EEG signal pre-processing for the P300 speller

No Author Given

No Institute Given

Abstract. One of the workhorses of Brain Computer Interfaces (BCI) is the P300 speller, which allows a person to spell text by looking at the corresponding letters that are laid out on a flashing grid. The device functions by detecting the Event Related Potentials (ERP), which can be measured in an electroencephalogram (EEG), that occur when the letter that the subject is looking at flashes (unexpectedly). In this work, after a careful analysis of the EEG signals involved, we propose a preprocessing method that allows us to improve on the state-of-the-art results for this kind of applications. Our results are comparable, and sometimes better, than the best results published, and do not require a feature (channel) selection step, which is extremely costly, and which must be applied to each user of the P300 speller separately.

1 Introduction

Brain signals detected using non-invasive methods such as electroencephalograms (EEG) (Figure 1) provide a very rough summary of the overall activity of the brain at different locations of the scalp. Event Related Potentials are relatively strong signals that can be detected when an event that is significant to the subject occurs. The P300 ERP (which stands for Positive peak at 300ms) is thought to occur when such event is both relevant to the task that the subject is performing, and unexpected. This principle has been applied to construct the so-called "P300 speller" (see Figure 2), which allows a subject to spell text by focusing on each individual letter, one at a time, and waiting for it to flash on a screen. If such flashes are unpredictable, a P300 occurs, which hints the device as to which letter the subject is looking at. In practice, since the noise and interference dominate the signal, P300 events are very hard to detect. Therefore, each letter must usually be flashed several times before an automatic decision can be made. Some devices arrange the letters on a rectangular grid and flash entire rows and columns at a time, which increments the number of times that each letter is flashed per time unit.

In 2006, an open challenge called the BCI Competition III was proposed. Its goal was to obtain the best possible performance (in correct letter classification) on a dataset obtained using a speller on two different subjects. The winner of the competition was the method proposed in [1], which combines several mainstream machine learning techniques. The method will be described in full detail in Section 2.

In this work, we perform an in-depth signal processing-oriented analysis of the EEG signals produced in P300 speller systems. In particular, we focus on the ones obtained from the BCI Competition.[1] The driving question behind our work is: how much can we simplify and/or robustify a speller system by applying a priori knowledge about the EEG signals involved? The result of this work is twofold: first, we are able to improve on the state-of-the-art by exploiting such prior information instead of relying on a pure black-box approach such as [1]; second, we provide evidence supporting the hypothesis that there is a significant amount of underlying information, beyond the P300 ERPs, that is needed for a successful discrimination between positive and negative events. The latter conclusion is obtained by classical signal-theoretic results from synchronous detection theory.

2 Background

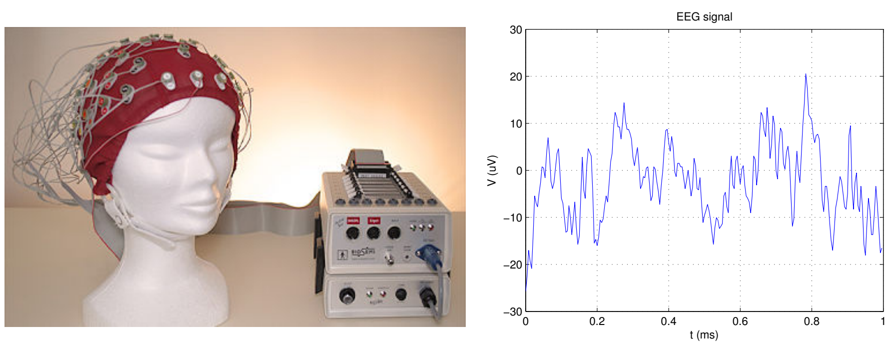

Fig. 1: Left: sample EEG measurement device. Right: typical EEG signal from one channel.

In this section we describe common aspects of EEG signals, the P300 speller, and the approach followed in [1] to infer a letter to be spelled from the EEG signals read from the scalp of the subject.

Figure 1 shows a typical EEG measurement device. The EEG signal is captured by several electrodes distributed over the scalp of a subject. These electrodes measure the electromagnetic field, at various points on the surface of the scalp, that is produced by the neural activity of the brain. The distribution of such points varies from device to device although some standards exist. The system discussed in this work adheres to the 10-20 standard for EEG electrode location [2]. The signal measured at each electrode is called a *channel*. Due to the conductive interface between the signal to be measured and the transducing

[1] Dataset available at `http://www.bbci.de/competition/iii`.

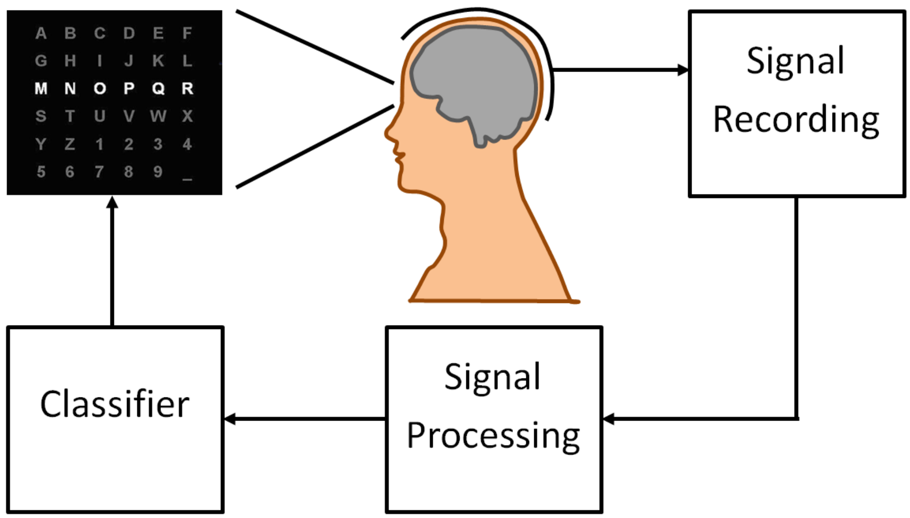

Fig. 2: P300 speller diagram. All possible letters are laid out on a square grid, displayed on a computer screen. All rows and colums are flashed, one at a time, in random order, while the subject stares at the desired letter. Meanwhile, the neural activity of the subject is captured using an EEG device, pre-processed, and then fed to a classification system which infers the letter that the subject is looking at.

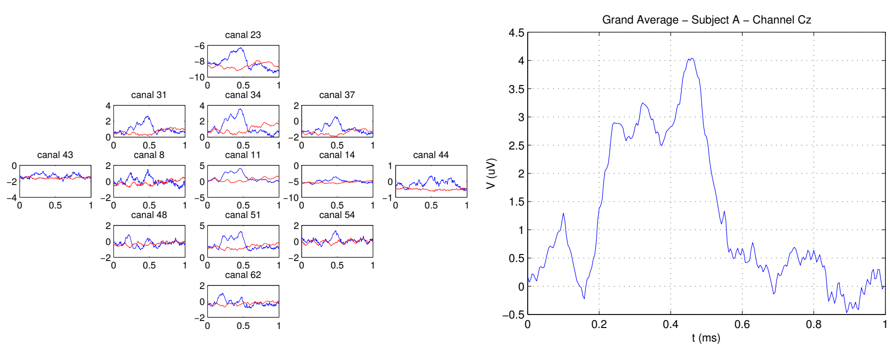

Fig. 3: Average P300 waveform for subject A. Left: on all 64 channels. Right: detail for the Cz channel, which is one channels where the P300 ERP manifests itself with more strength. In this case, on the Cz channel, the maximum potential is not achieved at 300ms, but rather at 450ms. However, the overall peak seems to be centered around 300ms.

electrodes, the resulting signal is low-pass filtered both in space (as a function of the position on the scalp), and time. Finally, due to the very small potentials involved, and the high amplification needed, the resulting signal to noise ratio (SNR) on all channels is usually very low. Other problems derived from the measurement mechanism includes common noise accross electrodes (channels). (We refer the reader to [3, 4] for details on the subject.)

The speller device used in the BCI Competition III (depicted in Figure 2) consists of a screen with a set $\mathcal{Y}$ of 36 characters arranged on a 6×6 grid, coupled to an EEG measurement machine with 64 channels distributed according to the 10-20 standard [2]. For the competition, the following experiment was performed on two different subjects, which we call "subject A" and "subject B". While a subject stares at some specific letter on the screen, each of the 6 rows and each of the 6 columns is flashed separately. This cycle of 12 flashes is repeated 15 times, for a total of 12×15 flashes, where a different random order is selected each time for the flashing of rows and columns. This procedure in turn is repeated for a series of 185 letters; we refer to each of these 185 repetitions as an *epoch*. The first 85 epochs are reserved for training; the remaining 100 are exclusively for testing.

Beginning with each flash, the EEG signal of the 64 channels is sampled for a duration of one second, at a precision of 12 bits per channel, at a sampling frequency of 240Hz. The resulting matrix of 240×64 signal samples constitutes one *data sample*, which we denote by $\mathbf{X} = \{x_{ik}\}$, with x_{ik} being the voltage measured for channel k at discrete time i (relative to the beginning of the flash). Each data sample $\mathbf{X}_j$ (where j denotes a data sample time index) is labeled with the letter $Y_j \in \mathcal{Y}$ that the subject is looking when the data is sampled.

The method proposed in [1] consists of a combination of various machine learning techniques, together with a standard pre-filtering of the signals. To begin with, the method considers only the first 667ms of the signal, discarding the remainig 333ms. It then applies a low-pass filter of cutoff frequency $f_c = 10$Hz followed by a subsampling of $12 : 1$, after which each data sample $\mathbf{X}_j$ is reduced to an 64×14 matrix. The system is trained on each subject separately, using the 85 training epochs of the dataset, and tested, only on that same subject, with the 100 testing epochs of the dataset.

Training of all parameters is done via a cross-validation[5]/classifier aggregation variant where the training subset is divided into 17 segments, and each segment is used to train a different (linear) Support Vector Machine (SVM) [6, 7]. This training includes the choice of the optimum parameter "C", as well as the optimum subset of channels (columns of the data samples $\mathbf{X}$) from which to train the SVM, and of course the best SVM for that setting.

Training proceeds as follows. The subset of channels is chosen via backward selection. In turn, for each candidate subset, differen SVMs are trained using different values of C, and the best one is kept. In all cases, the cost function to be minimized is the error rate on the remaining 16 subsets.

Finally, the best 17 SVMs are combined into one classifier by linearly adding their scores, and selecting the letter with the highest associated cumulative score.

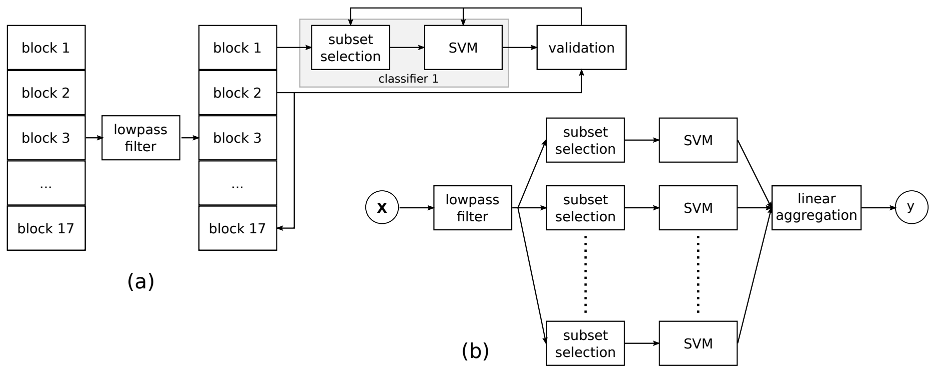

Fig. 4: Classifier architecture proposed in [1]. a) The training dataset is divided into 17 equal-sized, non-overlapping subsets, and 17 SVMs are trained with each one of them. The parameters of each SVM, including the selection of the best subset of channels on which to apply the SVM, is learned independently, using the other 16 subsets as validation data. b) For classification, the output of all 17 SVMs is linearly added to produce an average score, which is then used to select the candidate letter.

A diagram of the architecture just described is shown in Figure 4.

From the above description, two things should be immediatly clear. First, the training procedure is notoriously costly, as each step in the backward selection of each SVM consists in turn of the training and testing of several SVMs. (Once trained, however, detection is very fast, as only a few linear operations are required). Second, the total number of parameters is quite high, which makes the obtained detector extremely overfitted to a particular user. Although cross-performance between subjects was not the goal of the competition, it is nevertheless interesting to see how universal such system could be.

3 Adding a-priori information to improve P300 spellers

As mentioned in the introduction, the focus of this work is on a priori information about EEG signals for P300 speller detection. The a priori information that is usually assumed about EEG signals (see [4] for a review on the subject) includes, as is generally the case, a characterization of what is signal, and what is noise. The noise, as in most applications, is assumed white and uncorrelated. The signal of interest, on the other hand, is considered a band-limited linear superposition of various sub-signals related to specific neural phenomena such as alpha and beta waves, electrooculomotor (EOG) impulses, and ERPs.

In the case of P300 spellers, as their name suggests, the main hypothesis behind their design is that positive events (that is, “the row or column that the user is looking at flashes”) produce a positive ERP 300ms after the flash occurs.

Incidentally, another component that is usually present in P300 speller systems is the so-called Steady State Visual Evoked Potential (SSVEP), which occurs in response to a periodic visual stimulus. In the case of the BCI Competition speller experiment, the row and column flashes, which are produced at a constant rate of $5.7Hz$, are the cause of such sub-signals. Clearly, for a speller application, such SSVEP is to be considered interference.

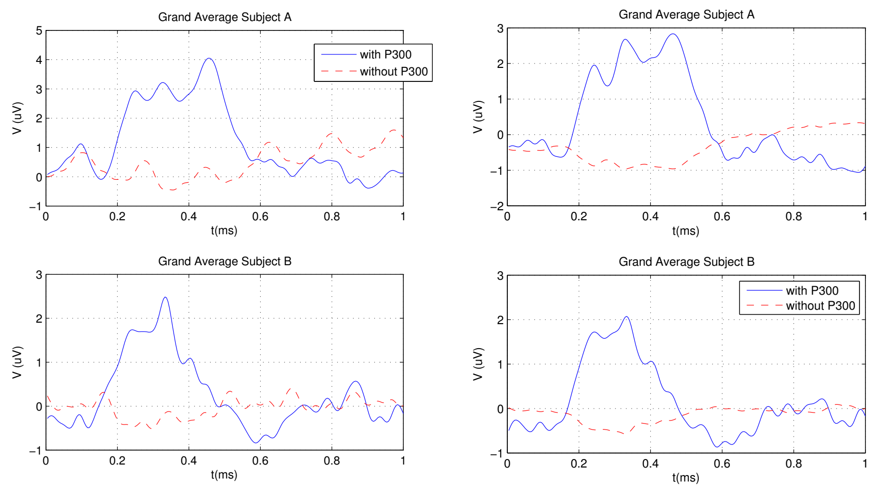

Fig. 5: Removal of SSVEP: Left: Average positive (blue, continuous) and negative (red, dotted) signals for subjects A (above) and B (below). Right: same signals after SSVEP removal. The SSVEP can be clearly seen as a periodic component on the average negative signals on both graphs on the left. Notice that both the average P300, as well as the SSVEP, vary significantly between both subjects.

According to the above scenario, and using [1] as the reference method, we propose three approaches to exploit the existing (or assumed) a priori information about EEG signals, with the hope to improve the speller performance:

1. A synchronous detector of the P300 pulse waveform
2. Pre-filter the signal using the P300 waveform as a matched filter and feed the result to the speller of [1]
3. Remove the SSVEP from the EEG signal and feed the result to the speller of [1]

In the first case, we constructed a synchronous detector by modeling the overall P300 pulse waveform (one per channel) from the grand average of all positive events (see Figure 3). Denote by $\mathcal{Y}^r$ the characters of $\mathcal{Y}$ on the r-th row, and by $\mathcal{Y}_c$ the characters on the c-th column. Denote by $r_j \in \{0, 1, 2, \ldots, 6\}$ and

$c_j \in \{0, 1, 2, \ldots, 6\}$ the index of the row or column flashed during sample j (if a row is flashed, we let $c_j = 0$ and vice versa). The P300 waveform is estimated as the average signal measured each time either $\mathcal{Y}^r$ or $\mathcal{Y}_c$ contains Y_j:

$$\mathbf{Z} = \sum_j \mathbf{X}_j \mathbf{1}\left(\{Y_j \in \mathcal{Y}^{r_j}\} \cup \{Y_j \in \mathcal{Y}_{c_j}\}\right), \tag{1}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function associated to an event.

Note that we are making a strong assumption here: that the shape and position of the pulse is always the same. Deviations from such assumptions may deteriorate the estimation of the matched filter $\mathbf{Z}$. The detection procedure, according to synchronous detection theory, is to measure the filter response at the peak of the matched filter. Note that the filter $\mathbf{Z}$ is multi-channel, each column of it being a classical one-dimensional matched filter (for example, the one corresponding to the Cz channel is shown in Figure 3 on the right):

$$\zeta_j = \sum_k \sum_{i=1} (\mathbf{X}_j)_{ik} \mathbf{Z}_{ik} \,. \tag{2}$$

Denote by J a given epoch. Similar to (1), the overall score for a candidate letter Y occuring during epoch J is given by

$$\zeta(Y) = \sum_{j \in J} \zeta_j \mathbf{1}\left(\{Y_j \in \mathcal{Y}^{r_j}\} \cup \{Y_j \in \mathcal{Y}_{c_j}\}\right). \tag{3}$$

As evidenced by the results in Table 1, the above procedure yields very poor results, which point out the weaknesses behind the basic assumptions about the P300 ERP in its role for detecting significant events. This may occur at two levels: either the P300 ERP is too variable itself (besides what can be assumed interference) to be summarized as an average waveform, either in shape or in location, or there is more information besides what may be called "P300" that is related to a positive event. The second detector proposed, which pre-filters the EEG signals prior to introducing it into the machinery proposed in [1], supports the above conclusion. Although synchronicity is not required in this case, variations in the occurence of the P300 peak may introduce a significant blur in the resulting matched filter, with a negative impact on the overall process.

The third variant is based on the observation that the periodic flashes that occur throughout the entire experiment induce a Steady State Visual Evoked Potential, which manifests itself as a periodic waveform of the same frequency as the flashing rate; this is clearly visible in Figure 5, left column. We remove this interference by estimating the periodic component of the signal with period $5.7Hz$ and then substracting that component from the original signal. The result can be observed on the left column of Figure 5.

4 Results, discussion and conclusions

By performing the aforementioned operation as a pre-processing step to the speller of [1], we observe gains in several aspects. The most important one is that we are able to significantly improve upon the performance of [1] when no channel selection is performed (and all channels are used); to give some perspective, using the implementation provided by the authors of [1], this reduces the training time from over an entire day to a few minutes. Moreover, for subject A, we even improve on the *best* result that can be obtained after the selection procedure. For subject B, the performance drops slightly (only two more samples are missclassified). When combining our pre-filtering with the full training of [1], we maintain the performance on subject A, and come closer to that of subject B. As such small differences could easily be due to random fluctuations, we conclude that the pre-filtering method proposed is able to produce essentially the same results as the original algorithm, while reducing its training time dramatically. Given that this training must be performed on each new subject, such reduction is clearly welcome.

Subject	A	B	C	D	E	F
A	97	33	83	96	94	**98**
B	**96**	34	61	95	92	94

Table 1: Summary of results, given as the number of correct letter identification obtained on the BCI Competition III testing dataset, which consists of 100 epochs. A: results from [1]; B: synchronous detector results, C: results obtained with [1] when the matched P300 filter is used to pre-filter the input; D: method from [1] when the SSVEP component is removed from the input; E: [1] with no channel selection; F: [1] with no channel selection, with the SSVEP component removed from the input

References

1. Rakotomamonjy, A., Guigue, V.: BCI Competition III: Dataset II - ensemble of SVMs for BCI P300 speller. IEEE Trans. Biomed. Eng. (2007)
2. American Electroencephalographic Society: Guidelines for standard electrode position nomenclature. J. Clin. Neurophysiol. **8** (1991) 200–202
3. Vidal, J.: Toward direct brain-computer communication. Annual Review of Biophysics and Bioengineering **2** (1973) 157–180
4. Wolpaw, J., Birbaumer, N., Heetderks, W., McFarland, D., Peckham, P., Schalk, G., E.Donchin, a Quatrano, L., Robinson, C., , Vaughan, T.: Brain-computer interface technology: a review of the first international meeting. IEEE trans. on rehabilitation eng. **8** (2000) 164–73
5. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence – Volume 2, Morgan Kaufmann (1995) 1137–1143
6. Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning **20**(3) (1995) 273–297
7. Smola, A., Schölkopf, B.: A tutorial on support vector regression. Statistics and Computing **14** (2004) 199–222