COMPARISON OF LINEAR CLASSIFICATION METHODS FOR P300 BRAIN-COMPUTER INTERFACE ON DISABLED SUBJECTS

Nikolay V. Manyakov, Nikolay Chumerin, Adrien Combaz, Marc M. Van Hulle

Laboratory for Neuro- and Psychofysiology, K.U.Leuven, Herestraat 49, POBox 1021, 3000 Leuven, Belgium
{NikolayV.Manyakov, Nikolay.Chumerin, Adrien.Combaz, Marc.VanHulle}@med.kuleuven.be

Abstract: In this paper, we investigate the accuracy of linear classification techniques for a P300 Brain-Computer Interface used in a typing paradigm. Fisher's Linear Discriminant Analysis (LDA), Bayesian Linear Discriminant Analysis (BLDA), Stepwise Linear Discriminant Analysis (SLDA), linear Support Vector Machine (SVM) and a method based on Feature Extraction (FE) were compared. Experiments were performed on patients suffering from Amyotrophic Lateral Sclerosis (ALS), middle cerebral artery (MCA) stroke and Subarachnoid Hemorrhage (SAH), in on-line and off-line mode. Our results show that BLDA yields a significantly higher accuracy than the other linear techniques we have compared, at least for our group of subjects.

1 INTRODUCTION

Research on *brain-computer interfaces* (BCIs) has witnessed a tremendous development in recent years (Sajda et al., 2008), and has enjoyed much attention even in popular media. Although a lot of research was done on invasive BCIs, leading to brain implants decoding neural activity directly, which are primarily tested on animals, noninvasive BCIs, *e.g.*, based on *electroencephalograms* (EEG) recorded on the subject's scalp, have recently enjoyed an increasing attention since they do not require any surgical procedure, and can therefore be more easily tested on human subjects. Several noninvasive BCI paradigm have been described in the literature, but the one we concentrate on, relies on the *event-related potential* (ERP, a stereotyped electrophysiological response to an internal or external stimulus (Luck, 2005)).

One of the most explored ERPs is the P300. It can be detected while the subject is shown two types of events with one occurring much less frequently than the other ("rare event"). The rare event elicits an ERP consisting of an enhanced positive-going signal component with a latency of about 300 ms after stimulus onset (Luck, 2005). In order to detect the ERP, the recording of one trial is usually not enough, and the recordings of several trials need to be averaged. Averaging is required because the recorded signal is a superposition of the activity related to the stimulus and all other ongoing brain activity. By averaging, the activity that is time-locked to a known event (*e.g.*, the onset of the attended stimulus) is extracted as an ERP, whereas the activity that is not related to the stimulus onset is expected to be averaged out. The stronger the ERP signal, the fewer trials are needed, and *vice versa*.

There has been a growing interest in the ERP detection problem, as witnessed by the increased availability of BCIs that rely on ERP detection. A notorious example is the P300 mind-typer (Farwell and Donchin, 1988), and with which subjects are able to type words and sentences on a computer screen. This application meets the BCI's primary goal, namely, to improve the quality of life of neurologically impaired patients suffering from pathologies such as: amyotrophic lateral sclerosis, brain stroke, brain/spinal cord injury, cerebral palsy, muscular dystrophy, *etc*. But, as is mostly the case with BCI research, they have been tested primarily on *healthy* subjects. Only very few attempts have been made on *patients* (Nijboer et al., 2008; Sellers and Donchin, 2006; Piccione et al., 2006; Hoffmann et al., 2008; Silvoni

et al., 2009; Sellers et al., 2010). Several of these patient tests (Nijboer et al., 2008; Sellers et al., 2010) deal with P300-based on-line typing, however, since only very few patients were tested, it is still must be investigated whether the P300 mind-typer is suited for them.

In addition, studies that report on the performance of different P300 classifiers were only made for healthy subjects. It, thus, remains to be seen how the comparison will look like for disabled subjects, and how this will affect the choice of the best classifier. This is indeed an important question since the P300 responses from healthy subjects and disabled patients can be quite different (Sellers and Donchin, 2006). Thus, the results of the classification performance comparison for healthy subjects could possibly not be valid for disabled ones. In addition, the outcomes of the comparison, performed on healthy subjects, also lead to slightly different conclusions. In (Krusienski et al., 2006) a comparison of several classifiers (Pearson's correlation method, Fisher's linear discriminant analysis (LDA), stepwise linear discriminant analysis (SLDA), linear support-vector machine (SVM) and Gaussian kernel support vector machine) was performed on 8 healthy subjects. It was shown that SLDA and linear SVM render the best overall performance. In (Mirghasemi et al., 2006) it was shown that, among linear SVM, Gaussian kernel SVM, multi-layer perceptron, LDA and kernel LDA, the best performance was achieved by LDA. Based on these studies, albeit different sets of classifiers were used in the comparison, one can conclude that linear classifiers work better than nonlinear ones, at least for the P300 BCI. This statement is also supported by other researchers (*e.g.*, in (Lotte et al., 2007)).

In this paper, we report on tests performed on a group of partially disabled patients suffering from Amyotrophic Lateral Sclerosis (ALS), Middle Cerebral Artery (MCA) stroke, and Subarachnoid Hemorrhage (SAH). We compare several linear techniques for P300 BCI classification. In addition to the linear techniques mentioned above, we also add two more methods (*i.e.*, Bayesian linear discriminant analysis and a method based on feature extraction). Thus, in our study we compare a much more extensive set of linear classification techniques, and perform our comparison on disabled patients, instead of healthy subjects, both of which distinguishes our approach from others.

2 METHODS

2.1 EEG data acquisition

The EEG recordings were performed using a prototype of an ultra low-power 8-channels wireless EEG system. The wireless EEG system was developed by IMEC[1] and built around their ultra-low power 8-channel EEG amplifier chip (Yazicioglu et al., 2006). The data are transmitted with a sampling frequency of 1000 Hz for each channel. We used a brain-cap with large filling holes and sockets for active Ag/AgCl electrodes (ActiCap, Brain Products). The recordings were made with eight electrodes located primarily on the parietal pole, namely at positions Cz, CPz, P1, Pz, P2, PO3, POz, PO4, according to the international 10–20 system. The reference electrode and ground were placed on the left and right mastoids.

2.2 Experiment design

Twelve subjects, naïve to BCI applications, participated in the experiments (ten male and two female, aged 37–66 with an average age of 51.25). The subjects were suffering from different types of brain disorders. The experimental protocol was approved by the ethical committee. After the recordings were made, four subjects were excluded from further classifier comparison, since their performance was close to chance level, which could be due to the nature of their brain disorder or because they did not understand the experiment. The information about the patients (EEG data of which where used for the analysis) including their diagnoses, age and gender is presented in Table 1.

We have used the same visual stimulus paradigm as the one used in the first P300-based speller, which was introduced by Farwell and Donchin in (Farwell and Donchin, 1988): a matrix of 6×6 symbols. Each experiment was composed of a training and several testing stages. During both stages, columns and rows

Figure 1: Typing matrix of the Mind Speller. Rows and columns are flashed in random order; one trial consists of flashing all six rows and all six columns. The intensification of the third column (left panel) and the second row (right panel) are shown.

Table 1: Information about the patients.

Patient ID	Age	Gender	Diagnosis
subject 1	43	M	Amyotrophic lateral sclerosis (2002). Moderate bulbar palsy. Severe weakness of upper and lower limbs and spasticity in lower limbs.
subject 2	51	M	Right MCA stroke (2008) with hypertension (stage II) and mild left hemiparesis.
subject 3	58	M	Spontaneous SAH and secondary intracerebral hemorrhage in the right hemisphere (2002) with hypertension (stage III) and severe left hemiparesis.
subject 4	54	F	Left MCA stroke (2005) with mild motor aphasia and right hemiparesis.
subject 5	52	M	Posterior circulation stroke (2002). Right hemiparesis with dysarthria.
subject 6	54	M	Left MCA stroke (16.10.2009) with right hemiparesis and motor aphasia.
subject 7	36	M	Acute left MCA stroke with partial motor aphasia, right hemisensory loss.
subject 8	65	M	Right MCA stroke (2008) with hypertension (stage III) and mild left hemiparesis.

of the matrix were intensified (see Figure 1) in a random manner. The intensification duration was set to 100 ms, followed by a 100 ms of no intensification. Each column and each row flashed only once during one trial, so each trial consisted of 12 stimulus presentations.

During the training stage, 11 symbols, taken from the typing matrix, were presented to the subject. For each symbol, 10 intensification for each row/column were performed. The subject was asked to count the number of intensifications of the corresponded symbol. The counting was used only for keeping subject's attention on the symbol.

The recorded data was filtered (in the 0.5–15 Hz frequency band with a fourth-order zero-phase digital Butterworth filter) and properly cut into signal tracks. Each of these tracks consisted of 1000 ms of recording, starting from the stimulus onset. Then, each of these tracks was downsampled, by retaining every 25th sample, and assigned to one of two possible groups: *target* and *nontarget*, according to the stimuli that they were locked to. For classifier training, we constructed a set of 1000 target-, and the same amount of non-target averaged brain responses, where the averages were taken based on k randomly selected responses from the corresponding groups. The number k was equal to the number of intensification sequences (trails), for each stimulus, during the testing stage.

Amplitude values at specific moments in time, of the downsampled EEG signal, restricted to the interval 100–750 ms after stimuli onset, were taken as features. All these features were normalized to their Z-score through the estimation of $f_{n,t} = (x_n(t) - \overline{x_n(t)})/\sigma_{x_n(t)}$, where $x_n(t)$ is the EEG amplitude of n-th channel (electrode) at time t, after the stimulus onset, $\overline{x_n(t)}$ the average of $x_n(t)$ and $\sigma_{x_n(t)}$ the standard deviation for all training examples of both the target and nontarget recordings of the training set. Combining all those features, we obtained a feature vector $\mathbf{f} = [f_1, ..., f_N]^T$, which was used as input for the linear classifier $w_1 f_1 + w_2 f_2 + \cdots + w_n f_n + b = \mathbf{w}^T \mathbf{f} + b$ (see further) [2]. After substitution of the feature vector $\mathbf{f}$ into the abovementioned equation, we obtain a distance (multiplied by factor $\sqrt{\mathbf{w}^T \mathbf{w}}$) from the point in feature space to the separating hyperplane, with the sign indicating to which side of the hyperplane the point belongs, *i.e.*, the target or non-target class.

After training the classifier, each subject performed several on-line test sessions during which (s)he was asked to *mind-type* a few words. The typing performance (ratio of correctly typed symbols) was used for estimating the classification accuracy. For these on-line test sessions, we used the classifier that was trained on data averaged over 15 trials. Thus, each subject attempted to type a symbol based on 15 row/column intensifications. During typing, the EEG data was stored for further off-line analysis based on a smaller amount k of trials (in this case we used all k-combination of 15 trials for each typed letter for assessing the accuracy).

The testing stage differs from the training stage by the way the signal tracks were grouped. During training, the system "knows" exactly which one of 36 possible symbols is attended by the subject at any moment of time. Based on this information, the collected signal tracks can be grouped into only two categories: target (attended) and non-target (not attended). However, during testing, the system does not know which symbol is attended by the subject, and the only meaningful way of grouping is by stimulus type (which in

---

[2]Since we use Z-scores as features, and since we use a balanced training set (equal numbers of target and non-target responses), the parameter b should be close to zero.

the proposed paradigm can be one of 12 types: 6 rows and 6 columns). Thus, during the testing stage, for each trial, we had 12 tracks (from all 12 groups) of 1000 ms EEG data recorded from each electrode. The averaged EEG response for each electrode was determined for each group. The selected features of the averaged data were then fed into the classifier. As a result, the classifier produces 12 (for each row/column) values $(c_1,\ldots,c_{12})$ which describe the distance to a separating hyperplane in the feature space together with the sign. The row index i_r and the column index i_c of the classified symbol were calculated as:

$$i_r = \arg\max_{i=1,\ldots,6}\{c_i\}, \text{ and } i_c = \arg\max_{i=7,\ldots,12}\{c_i\} - 6.$$

The symbol on the intersection of the i_r-th row and i_c-th column in the matrix was then taken as the result of the classification and presented, as a feedback, to the subject in the on-line session.

3 CLASSIFICATION METHODS

3.1 Fisher's Linear Discriminant Analysis

Fisher's Linear Discriminant Analysis (LDA) is one of the most widely used classifiers in P300 BCI systems (Krusienski et al., 2006; Panicker et al., 2010). It was reported that it can even outperform other classifiers (Mirghasemi et al., 2006). Its main idea is to find a projection from the N-dimensional feature space onto a one dimensional space $\mathbf{w}^T\mathbf{f}$ for which the ratio of the variance between the two classes (target and non-target) *vs.* the variance within the classes is maximal. This 'optimal' projection is estimated as $\mathbf{w} = (\Sigma_{-1} + \Sigma_{+1})^{-1}(\mu_{+1} - \mu_{-1})$, where Σ and μ define the covariances and the means of the two classes (target and non-target) that need to be separated.

3.2 Stepwise Linear Discriminant Analysis

Stepwise Linear Discriminant Analysis (SLDA) was used in the patient studies of P300 BCI (Nijboer et al., 2008; Sellers and Donchin, 2006). It can be considered as an extension of LDA with an incorporated filter feature selection. SLDA adds and removes terms from a linear discriminant model based on their statistical significance in regression, thus, producing model that is adjustable to the training data. It was shown that SLDA performs equally well or even better than several other classification methods

in P300 BCI (Krusienski et al., 2006). For our comparison analysis, we have used the same procedure as in (Krusienski et al., 2006) (in the forward step, the entrance tolerance p-value < 0.1; in the backward step, the exit tolerance p-value > 0.15). The process is iterated until convergence, or until it reaches a predefined number of 60 features.

3.3 Bayesian Linear Discriminant Analysis

Bayesian Linear Discriminant Analysis (BLDA) was used for P300 BCI in patients (Hoffmann et al., 2008). It is based on a probabilistic regression network. Assume that the targets t_i (in the case of a classification problem these are $+1$ and -1) are linearly dependent on the observed features $\mathbf{f}^i = [f_1^i,\ldots,f_N^i]^T$ with an additive Gaussian noise term ε_n: $t_i = \mathbf{w}^T\mathbf{f}^i + \varepsilon_i$. Assuming further an independent generation of the examples from a data set, the likelihood of all data is $p(\mathbf{t}|\mathbf{w},\sigma^2) = \prod_{i=1}^N (2\pi\sigma^2)^{-1/2}\exp\left(-\frac{(t_i - \mathbf{w}^T\mathbf{f}^i)^2}{2\sigma^2}\right)$. Additionally to this, we have to introduce a prior distribution over all weights as a zero-mean Gaussian $p(\mathbf{w}|\alpha) = \prod_{j=1}^n \left(\frac{\alpha}{2\pi}\right)^{1/2}\exp\left(-\frac{\alpha}{2}w_j^2\right)$. Using Bayes's rule, we can define the posterior distribution $p(\mathbf{w}|\mathbf{t},\alpha,\sigma^2) = (p(\mathbf{t}|\mathbf{w},\sigma^2)p(\mathbf{w}|\alpha))/p(\mathbf{t}|\alpha,\sigma^2)$, which is Gaussian with mean $\mu = (\mathbf{F}^T\mathbf{F} + \sigma^2\alpha\mathbf{I})^{-1}\mathbf{F}^T\mathbf{t}$ and covariance matrix $\Sigma = \sigma^2(\mathbf{F}^T\mathbf{F} + \sigma^2\alpha\mathbf{I})^{-1}$, where $\mathbf{I}$ is an identity matrix and $\mathbf{F}$ is a matrix with each row corresponding to a training example in feature space, $\mathbf{t}$ a column-vector of true labels (classification) for all corresponding training examples. As a result, our separation plane will have the form $\mu^T\mathbf{f}$. This solution is equivalent to a penalized least-square estimate $E(\mathbf{w}) = \frac{1}{2\sigma^2}\sum_{i=1}^N (t_i - \mathbf{w}^T\mathbf{f}^i)^2 + \frac{\alpha}{2}\sum_{j=1}^n w_j^2$ (Tipping, 2004).

3.4 Linear Support Vector Machine

In P300 BCI research, Support Vector Machine (SVM) is regarded as one of the more accurate classifiers (Thulasidas et al., 2006; Krusienski et al., 2006). The principal idea of a linear SVM is to find the separating hyperplane, between two classes, so that the distance between the hyperplane and the closest points from both classes is maximal. In other words, we need to maximize the margin between the two classes (Vapnik, 1995). Since it is not always the case that the two classes are linearly separable, the linear SVM idea was also generalized to the case where data points are only required to fall within the margin (and even are on the wrong side of the decision boundary) by adding a regularization
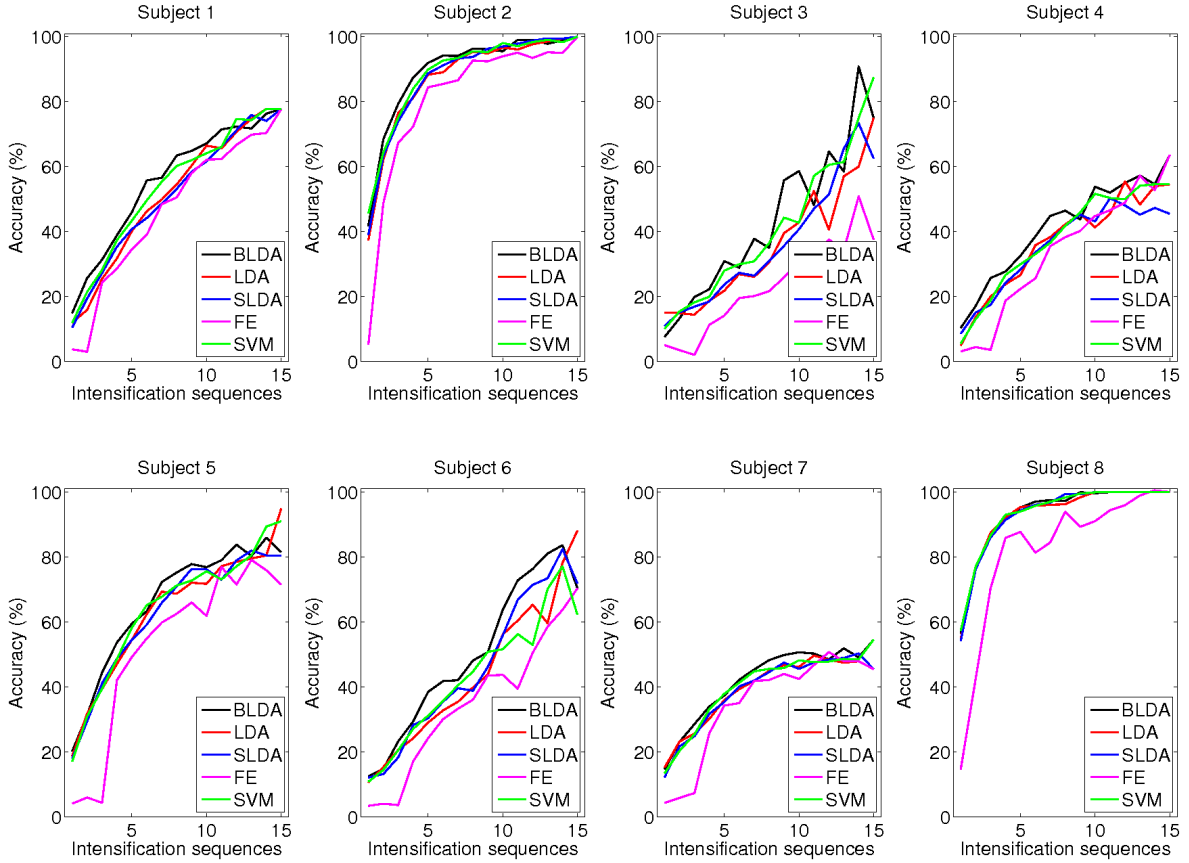
Figure 2: Classification accuracy as a function of the number of intensifications for every subject, and for all discussed classification methods.

term. For our analysis, we used use method proposed in (Combaz et al., 2010), which uses linear least-squares SVM (Suykens et al., 2002) to solve the minimization problem $\min_{\mathbf{w},b,\mathbf{e}}(\frac{1}{2}\mathbf{w}^T\mathbf{w}) + \gamma\sum_{i=1}^{N}e_i^2$ with respect to $y_i(\mathbf{w}^T\mathbf{f}^i + b) = 1 - e_i$, $i = 1,...,n$, where $\mathbf{f}^i$ corresponds to training points in feature space, and y_i is the associated output ($+1$ for the responses to the target stimulus and -1 for the non-target stimulus). The regularization parameter is estimated through a line search on cross-validation results.

3.5 Method based on Feature Extraction

Another classification method in P300 BCI research (Chumerin et al., 2009) relies on the one-dimensional version of a linear *feature extraction* (FE) approach proposed by Leiva-Murillo and Artés-Rodrí-guez in (Leiva-Murillo and Artes-Rodriguez, 2007). The method searches for the "optimal" subspace maximizing (an estimate of) the mutual information be-

tween the set of projections $Y = \{\mathbf{w}^T\mathbf{f}^i\}$ and the set T of corresponding labels $t_i = \{-1,+1\}$. According to (Leiva-Murillo and Artes-Rodriguez, 2007), the mutual information between the set of projections Y, and the set of corresponding labels C can be estimated as: $I(Y,C) = \sum_{p=1}^{N_t} p(t_p)\left(J(Y|t_p) - \log\sigma(Y|t_p)\right) - J(Y)$, with $N_t = 2$ the number of classes, $Y|t_p$ the projection of the p-th class' data points onto the direction $\mathbf{w}$, $\sigma(\cdot)$ the standard deviation, and $J(\cdot)$ the negentropy, estimated using Hyvärinen's robust estimator (Hyvärinen, 1998).

4 RESULTS

The performance results are shown in Figure 2 for individual subjects, and in Figure 3 as a grand average among all subjects. In order to verify the statistical significance of the comparison, we used the nonparametric Friedman's test (Corder and Foreman, 2009) between each pairs of different methods to test the

difference in the medians of the accuracy results. We have found that the accuracy based on BLDA is significantly ($p < 0.001$) better than any other. Linear SVM is second. As for SLDA and LDA, there is no any significant difference between them.

We have also analyzed the mistyped (erroneously detected) symbols [results not shown]. We have found that, for all classification method, the misclassifications mostly occur for either a misclassified row or column, and the erroneously typed symbols are in close proximity on the screen to the desired ones.

We observed that some subjects were not comfortable with the visual stimulation protocol we used during the on-line sessions. This discomfort was expressed by the frequent (3–8 Hz) eye blinking of the subjects. For those subjects, we had to adapt the stimulation protocol in terms of the interstimuli interval, which was increased up to 300 ms (150 ms of intensification followed by 150 ms of no intensification). This shows that working with patients can be quite different.

5 CONCLUSIONS

We have compared five linear classification methods for a P300-based BCI, tested on disabled patients. We have found that BLDA yields significantly better results compared to the other classification methods we considered, with linear SVM as the second one in accuracy. These results can be helpful in deciding what classifier to use for patients. Additionally to this, since the classifiers could produce different outcomes, one could benefit from combining them using a co-training approach (Panicker et al., 2010), to improve the classification performance.
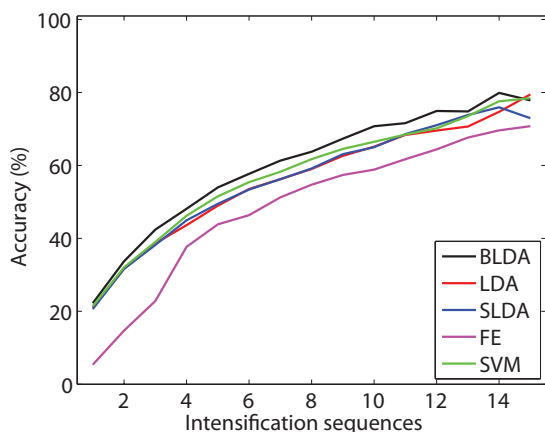

Figure 3: Average classification accuracy as a function of the number of intensifications for all discussed classification methods.

REFERENCES

Chumerin, N., Manyakov, N. V., Combaz, A., Suykens, J. A., Yazicioglu, R. F., Torfs, T., Merken, P., Neves, H. P., Van Hoof, C., and Van Hulle, M. M. (2009). P300 detection based on feature extraction in on-line brain-computer interface. *Lecture Notes in Computer Science*, 5803:339–346.

Combaz, A., Chumerin, N., Manyakov, N. V., Suykens, J., and Van Hulle, M. M. (2010). Error-related potential recorded by eeg in the context of a p300 mind speller brain-computer interface. In *Machine Learning for Signal Processing, IEEE Workshop on*, pages 65–70, Kittilä, Finland.

Corder, G. and Foreman, D. (2009). *Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach*. New York, J. Wiley.

Farwell, L. and Donchin, E. (1988). Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical Neurophysiology*, 70(6):510–523.

Hoffmann, U., Vesin, J.-M., E. T., and Diserens, K. (2008). An efficient P300-based brain-computer interface for disabled subjects. *Journal of Neuroscience Methods*, 167:115–125.

Hyvärinen, A. (1998). New approximations of differential entropy for independent component analysis and projection pursuit. In *Proceedings of the 1997 conference on Advances in neural information processing systems*, pages 273–279. MIT Press Cambridge, MA, USA.

Krusienski, D., Sellers, E., Cabestaing, F., Bayoudh, S., McFarland, D., Vaughan, T., and Wolpaw, J. (2006). A comparison of classification techniques for the P300 Speller. *J. Neural. Eng.*, 3:299–305.

Leiva-Murillo, J. and Artes-Rodriguez, A. (2007). Maximization of mutual information for supervised linear feature extraction. *IEEE Transactions on Neural Networks*, 18(5):1433–1441.

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based Brain-Computer Interface. *Journal of Neural Engineering*, 4:R1–R13.

Luck, S. (2005). *An introduction to the event-related potential technique*. MIT Press Cambridge, MA.

Mirghasemi, H., Fazel-Rezai, R., and Shamsollahi, M. (2006). Analysis of P300 classifiers in Brain Computer Interface speller. In *Proceedings of the 28th IEEE EMBS Annual International Conference*, pages 6205–6208.

Nijboer, F., Sellers, E., Mellinger, J., Jordan, M., Matuz, T., Furdea, A., Halder, S., Mochty, U., Krusienski, D., Vaughan, T., Wolpaw, J., Birbaumer, N., and Kübler, A. (2008). A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, 119:1909–1916.

Panicker, R., Puthusserypady, S., and Sun, Y. (2010). Adaptation in P300 Brain-Computer Interface: A two-classifier co-training approach. *IEEE Trans Biomed Eng*, 57.

Piccione, F., Giorgi, F., Tonin, P., Priftis, K., Giove, S., Silvoni, S., Palmas, G., and Beverina, F. (2006). P300-based brain-computer interface: Reliability and performance in healthy and paralysed participants. *Clinical Neurophysiology*, 117:531–537.

Sajda, P., Müller, K.-R., and Shenoy, K. (2008). Brain-computer interfaces. *IEEE Signal Processing Magazine*, 25(1):16–17.

Sellers, E. and Donchin, E. (2006). A P300-based brain-computer interface: Initial test by ALS patients. *Clinical Neurophysiology*, 117:538–548.

Sellers, E., Vaughan, T., and Wolpaw, J. (2010). A brain-computer interface for long-term independent home use. *Amyotrophic Lateral Sclerosis*, pages 1–7.

Silvoni, S., Volpato, C., Cavinato, M., Marchetti, M., Priftis, K., Merico, A., Tonin, P., Koutsikos, K., Beverina, F., and Piccione, F. (2009). P300-Based Brain-Computer Interface Communication: Evaluation and Follow-up in Amyotrophic Lateral Sclerosis. *Frontiers in Neuroscience*, 3(60):1–12.

Suykens, J., Van Gestel, T., De Brabanter, J., De Moor, B., and Vanderwalle, J. (2002). *Least square support vector machines*. World Scientific, Singapore.

Thulasidas, M., Guan, C., and Wu, J. (2006). Robust classification of EEG signal for brain-computer interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(1):24–29.

Tipping, M. E. (2004). Bayesian inference: An introduction to principles and practice in machine learning. In Bousquet, O., von Luxburg, U., and Rätsch, G., editors, *Advanced Lectures on Machine Learning*, pages 41–62. Springer.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag.

Yazicioglu, R., Merken, P., Puers, R., and Van Hoof, C. (2006). Low-power low-noise 8-channel EEG front-end ASIC for ambulatory acquisition systems. In *The 32nd European Solid-State Circuits Conference. Proceedings of*, pages 247–250.