (1995) 327—333.
[30] Y. Yovell, H.A. Sackeim, D.G. Epstein, J. Prudic, D.P. Devanand, M.C. McElhiney, J.M. Settembrino, G.E. Bruder, Hearing loss and asymmetry in major depression, J. Neuropsychiatr. Clin. Neurosci. 7 (1995) 82—89.
[31] R.A. Atchley, S.S. Ilardi, A. Enloe, Hemispheric asymmetry in the processing of emotional content in word meanings: the effect of current and past depression, Brain Lang. 84 (2003) 105—119.
[32] C.S. Emerson, D.W. Harrison, D.E. Everhart, Investigation of receptive affective prosodic ability in school-aged boys with and without depression, Neuropsychiatr. Neuropsychol. Behav. Neurol. 12 (1999) 102—109.
[33] J.W. Stewart, F.M. Quitkin, P.J. McGrath, G.E. Bruder, Do tricyclic responders have different brain laterality? J. Abnormal Psychol. 108 (1999) 707—710.