# AN APPLICATION OF FEATURE SELECTION TO ON-LINE P300 DETECTION IN BRAIN-COMPUTER INTERFACE

*Nikolay Chumerin[1], Nikolay V. Manyakov[1], Adrien Combaz[1], Johan A.K. Suykens[2], Marc M. Van Hulle[1]*

[1] K.U.Leuven, Laboratorium voor Neuro- en Psychofysiologie,
Campus Gasthuisberg, Herestraat 49, B3000 Leuven, Belgium
[2] K.U.Leuven, ESAT-SCD, Kasteelpark Arenberg 10, B3001 Heverlee, Belgium

E-mail:{Nikolay.Chumerin, NikolayV.Manyakov, Adrien.Combaz, Marc.VanHulle}@med.kuleuven.be,
Johan.Suykens@esat.kuleuven.be

## ABSTRACT

We propose a new EEG-based wireless brain computer interface (BCI) with which subjects can "mind-type" text on a computer screen. The application is based on detecting P300 event-related potentials in EEG signals recorded on the scalp of the subject. The BCI uses a linear classifier which takes as input a set of simple amplitude-based features that are optimally selected using the Group Method of Data Handling (GMDH) feature selection procedure. The accuracy of the presented system is comparable to the state-of-the-art systems for on-line P300 detection, but with the additional benefit that its much simpler design supports a power-efficient on-chip implementation.

## 1. INTRODUCTION

Research on *brain computer interfaces* (BCIs) has witnessed a tremendous development in recent years (see, for example, the editorial in Nature [1]), and is now widely considered as one of the most successful applications of the neurosciences. BCIs can significantly improve the quality of life of neurologically impaired patients with pathologies such as: amyotrophic lateral sclerosis, brain stroke, brain/spinal cord injury, cerebral palsy, muscular dystrophy, etc.

Brain computer interfaces are either *invasive* (intra-cranial) or *noninvasive*. The first ones have electrodes implanted into the premotor- or motor frontal areas or into the parietal cortex (see review in [2]), whereas the noninvasive ones mostly employ *electroencephalograms* (EEGs) recorded from the subject's scalp. The noninvasive methods can be further subdivided into three groups. The first group explores visually evoked potentials (VEPs) and they can be traced back to the 70s, when Jacques Vidal constructed the first BCI [3]. This system was used for controlling a cursor on a screen based on the estimation of the subject's direction of gaze. The gaze was estimated from the detection of the harmonics $f$, $2f$ and $3f$ in the Fourier transform of the EEG while the subject was observing the stimuli flickering at a frequency $f$. This idea was further improved by Sutter [4], Middendorf [5], among others.

The second group of noninvasive BCIs rely on the detection of imaginary movements of the right or the left hand. These methods exploit *slow cortical potentials* (SCP) [6, 7], *event-related desynchronization* (ERD) on the mu- and beta-rhythm [8, 9], and the *readiness potential* (bereitschafts-potential) [10]. The detection of other mental tasks (*e.g.*, cube rotation, subtraction, word association [11]) also belong to this group.

The third noninvasive group are the BCIs that rely on the 'oddbal' evoked potential in the parietal cortex. An *event-related potential* (ERP) is a stereotyped electrophysiological response to an internal or external stimulus [12]. One of the most known and explored ERP is the P300. It can be detected while the subject is classifying two types of events with one of the events occurring much less frequently than the other ("rare event"). The rare events elicit ERPs consisting of an enhanced positive-going signal component with a latency of about 300 ms [13]. In order to detect the ERP in the signal, one trial is usually not enough and several trials must be averaged. The averaging is necessary because the recorded signal is a superposition of *all* ongoing brain

activities. By averaging the recordings, those that are time-locked to a known event (*e.g.*, attended stimulus) are extracted as ERPs, whereas those that are not related to the stimulus presentation are averaged out. The stronger the ERP signal, the fewer trials are needed, and *vice versa*. Figure 1 shows an example of two EEG responses (blue and green curves), for the attended target (top panel) and the nontarget (bottom panel) stimuli, as well as the average response (red curve). The BCI system described in this article is an elaboration of the P300-based BCI but with emphasis on a simple design for a power-efficient on-chip implementation, which must use a computationally cheap classification scheme. To this end in the presented system we exploit a linear classifier which takes as input a small set of simple amplitude-based features. The features are selected using the *Group Method of Data Handling* (GMDH) [14] feature selection method, which has not been used in the BCI domain so far.

There has been a growing interest in the ERP detection problem, witnessed by the increased development of BCIs able to transfer brain signals to computers using ERPs (for example the P300 mind-typer [15, 16, 17]). The requirements for ERP detection have also increased. Now the task is not only to be able to detect and classify ERPs but also do it as fast and accurately as possible. Ideally, one would like to be able to robustly detect ERPs from single trials. Unfortunately, this is still beyond reach.

A number of off-line studies have been reported that improve the classification rate of the P300 speller [18, 19, 20], but not much work has been done on *on-line* classification (which is out of scope of the BCI competition). To the best of our knowledge, the best classification rate for *on-line* mind-typers is reported in [17]. The results of our work are compared to the latter.

## 2. METHODS

### 2.1. Acquisition hardware

The EEG recordings were performed using a prototype of an ultra low-power 8-channel wireless EEG system, which consists of two parts: an amplifier coupled with a wireless transmitter (see Fig. 2) and a receiver (see Fig. 3). The prototype was developed and provided to us by the *Interuniversity Microelectronics Centre* (IMEC)[1]. We used a braincap with large filling holes and sockets for ring electrodes.

The IMEC wireless EEG system interfaces to the PC via an USB stick receiver (see Fig. 3), which uses a FTDI FT232BM serial USB converter. The EEG system can be accessed by a virtual serial port which behaves exactly like a conventional serial port, except that the communication
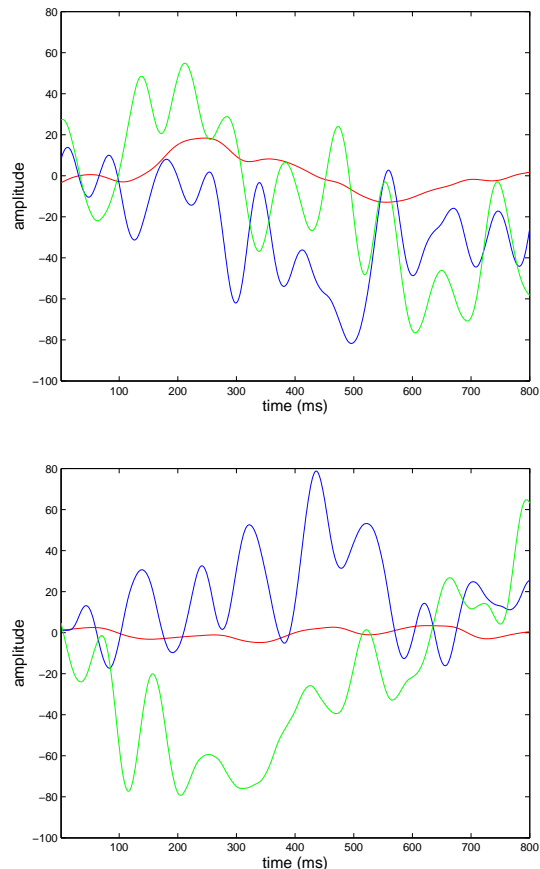
---

[1]http://www.imec.be/



**Fig. 1**. Examples of the event-related EEG responses to the target (top panel) and nontarget (bottom panel) stimuli recorded from the first subject at electrode CPz. The blue and green curves depict pairs of randomly-chosen EEG waveforms, while the red curves show signals averaged over the whole recording session. All recordings are filtered in the 0.5–15 Hz frequency band. Time $t = 0$ marks the stimulus onset.

software has to be able to handle non-standard (higher) baud rates.

After it is switched on, the EEG system works in a *configuration mode*. In this mode, the user can send configuration commands (*e.g.* EEG gain, bandwidth, impedance measurement) to the system. After sending the configuration commands, the system can be switched (or after 30 seconds have elapsed without sending any commands, it switches automatically) into the *measurement mode*. In this mode, every 2 milliseconds (500 Hz), the EEG system transfers a data frame. Each frame is sized 27 bytes, and consists of a synchronization byte, a frame counter byte, a battery voltage byte and two samples of EEG data (each sample has 8 values, each value is stored in 12 bits), so the actual

**Fig. 2**. Wireless 8 channel EEG device (amplifier and transmitter).



**Fig. 3**. USB stick receiver, plugged into the extension cable.

sampling rate is 1000 Hz per channel.

A comprehensive technical description of the IMEC wireless EEG system can be found in [21].

## 2.2. Acquisition procedure

Recordings were collected from eight electrodes in the occipital and parietal areas, namely in positions Cz, CPz, P1, Pz, P2, PO3, POz, PO4, according to international 10–20 system. The reference electrode and ground were linked to the left and right mastoids.

Each experiment started with a pause (approximately 90 s) needed for EEG signal stabilization. During this period, the EEG device transmits data but it is not recorded. The data for each symbol presentation was recorded in one recording session. As the duration of the session was known *a-priori* , as well as the data transfer rate, it was easy to estimate the amount of data transmitted during a session. We used this estimate, increased by a $10\%$ margin, as the size of the serial port buffer. To make sure that the entire recording session for one symbol fits completely into the buffer, we cleared the buffer just before recording. This trick allowed us to avoid broken/lost data frames, which usually occur due to a buffer overflow. Unfortunately, sometimes data frames still are lost because of a bad radio signal. In such cases, we used the frame counter to reconstruct the lost frames, using a simple linear interpolation.

## 2.3. Data-stimuli synchronization

Unlike a conventional EEG systems, the system we used does not have any external synchronization inputs. We tried to use one of the channels for this purpose (connecting it to a photo-sensor attached to the screen), but this scheme was not stable enough for long recording times. Finally, we came up with an "internal" synchronization scheme based on high-precision (up to hectananosecond) timing[2].

For the synchronization, we saved the exact time stamps of the start and end of the recording session, as well as the time stamps of stimulus onsets and offsets. Due to the fact that the reconstructed EEG signal has a constant sampling rate, it is possible to find very precise correspondences between time stamps and the data samples. We used this correspondence mapping for partitioning the EEG signal into signal tracks, for further processing.

## 2.4. Experiment design

Four healthy male subjects (aged 23–36 with average age of 31, three righthanded and one lefthanded) participated in the experiments. Each experiment was composed of a training and several testing stages.

We used the same visual stimuli paradigm as in the first P300-based speller, which was introduced by Farwell and Donchin in [15]: a matrix of $6 \times 6$ symbols. The only (minor) difference was in the type of symbols used, which in our case was a set of 26 latin characters, 8 digits and two special symbols '␣' (used instead of space) and '¶' (used as an *end of input* indicator).

During the training and testing stages, columns and rows of the matrix were intensified (see Fig. 4) in a random manner. The intensification duration was 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.

As it was mentioned above, one trial is not enough for robust ERP detection, and we adopted the common practise of averaging the recordings over several trials before performing the classification of the (averaged) recordings.

During the training stage, all 36 symbols from the typing matrix were presented to the subject. Each symbol had 10 trials of intensification for each row/column (10-fold averaging). The subject was asked to count the number of

---

[2]TSCtime high-precision time library by Keith Wansbrough.

intensifications of the corresponded symbol. The counting was used only for keeping subject's attention to 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. Note that subsequent tracks overlap in time, since the time between two consequent stimuli onsets is 200 ms. Then, each of these tracks was downsampled to 30 tabs and assigned to one of two possible groups: *target* and *nontarget* (according to the stimuli, which they were locked to).

Amplitude values at certain moments in time of the downsampled EEG signals were taken as features. All these features were normalized to $[0, 1]$ through the estimation of $f_{n,t} = \frac{x_n(t) - x_{\min,n}(t)}{x_{\max,n}(t) - x_{\min,n}(t)}$, where $x_n(t)$ is the EEG amplitude of $n$-th channel (electrode) at time $t$ after the stimulus onset, $x_{\max,n}(t)$ and $x_{\min,n}(t)$ are the maximal and minimal values of the EEG amplitudes of the $n$-th channel at a moment of time $t$ after stimulus onset among all target and nontarget recordings from the training set. Having combined such features from all EEG recording channels, feature selection was performed using the *Group Method of Data Handling* (see further for a description). As an external criterion, a 5-fold cross validation was used. A *linear discriminate analysis* (LDA) was chosen for the classification. No more than 20 features were selected for the classifier. Basing on LDA we have also estimated coefficients $a_i$ and $b$, where $i = 1, \ldots, n$ and $n$ is a number of the selected features $f_i$, of the hyperplane $a_1 f_1 + a_2 f_2 + \cdots + a_n f_n + b = 0$, which separate two subsets of the data (namely the target and nontarget subsets). After substitution of the feature values $f_i$ into the right hand side of the abovementioned equation, we obtain a distance (multiplied by factor $\sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$) from the point $(f_1, f_2, \ldots, f_n)$ in feature space to the separating hyperplane with a sign, indicating one of two subspaces from both sides of the hyperplane. This sign is an indicator of belonging to one of the trained groups *e.g.*, '+' for target and '−' for non-target.

After training, all coefficients $a_i$ and $b$, together with the amplitude position in time (time offset), the selected electrodes, and the normalization coefficients (namely $x_{\max,n}(t)$ and $x_{\min,n}(t)$), were taken as features for the on-line classification.

After training the classifier, each subject performed several test sessions and was asked to *mind-type* a few words (about 30–50 symbols), the performance of which was used for estimating the classification accuracy. For each test session, we used the classifier that had been trained on data averaged over a given number of trials. The number of trials ($k$) that was used for averaging varied from 2 to 10. The classification accuracy for each value of $k$ was measured.

The testing stage differs from the training stage not only



**Fig. 4**. Typing matrix of the mind-typer. 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.

by the classification step, but also by the way of grouping the signal tracks. 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 the proposed paradigm can be one of 12 types: 6 rows and 6 columns). So, 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.

## 2.5. Feature selection

In order to optimize the set of features, by selecting a subset of them, we use a feature selection procedure called the *Group Method of Data Handling* (GMDH) [14], which is a breadth-first search algorithm working as a wrapper that minimizes the hold-out error. This algorithm constructs, for each iteration $i$, the set $S_i$, of cardinality $n$, of the best subsets $C_{ij}$ (where $j = 1, \ldots, n$). This means that $S_i = \{C_{i1}, C_{i2}, \ldots, C_{in}\}$ (in the first step $S_1$ consist of the $n$ best discriminative features). Each of these subsets $C_{ij}$ consists of $i$ features from the whole feature space with dimension
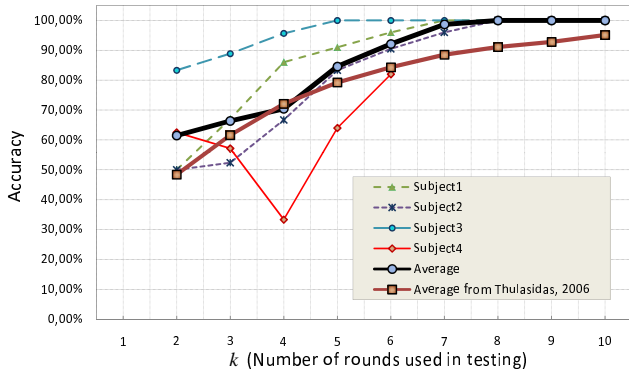
**Fig. 5**. Accuracy of classification for different subjects as a function of the number of trials used in testing. Averaged result and result from [17] are also plotted.

$N$. The transition from the $i$-th iteration to the next $(i + 1)$-th, causes a new set of $n(N - i)$ groups of features to be constructed by generating for each of the $n$ subsets $C_{ij}$, a collection of subsets consisting of the entire subset of $C_{ij}$, with the addition of one of the $(N - i)$ missing features. From the subset received in this way, the best $n$ subsets are chosen by an external criterion, to generate a new set $S_{i+1}$. As a stopping criterion, the absence of an increase in performance in subsequent $d$ iterations is used. As a result, we take the best subset in the latter $d$ iterations.

## 3. RESULTS AND DISCUSSION

The performance of each subject in mind-typing with our system is displayed in Fig. 5, where the percentage of correctly-typed symbols is plotted versus the number of trials $k$ used for averaging. The average performance of all subjects, as well as the average performance of the bestto our knowledge on-line mind-typing system described in the literature [17], are also plotted (this should not be confounded with the BCI competition, which is about off-line classification [18, 19, 20]). It should be mentioned that the mind-typing system of Thulasidas and co-workers is based on a *support-vector machine* (SVM) classifier, where model selection (for kernel parameter and regularization constant) is done using a grid-search procedure, and does not use feature selection. The training of the SVM classifier takes substantially longer time than the feature selection and the training of the linear classifier used in our system. Another consideration is that the on-chip implementation of the SVM classifier is much more complex than our solution, due to the presence of nonlinearities (kernel-based functions).

As it is clear from Fig. 5, the performance strongly depends on the subject. From our experiments, we found that:

- the accuracy decreases with increasing subject age,

- the more "emotional" subject is, the more detectable his/her P300 is.

But we hasten to add that it is impossible to draw any statistically-grounded conclusions from only four subjects. For this, many more experiments needed to be performed.

## 4. CONCLUSION

The brain-computer interface (BCI) presented in this article allows the subject to type text by detecting P300 potentials in the recorded EEG signals. The system consists of a linear classifier that uses a limited number of normalized amplitude-based features as input. The simplicity of the proposed system supports an efficient on-chip implementation (*e.g.*, on ASIC chip). The developed in Matlab software can successfully handle data frame losses, which often occur during wireless transmission.The results of this study shows that, in the field of BCIs based on event-related potentials (ERPs), even simple solutions can successfully compete with the state of the art, given that a feature selection is performed.

## 5. REFERENCES

[1] "Editorial comment: Is this the bionic man?," *Nature*, vol. 442, no. 7099, pp. 109, July 2006.

[2] B. Pesaran, S. Musallam, and R.A. Andersen, "Cognitive neural prosthetics," *Current Biology*, vol. 16, no. 3, pp. 77–80, 2006.

[3] J.J. Vidal, "Toward direct brain-computer communication," *Annual review of Biophysics and Bioengineering*, vol. 2, no. 1, pp. 157–180, 1973.

[4] E.E. Sutter, "The brain response interface: communication through visually-induced electrical brain responses," *Journal of Microcomputer Applications*, vol. 15, no. 1, pp. 31–45, 1992.

[5] M. Middendorf, G. McMillan, G. Calhoun, and K.S. Jones, "Brain-computer interfaces based on the steady-state visual-evoked response," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 211–214, 2000.

[6] A. Kübler, B. Kotchoubey, J. Kaiser, J.R. Wolpaw, and N. Birbaumer, "Brain-computer communication: unlocking the locked in," *Psychological Bulletin*, vol. 127, no. 3, pp. 358–375, 2001.

[7] N. Birbaumer, A. Kubler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchoubey, N. Neumann, and H. Flor, "The

thought translation device (TTD) for completely paralyzedpatients," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 190–193, 2000.

[8] J.R. Wolpaw, D.J. McFarland, and T.M. Vaughan, "Brain-computer interface research at the Wadsworth Center," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 222–226, 2000.

[9] G. Pfurtscheller, C. Guger, G. Müller, G. Krausz, and C. Neuper, "Brain oscillations control hand orthosis in a tetraplegic," *Neuroscience letters*, vol. 292, no. 3, pp. 211–214, 2000.

[10] B. Blankertz, G. Dornhege, M. Krauledat, K.R. Müller, and G. Curio, "The non-invasive Berlin brain–computer interface: fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.

[11] J. del R. Millán, F. Renkens, J. Mouriño, and W. Gerstner, "Noninvasive brain-actuated control of a mobile robot by human EEG," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1026–1033, 2004.

[12] S.J. Luck, *An introduction to the event-related potential technique*, MIT Press Cambridge, MA:, 2005.

[13] W.S. Pritchard, "Psychophysiology of P300," *Psychological Bulletin*, vol. 89, no. 3, pp. 506, 1981.

[14] A.G. Ivakhnenko, "The group method of data handling — a rival of the method of stochastic approximation," *Soviet Automatic Control*, vol. 13, no. 3, pp. 43–55, 1968.

[15] L.A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[16] E. Donchin, K.M. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 174–179, 2000.

[17] M. Thulasidas, C. Guan, and J. Wu, "Robust classification of EEG signal for brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 1, pp. 24–29, 2006.

[18] Kaper M., Meinicke P., Grossekathoefer U., Lingner T., and Ritter H., "BCI Competition 2003 - Data set IIb: support vector machines for the P300 speller paradigme," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1073–1076, 2004.

[19] Serby H., Yom-Tov E., and Inbar G.F., "An improved P300-based brain-computer interface," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, pp. 89–98, 2005.

[20] N. Xu, Gao X., Hong B., Miao X. Gao S., and Yang F., "BCI Competition 2003 - Data set IIb: enhancing P300 wave detection using ICA-based subspace projections for BCI applications," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 2, pp. 1067–1072, 2004.

[21] R.F. Yazicioglu, T. Torfs, P. Merken, J. Penders, V. Leonov, R. Puers, B. Gyselinckx, and C. Van Hoof, "Ultra-low-power biopotential interfaces and their applications in wearable and implantable systems," *Microelectronics Journal*, 2008.