New Methods for the P300 Visual Speller

Lab Report 1
on the first lab rotation from 29.07.2006 - 25.09.2006

Dept. Empirical Inferences
MPI for Biological Cybernetics, Tübingen

Supervisor: Jeremy Hill

Felix Bießmann

November 3, 2006

Abstract

Brain-Computer Interfaces (*BCI*'s) enable us to infer intentional control signals from brain activity. The Visual Speller is a *BCI* based on event related potentials (*ERP*'s) in the electroencephalogram, such as the P300 (a positive deflection in the EEG about 300 ms after a rarely occuring stimulus). In the classical paradigm one trial (i.e. prediction of one symbol) consists of successive highlightings of one or more symbol(s) on a visual grid presented to the subject. The stimulus events in which the symbol of interest was highlighted will result in an enhanced ERP. This ERP, being stronger than the ERP's elicited by non-target stimulus events, can be used for prediction of the letter the subject was focussing on using some machine learning algorithm, for example the support vector machine. The more symbols are highlighted simultaneously the faster the speller could potentially work. A stimulus code that uses few events per trial (and thus shows many symbols at once) is called dense. The tradeoff against code density is that the signal to noise ratio becomes worse with increasing stimulus frequency: the P300 signal is reported to be strongest when the target symbol frequency is lowest. The stimulus code in which only one symbol per stimulus event is presented, is a maximally sparse code.

It has been proposed that high bitrates of information transfer in a visual speller can best be achieved with sparse stimulus codes. However sparse codes have long trial durations. In order to improve the information transfer rate, we tried to use denser stimulus codes that present fewer stimulus events per trial. To investigate the effect of stimulus type on classification accuracy and the interdependence of stimulus code and type, we explored new stimulus types including ones exploiting recent findings in neuropsychology, such as change blindness and isoluminant color motion. We show that, using appropriate stimuli, denser codes, and hence fewer stimulus events, yield sufficient classification accuracy to achieve competitive bitrates.

Contents

1 Introduction **4**

2 Methods **5**
2.1 Stimulus preparation . . . 5
2.1.1 Sparse vs. dense codes . . . 5
2.1.2 Classical Stimulus . . . 6
2.1.3 Change Blindness . . . 7
2.1.4 Isoluminant Color Motion . . . 8
2.2 EEG recordings . . . 9
2.3 Analysis . . . 9
2.3.1 Preprocessing . . . 9
2.3.2 Classification . . . 10
2.4 Performance measures . . . 11

3 Results **12**
3.1 Averaged ERP's . . . 12
3.2 Classification Accuracy . . . 13
3.3 Information Transfer Rate . . . 14

4 Discussion **17**

5 Conclusion **19**

1 Introduction

Brain-Computer interfaces (*BCI*) are a promising research area. *BCI*'s enable people to communicate with a computer via brain activity. They are potentially useful for patients suffering from amyotrophic lateral sclerosis [18], a neuronal disease that causes motor neurons to degenerate. In severe cases, communication is practically impossible since it relies on muscular activity, although the patients are still conscious. Brain-computer interfaces might allow them to have a rudimentary form of communication with the outside world.

There are two major kinds of *BCI*'s: *endogenous* are those *BCI*'s, that are self-initiated by a specific thought (e.g. in motor imagery [6], where patients imagine the movement of either left or right arm). In *exogenous BCI*'s in contrast, an external stimulus triggers a certain type of brain activity which can be modulated by the user's attention. This report focuses on the improvement of an exogenous *BCI*, the visual speller, that was invented by Farewell and Donchin [4]. They used a rectangular grid of letters rows and columns of which were highlighted successively. Synchronized with the stimulus presentation, EEG data was recorded. When the letter which the subject was thinking of was presented, a positive deflection in the EEG was detected 300 ms after stimulus presentation (P300). The P300 response is ususally elicited by an oddball paradigm, that is, when surprising changes at the target symbols occur suddenly. This event related potential was used to determine the time of the stimulus that resulted in the P300 and thus the letter the subject was thinking of could be retrieved. Theoretically, one could also think of other ERP's to be useful for classification. But since the P300 is modulated by attention, it is among the best studied ERP's for exogenous BCI's. Other ERP's seem to be associated more with stimulus deviance, task relevance evaluation or simply stimulus novelty. The scientific community coined terms like P300, N200, P200, novelty P3 or just "late positive complex" to describe those ERP's [13], [14], [5], [4] and to distinguish between their different functional aspects.

Concerning BCI's, it actually does not matter, which of those components are being used for classification. For a brain-computer interface, it is most important that there is at least one of those components that is robust enough to distinguish, with the help of a classifier, those stimulus epochs, where the symbol of interest has been presented to the subject, from those other epochs, where the symbol of interest was not contained in the stimulus.

The main aim of this work was to investigate which influence stimulus frequency and stimulus code length have on the signal to noise ratio of ERP's like the P300. To increase the information transfer rate, we tried to compare the influence of the stimulus type used to evoke the ERP's on accuracy and bitrate when the trials are either long (sparse stimulus code, low symbol frequency) or short (dense stimulus codes, high symbol frequency). The idea is that new stimulus types, that enhance the perceptual impact of stimulus events in the focus of attention but reduce the perceptual impact of stimulus changes outside the focus of attention, should also result in a relative enhancement of the ERP's that are elicited by target symbol changes. This enhancement of attended vs. unattended symbols could be advantagous when using dense codes, because distracting stimulus changes at non-target locations should have less effect not only on the perceptual but also on the neurophysiological level and thus on the EEG. In other words, using new types of stimuli, we hope to increase the ERP's of attended stimulus events relative to unattended ones.

Figure 1: sequence of one trial, consisting of 4 subtrials (repeated trials with the same target letter); the number of stimulus events within one subtrial was dependent on the stimulus code length

2 Methods

2.1 Stimulus preparation

We used four different stimulus types in combination with 3 different stimulus codes. All stimuli were superimposed on a four-by-three grid of numbers, like on a telephone keypad, see fig. 2. The stimulus type refers to the appereance of the symbols and the change in their appearance that will elicit the ERP's. The stimulus code refers to the number of stimuli needed until every letter on the grid underwent the stimulus type specific change at least once. The nomenclature for the experiments is the following: For each symbol, the speller performed one trial. Within this trial, the speller could use 1-4 repetitions for classification, so called subtrials. Taking more repetitions into account, the P300 signal is increased, taking less subtrials, the bitrate is increased since less time is needed for one symbol. So a trial consisted of 4 repetitions of the same target letter. An example of a single trial is depicted in fig. 1.

The general structure of a trial was the following: One trial comprised 4 subtrials, i.e. repetitions of the same target letter. The target letter on which the subjects had to focus on was indicated by a pink square superimposed on the target symbol shown before each trial. The subjects were instructed to concentrate on the target letter by counting the stimulus events, i.e. the changes occurring at that particular symbol, during the trial. This covert response was shown to be sufficient for the subjects to concentrate on the task relevant target symbols in such a way that a P300 is generated [4].

2.1.1 Sparse vs. dense codes

To find the best code we varied the stimulus code from a very dense to a single code and used the row-column code for comparison. The latter is the code used in [4], it consists of successive changes in all columns and rows of the visual grid. Since our grid was four-by-three, we needed 7 stimulus events to show every row and column changing at least once. So the code length of the row-column code was 7. It has been reported that the best bitrates can be obtained with a sparser code [3], where every symbol is changed only once in a subtrial thereby reducing the stimulus event frequency of each symbol. This complies with the finding that the ERP strengths are inversely

related to the frequency of the respective stimulus event [13], [14], [5]. The more often a stimulus target is shown the noisier the corresponding ERP's and thus the worse becomes classification accuracy.
When each symbol was changed within a subtrial only once, we need one stimulus event for every symbol, i.e. 12 symbols per subtrial. However, leaving aside the effect of stimulus frequency on the ERP, the code could be theoretically reduced, in our case to 4 stimulus events. This is because we have only 12 symbols to distinguish and we have (assuming perfect classification) 1 bit of information transferred after each stimulus, that is, we either have a P300 (or any ERP associated with the symbol of interest) or not. In an optimally dense code the lower bound for the code length L_{code} is:

$$L_{code} \geq log_2(Number\ of\ symbols) \tag{1}$$

So if we were to ask yes-no questions to come up with the right symbol, we would have to pose not more than 4 questions to uniquely determine which of the 12 symbols shown was the one that the subject was focussing on, since $log_2(12) = 3.585$. As the number of stimulus events has to be an integer, we have to round that value up and get 4. This would allow us to distinguish 16 symbols if we would use those 4 bits for a binary coding scheme, but we restricted ourselves to the 12 symbols on the telephone grid. A comparison of the different stimulus types is shown in table 2.1.1.

stimulus code	dense	row-column	single
$Length_{code}$	4	7	12
(in general)	$ceil(log_2(Number\ of\ symbols)$	rows + columns	Number of Symbols

Of course the bitrate is also dependent on the classification accuracy, thus if there occur disproportionately more errors with dense codes, the bitrates will not increase. In order to improve the signal to noise ratio of the ERP's and thus also the accuracy of the speller, we tried new stimulus types. This we hoped would allow us to get better accuracies that would allow us to use dense stimulus codes without sacrificing too much accuracy.

2.1.2 Classical Stimulus

To compare our results with other publications, we used the classical Farewell and Donchin [4] (FD) stimulus type as a control. In their paradigm, the letters were just successively highlighted, see fig. 2b). For all stimuli, we tried to use the same duration (250 ms) for one stimulus event, which is in the FD type experiment one highlighting of one or more letters. A stimulus event consisted of the change in the visual grid (at one symbol for the single code and at multiple symbols for denser codes) plus a resting phase during which nothing was changing in the stimulus. The interstimulus interval (ISI) was defined in [10] as the time from stimulus onset to next stimulus onset. The symbol was highlighted for 150 ms followed by a 100 ms rest phase, this gives an ISI of 250. The ISI hence is synonymous for stimulus event-duration. The reason why we chose such a short interstimulus interval is that if we want to get higher bitrates, we have to reduce the time needed for one stimulus event. Implicitly this can be done by reducing the resting phase. However, we have to assure, that this phase is not too short for the P300 to develop. We chose the ISI duration according to recent findings by Kaper and Ritter [9], [10], who investigated how the ISI length affects accuracy. Their results show, that even if the ISI was significantly shorter than 600 ms (the interval, in which the P300 has fully developed and decayed), classification

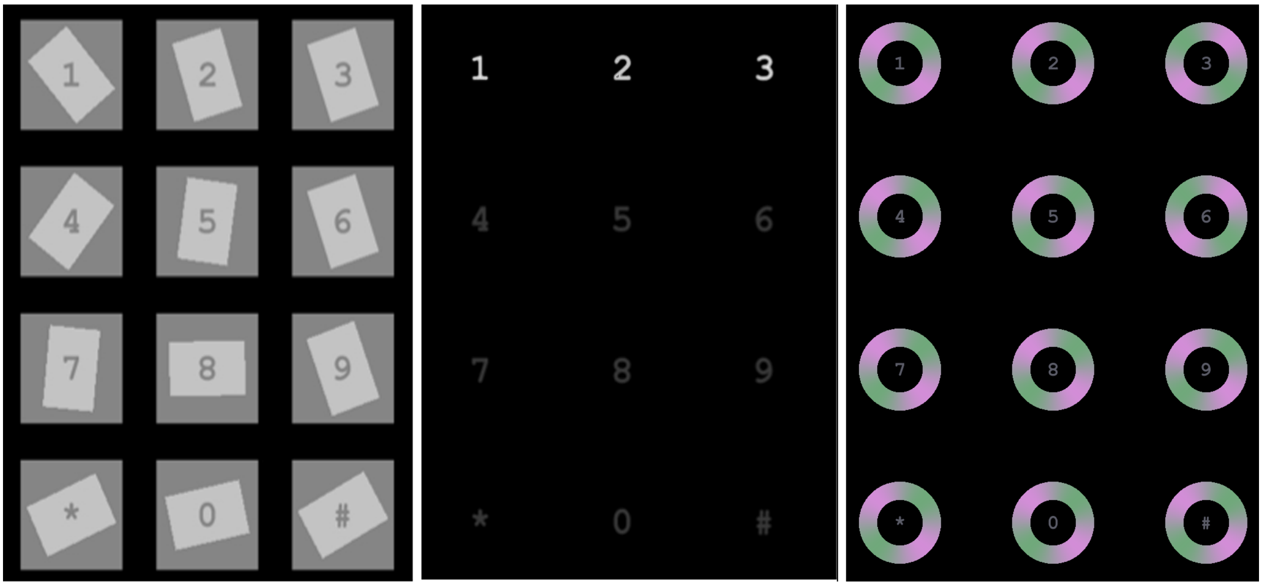

Figure 2: *left*: Change blindness stimulus (CB); between a full flip of the rectangles, a grey screen was flashed on and off; control stimulus (Change-non-blindness, CnB) was the same but without the grey flash; *middle*: Stimulus type taken from the classical Farewell and Donchin P300 speller (FD); rows and columns were successively highlighted; *right*: Isoluminant color stimulus (ICM), the circles around the symbols were coloured with faint (low luminance, low colour contrast) gratings from pink to green; during a stimulus event, the colour swapped from green to pink and vice versa, that is a standing wave was shown

accuracies of 80% can be achieved. For some subjects they report nearly perfect accuracies even with ISI's of 150 ms. That is why we chose an ISI of 250 ms for the FD stimulus type. For technical reasons (rounding inaccuracies and framerate of the monitor) the exact stimulus event duration was 257 ms for the FD type.

2.1.3 Change Blindness

One of the new stimulus types we used was inspired by the phenomenon of *change blindness* (see e.g. [20],[19],[15]). When you present in short sequence two pictures that are identical except for one feature, the change in the image will be seen immediately. However, when there is in between the presentation of the two images a short interval with zero contrast, i.e. a grey screen, the change will not be detected that easily. After scanning the whole picture, observers usually find the change in spite of this effect, and after having it noticed once, they will detect it immediately again. But the attention is not drawn to the change as in the condition without the grey screen in between. This property of *change blindness*, namely that our attention does not automatically jump to the location where there is some change happening in the visual scene, we wanted to use to increase the signal to noise ratio of the visual speller. What we consider as *signal* is the P300 (elicited by the symbol of interest) whereas noise is basically every signal that is not correlated with the symbol of interest. By showing the intermediate grey screen, the observers should be made insensitive for changes outside of their focus of attention. Our hypothesis is that this might enhance the P300 evoked by the "correct" stimulus. To be able to show a grey flash in between a stimulus, we used not the FD type stimulus but a white rectangle that flipped its orientation by 90° (fig. 2, left). Between the two frames the grey screen was shown. The control condition for the change blindness (CB) stimulus was the same but the grey screen during the flip was omitted. We will refer to this condition as change-non-blindness (CnB). The grey flash presentation was set to 100 ms. Before the flash, 50 ms of the flip were shown and 50 ms after the grey flash, the rest phase was set to 50 ms. This sums up to 250 ms in total. Due to technical limitations the effective stimulus event duration was 214 ms for both the CB and the CnB condition. During the latter, the 100 ms of the flash were added to the inter-stimulus interval.

2.1.4 Isoluminant Color Motion

We also explored another stimulus based on color motion as investigated by Seiffert and Cavanagh [16]. They reported two types of motion perception. One, referred to as first-order motion perception, was more sensitive to target velocity, whereas the other, second-order motion perception, was more sensitive to target position. While first-order motion in their experiments was based on luminance contrast and local motion detection, second-order motion was elicited by isoluminant color contrasts and required solving the correspondence problem. Since the latter motion percept is based on position tracking and not low-level luminance contrast analysis, we concluded that second-order motion needs attention at the site of interest. Low-level luminance contrast detection is already performed at early stages of the visual processing hierarchy, like retinal ganglion cells. Therefore, distracting stimuli are present at all stages of the hierarchy. On the other hand, attention driven detection like second-order motion should be less susceptible to distracting stimuli. Hence, as in the case of change blindness, the stimulus of interest should be less affected by noise, i.e. changes happening at other locations than the focus of attention. Since the luminance output of each colour Lum_{col} is not a linear function of the voltage V applied to each pixel, we had to calibrate the monitor in order to have isoluminant colours. The three colours were linearized by fitting a first order exponential function to the red, green and blue pixels of the monitor:

$$Lum_{col} = \alpha + \beta V^{\gamma} \tag{2}$$

We conducted a short series of psychophysical measurements before the actual experiment during which we determined the coefficients of this function at varying luminance levels. Subjects were instruced to assimilate the luminance of two halves of a colored square. We fitted the function yielding Lum_{col} of the respective colour using those coefficients. To maximize the effect of second order motion, we showed a grating of isoluminant colours that drifting from left to right at about 1° above a fixation cross (which is about the same distance from the symbol on the visual grid to the actual ICM stimulus, see fig. 2). Since the isoluminant colour motion effect is most prominent at low colour contrast, we tried to make the contrast as low as possible without losing the motion percept. While the subject fixated at 1° below the moving grating, we varied the colour contrast, until it was perceived somewhat slower. This was taken as an indication for a second-order motion percept [16]. The contrast levels used were *red*=0.588-0.843, *green*=0.843-0.588 and *green*=0.588-0.843. Colors were varied between these luminance levels in a circle around the symbols of the visual grid. The resulting percept was a standing wave that changed its colour from a faint pink to green. Stimulus frequency and colour contrast have an effect on the motion percept [17]. With increase in temporal frequency or colour contrast, the motion percept becomes more and more distorted by interference with other motion detection mechanisms, thus we kept both parameters low. The colour contrast was fixed at a low level as described earlier. The temporal frequency was set to 2 Hz, which is the value where luminance contrast motion is least interfering with second-order motion [17]. Stimulus event duration thus was set to 350 ms plus 150 ms rest phase and the spatial frequency was set to 2 periods per cycle.

2.2 EEG recordings

Recordings were conducted with a QuickAmp amplifier manufactured by Brain-Vision. The software for synchronizing the EEG with the stimuli was custom made. We used for four of the 10 subjects the whole set of the 128 electrodes. For the last 6 subjects we used only a subset (76) of the electrodes. We made sure that impedance was kept below 5 $K\Omega$ before starting the session. For some channels the impedance could not be brought underneath the threshold. Those were removed from the recordings afterwards. Apart from the electrodes mounted on the cap, we used one electrode to record electro-occulograms to be able to tell, wether the EEG signal was distorted by eye movements. Subjects were sitting in a normal office rather than a real psychophysical setup, since we wanted to ensure our analytical methods were robust to everyday levels of noise and disturbance.

2.3 Analysis

The analysis was done in Matlab. The framework for the code I used was provided by my supervisor. For the preprocessing, different configurations were tested, the final ones were the following.

2.3.1 Preprocessing

After removing the high impedance channels, the EEG data was separated into its independent components (*IC's*) using the FastICA algorithm [7]. Trials that were corrupted by artefacts were detected by visual inspection. Those trials have been removed. Criteria for removal were either dominant activity in the occulogram and components in the frontal regions or just noise spread across electrodes. Then, linear trends were subtracted from the data by fitting a first order polynomial to each component and subtracting it. The next step was to bandpass the signal within the band of 0.1-30 Hz and downsample it according to the Nyquist frequency of the bandpass band. The filter cutoffs were set as hamming windows rising from 0.1 Hz to 1 and decaying from 25 to 30 Hz. Finally the trials were split in windows of 600 ms after stimulus event onset.

2.3.2 Classification

For classification we used linear support vector machines (SVM's) [22]. Given a training data set consisting of data points $\mathbf{x} \in R^N$, where R^N is the input space (in our case: N is the number of channels $\times$ number of timesamples) and corresponding labels $y \in \{-1;1\}$, where -1 refers to EEG activity during stimuli that did not contain changes at the target symbol and 1 to EEG patterns occurring during stimulus events that contained changes at the target symbol. Our classifier should ideally find a mapping f($\mathbf{x}^*$) from the real numbered test data $\mathbf{x}^* \in R^N$ onto the set of (unkown) testlabels y^*, where * indicates that the datum is taken from the test and not the training set.

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b \tag{3}$$

$$f(\mathbf{x}) = \begin{cases} > 0: & \text{stimulus changed at target symbol} \\ < 0: & \text{stimulus changed elsewhere} \end{cases} \tag{4}$$

The weight vector $\mathbf{w}$ separates the negative from the positive data points. The distance between the weight hyperplane and the data points nearest to it is called the *margin* of the classifier. The size of the margin is often considered as an indicator of a classifiers generalization performance. That is, the larger the margin, the better the classifier should generalize the learned discriminatory capabilities to new, unseen test data. In a canonical representation, maximizing the margin is equivalent to minimizing the L_2-norm of the weight vector. While minimizing, the constraint $y(\mathbf{w} \cdot \mathbf{x} + b) \geq 1$ has to hold. When this constraint is violated because the data is not separable or too many datapoints lie within the margin, it is relaxed by introducing so called slack variables ξ_i each representing the error of the corresponding training datum $\mathbf{x}_i$. Those ξ_i's are also minimized while they are penalised during optimization by the regularization parameter C. So the goal of the svm training is to minimize the norm of the weight hyperplane $\frac{1}{2}||\mathbf{w}||$ to get a large margin and hence good generalization performance. At the same time we want to reduce the number of margin errors ξ_i, which is penalised by C:

$$\begin{aligned} min \quad & \frac{1}{2}||\mathbf{w}|| + C\sum_i \xi_i \\ s.t. \quad & y_i(\mathbf{w} \cdot \mathbf{x} + b) \geq 1 - \xi_i, \xi_i > 0 \forall i \end{aligned} \tag{5}$$

Usually this problem is solved in the dual representation, so the goal is to find the α_i that maximize the Lagrangian:

$$\begin{aligned} max \quad & \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \\ s.t. \quad & 0 \leq \alpha_i \leq C \forall \text{ i} \\ and \quad & \sum_{i=1}^{m} \alpha_i y_i = 0 \end{aligned} \tag{6}$$

The α_i that are nonzero correspond to those datapoints that are called support vectors of the separating hyperplane. Note that in the above formulation, the explicit representation of the weight vector is substituted by the inner products of data points. By solving this problem we can compute the weights $\mathbf{w} = \sum_i y_i \alpha_i \mathbf{x}_i$. Thus we can reformulate the separating function as:

$$f(\mathbf{x}) = \sum_{i=1}^{m} y_i \alpha_i \cdot (\mathbf{x}_i, \mathbf{x}_j) + b \tag{7}$$

We trained the SVM's in a leave-one-out fashion, that is, for the offline prediction of a symbol in one trial all other trials were used as training data. Within the training fold, we performed a 10-fold crossvalidation over a range of C-parameters of the SVM. In each fold, we balanced the training examples by discarding randomly the surplus of negative training examples from the training set. The 10 regularization hyperparameters chosen first were between $10^{-6} - 10^{3}$ on a logarithmic scale. We also tried a non-linear, Gaussian kernel with varying widths. Since this did not significantly improve the results, we present the results of the linear svm. After having trained the svm, we multiplied the raw svm outputs $\mathbf{o}$ with the stimulus code $\mathbf{F}$. The code consisted of s rows and e columns, where s is the number of symbols on the visual grid and e the number of stimulus events. Each entry of $\mathbf{F}_{se}$ was 1 if the sth symbol was changed in the eth stimulus event and -1 otherwise. For each column we computed one svm output. The svm outputs were subtracted from the flipgrid and the columns were summed up. The smallest value was then the symbol most similar to the one presented and was chosen by the the speller. So

$$S_p = argmin_s(\sum_e |F_{se} - o_e|) \tag{8}$$

Where S_p is the symbol predicted, F_{se} is the sth symbol in the eth stimulus event and o_e is the eth svm output value. So F_{se} where s is 1,...,s are all symbols at stimulus event e, i.e. it is a vector whereas o_e is just a scalar svm output.

2.4 Performance measures

To assess how good the visual speller performs, we compared the classification accuracy and the bitrates of our *BCI* with other publications. The accuracy is just the proportion of correctly predicted symbols. For the sake of comparability, we used the definition of [9] to compute the bitrates of information transfer:

$$BR = \frac{60}{t}(log_2 s + p_c log_2 p_c + (1 - p_c) log_2 \frac{(1 - p_c)}{(s - 1)}) \tag{9}$$

Where t is the time needed for one trial, s is the number of symbols on the grid and p_c is the probability of a correct trial, i.e. that the symbol predicted by the speller was the same as the subject was (or rather should be) focussing on. It has been argued that the definition given in [9], which is adopted from [8], was underestimating the true bitrate if the number of symbols on the grid is larger than 5 (see [11]). Kronegg et al. claim, that the definition given by Nykopp [12], yielded more reliable bitrates. Nykopp computed the bitrates as the

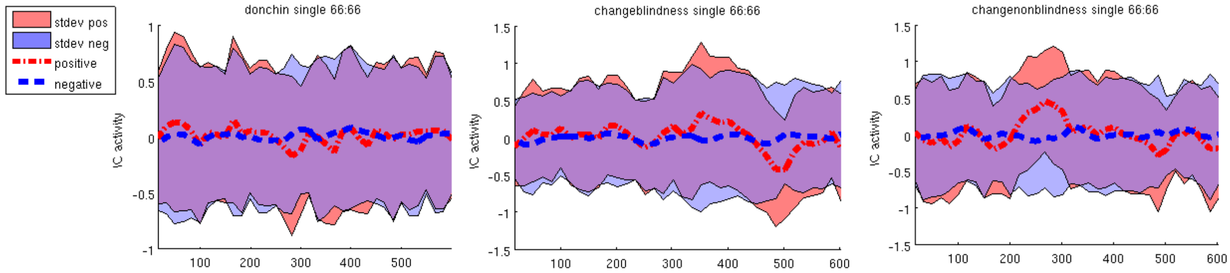

Figure 3: Grand average of FD (*left*), change blindness (*middle*) and change non-blindness (*right*) stimulus types; trials were taken from the same code condition (single) and independent components corresponding to occipito-parietal regions of the left hemisphere; positive trials (that contained the stimulus of interest) are plotted red, negative trials are plotted blue; standard deviations are superimposed in respective color, overlapping regions are purple; y-axis is dimensionless activity of the independent components

channel capacity based on the mutual information of input and output symbols [2]. However when using this definition, our results were not directly comparable with other publications (we achieved bitrates of more than 180 bits/min), so we decided to show only the bitrates according the Wolpaw et al. definition.

3 Results

3.1 Averaged ERP's

To examine the efficacy of the stimulus types used, we had a look at the averaged EEG trials and the differences in brain activity comparing positive trials and negative ones. The source of P300 signal has been reported to be located in a centro-parietal position [13]. To find differences in brain activity corresponding to the different stimulus types, we picked subsets of independent components of the EEG data separated by ICA. We chose independent components, that were useful for distinguishing positive trials from negative ones. The criteria for our choice were the following: we looked at the weights chosen by the support vector machines; the largest absolute values indicated the importance of the corresponding features. Another criterium were the receiver-operator characteristics (ROC) of each feature. The ROC was evaluated by looking at the area under the ROC curve of each spatio-temporal feature. For the luminance contrast conditions we found decisive features in the occipito-parietal region, while for the isoluminant color motion we took the independent components originating in the central region of the skull (Cz). Although we did find some discriminative features in the central regions for the luminance contrast motion paradigms, too, the features in the occipito-parietal areas were more decisive for the subject shown in 3. Moreover there was a considerable intersubject variability concerning the most decisive features. The comparisons of the different stimulus conditions are shown in fig. 3. The dashed lines show the means of positive (*red*) and negative (*blue*) trials, the area around the lines shaded in the respective color is one standard deviation, overlapping areas are purple.

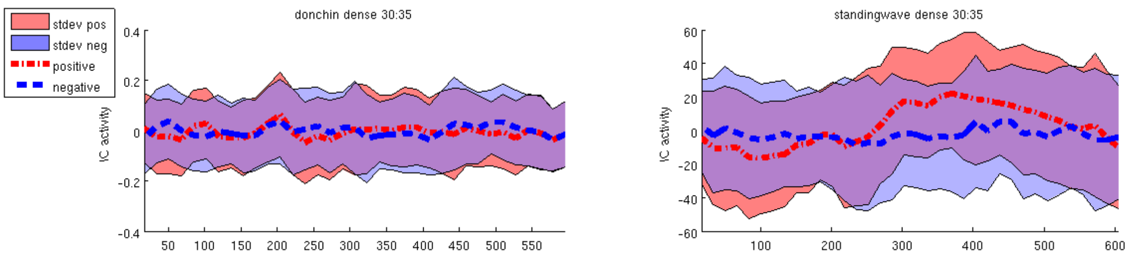

Figure 4: *left:* Average over all trials of the FD type and *right:* isoluminant color condition, stimulus code was dense; positive trials are plotted blue, negative trials green; note the remarkable difference in the scale, the isoluminant color motion (ICM) resulted in a peak activity in the plotted IC's that was about 600 times stronger than that of the FD type luminance color motion; origin of the component was around Cz

The isoluminant color condition is not directly comparable to the other three conditions, because the stimulus events had twice the duration. Another difference was that the stimulus evoked aftereffects, with which not all subjects felt comfortable. So some seemed to be rather irritated by the coloured illusions between the rather fast stimulus presentations. Still for some subjects the isoluminant color condition yielded very good results in the classification performed later on. The example shown in 4 compares FD type stimulus and the ICM stimulus. Plotted are the average over all trials in the isoluminant colour condition using a dense stimulus code at the independent components originating around electrode position Cz. The difference in means in the ICM example is large, when compared to the signal to noise ratio of the FD stimulus type. The P300 peak and duration differences are clearly visible in the ICM condition. Preceding the positive deflection of the EEG signal there is a distinct negative peak around 100 ms after the stimulus offset.

3.2 Classification Accuracy

To compare the influence of the new stimulus types and codes used on classification accuracy we examined the bitrates of the visual speller in each of the 12 experimental conditions. For some subjects the performance was very good. But for others, the bitrates and accuracy were not comparable to the results in recent publications, c.f. [9]. Some possible reasons for this are presented in the discussion part. The mean accuracy was rather poor, especially in all 3 stimulus code conditions of the classical FD type stimulus (see. fig 5 and 6).
Comparing the codes, the general assumption on ERP's elicited by an oddball paradigm, namely that stimulus frequency is inversely related to the quality of the ERP, seems to hold. In the case of the dense code, the accuracy was never perfect . This was not the case for the row column code. Still, the assumption that the longer the stimulus code, that is the more trials for prediction are used, the better the accuracy, does not hold for all stimulus types. The mean accuracy had its maximum not always in the single code condition. The best stimulus type in terms of classification accuracy was the CnB condition, throughout all stimulus code conditions. This was surprising, because we actually just used this stimulus type to have a control condition for the CB stimulus. Allthogh worse than its

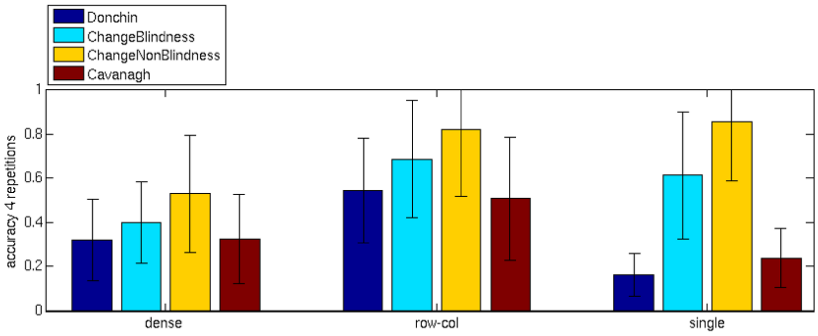

Figure 5: Mean classification accuracy ± one standard deviation as function of stimulus code length (4 subtrials used for classification); not for all stimulus types accuracy increases with number of stimulus presentations; interestingly, the single code condition resulted only in the CnB condition in the maximal mean accuracy

control, the CB condition was still better than the classical FD type stimulus. A comparison between the different stimulus types is plotted in fig. 6. The blue line indicates equal performance in terms of accuracy achieved with the respective stimulus type. The stimuli using luminance contrast motion (CB, CnB) outperform those that use luminance contrast flicker (FD) consistently. This can be seen in the plots in fig. 6: most subjects have higher accuracies in the former two than latter two paradigms and fall below the iso-accuracy line. In some cases the CnB condition needed only one trial to predict more than 90% of the digits correctly. Depending on the stimulus code, the number of stimulus events was 4 (dense) , 7 (RC) and 12 (single code). So assuming 250 ms for the duration of a single event, one trial was either L_{dense}=1 sec, L_{RC}=1.5 sec or L_{single}=3 sec. The exact durations, deviate a little from these values. For the isoluminant color condition, the trial took double the time.

3.3 Information Transfer Rate

Although prediction accuracy suffered from target frequency: even if the accuracy in the dense code condition and the RC code was lower than in the single code, it was still high enough to result in higher bitrates than the single condition. The increase in accuracy with a single code was not high enough to increase the bitrate in this condition significantly. This tendency was consistent through all conditions. The only remarkable difference between the classical FD type stimulus and the new stimuli we used was the fact that the bitrate of the former dropped from the dense to RC condition and was nearly the same for the RC and single condition, while the latter conditions (CB, CnB, isoluminant color motion) had similar bitrates in the dense and RC condition and lower bitrates in the single code condition (see fig. 7). For the CnB condition, we could achieve good bitrates, comparable to results reported in recent publications, see tables 1, 2 and [23], [24], [3], [9].

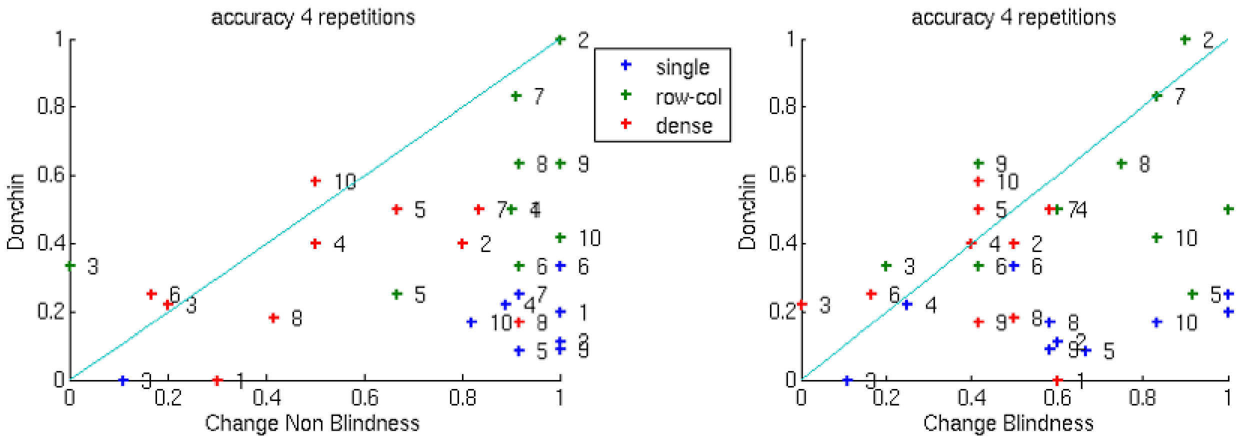

Figure 6: Comparison of classification accuracy (1 subtrial used for classification): the straight line represents equal performance with both stimulus types; subjects below the line were better in the stimulus conditions corresponding to the x-axis; *left:* x-axis shows performance in CnB condition, y-axis the performance in the classical FD condition *right:* y-axis shows performance in CB condition, y-axis again the FD condition; note that the new stimulus types using luminance contrast motion rather than flicker result in better accuracies;

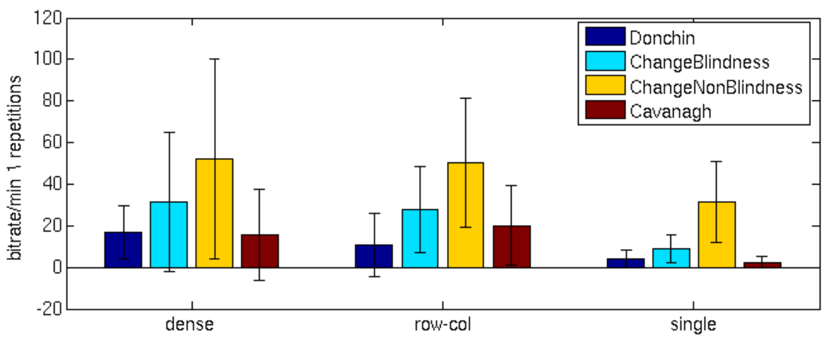

Figure 7: Mean bitrates ± one standard deviation as function of stimulus events per subtrial (1 repetition); note that all stimulus types had higher bitrates in the dense code condition

	dense code	RC code	single code
FD	16.54±12.8	10.83±15.3	3.86±4.2
CB	31.37±33.3	27.94±20.6	8.90±6.7
CnB	52.35±48.0	50.55±31.0	31.37±19.3
ICM	15.49±21.8	20.16±19.3	2.47±2.9

Table 1: Mean bitrates ± standard deviation of all 12 experimental conditions using 1 subtrial per trial; rows indicate the stimulus type, columns the stimulus code;

	dense code	RC code	single code
FD	6.57±6.2	7.13±6.1	0.44±0.6
CB	9.68±6.1	12.67±9.5	4.24±4.0
CnB	17.48±15.2	13.41±12.0	6.81±7.2
ICM	7.07±7.1	6.60±6.8	1.10±1.4

Table 2: Mean bitrates ± standard deviation of all 12 experimental conditions using all 4 subtrials available per trials; rows indicate the stimulus type, columns the stimulus code;

4 Discussion

Interestingly, the control condition of the change blindness paradigm yielded significantly better results than the change blindness condition itself (fig. 5 middle and right). This was counterintuitive considering that the CB stimulus was constructed such that subjects did not perceive changes going on at locations outside of their focus of attention. There are two stimulus properties, that could potentially explain that. Either the change blindness condition produces too much sensory noise by the continously flashing screen, or the resting phase between the stimulus events was too short in the CB condition, compared to the CnB condition. The phenomenological aspects of the first argument were also reported by the subjects themselves. The CB stimulus was the one most subjects felt least comfortable with. This was probably due to the grey flash in between the stimulus events. In some subjects (data not shown) we can see visually evoked potentials in the averaged EEG, oscillating phase locked with the frequency of the grey flashes. The other argument was that the resting phases were too short in the CB condition. Kaper and Ritter investigated the effect of the ISI on the classification accuracy and found that decreasing the ISI below the 600 ms border did surprisingly not affect the classification accuracy in some subjects [9]. Considering the latency and decay time of a typical P300 signal (it only decays about 600 ms after stimulus onset), one would think, that the P300 responses would be distorted by the previous one. Even when decreasing the ISI down to 150 ms (the resting phase was not mentioned, so it was most probably zero), some subjects achieved nearly perfect accuracy after only 4 trials. Appearently reducing the ISI below the critical value of 600 ms did not decrease neither the quality of the ERP's nor the classification results significantly. This encouraged us to reduce the ISI to 250 ms. This alone should not have affected our results too much. Allthough the ISI did not change across stimulus conditions, the resting phase did. This phase might possibly have had an influence on the quality of the P300 or other signals, that could have been used for classification. Unfortunately we

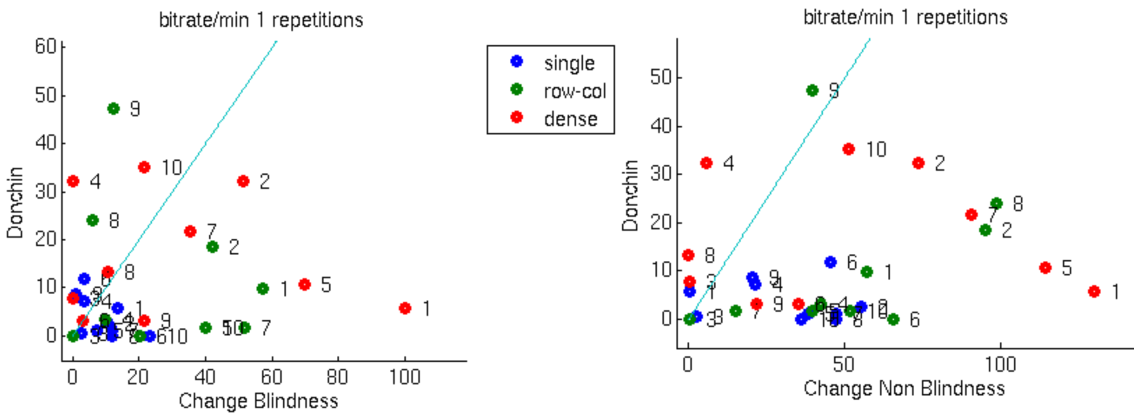

Figure 8: Bitrates of single subjects in bits/min using one subtrial for prediction; for many subjects, the speller yielded very poor results; however, there was a tendency for higher bitrates in the dense code conditions; the blue line with slope 1 indicated the performance if the experimental conditions would result in the same performance; the new luminance motion stimuli outperform the FD type, as most of the subjects fall underneath the line of equal performance

cannot directly compare our stimulus events with those reported in [9]. Since we sacrificed in the CB condition some time from the ISI to the grey flash needed for the change blindness effect, the resting phase lasted only 50 ms. This might have been too short. The resting phase could be a parameter that affects the qualitiy of those signals that are important for classification (P300, N200).
Leaving aside the unexpected negative effect of the CB stimulation, both stimuli can achieve accuracies and bitrates comparable to results in recent publications. Some subjects needed only one second (in the dense code condition) to transmit the correct symbol. In those subjects that had a classification accuracy above 80%, we looked at the average EEG and found P300 activity in most conditions. An example is shown in fig. 3b). However, intersubject variability was very large, in the classification results as well as in the features used for classification.
Surprisingly the CB/CnB stimuli outperformed the classical FD stimulus consistently. The superiority of the CnB/CB stimulus type compared to the classical luminance flicker paradigm of the FD speller cannot be generalized but must be rather due to our specific setup. The bitrates and accuracies achieved with the CB/CnB stimuli are comparable to other publications, however the FD speller results are not. The main difference between our stimuli and those of other publications that could have led to this outcome is the size of the symbols on the grid. In our stimuli, they were comparatively small ($\approx 1.5°$ of visual angle). Another speculation, allthough not explaining the huge difference compared to state-of-the-art visual spellers working with the FD speller paradigm as in [9], is that the difference could be due to the motion energy contained in the new luminance contrast motion stimuli. Assuming a low level motion detection mechanism consisting of spatio-temporal receptive fields (e.g. [1], [21]) reflected in the EEG, the difference in activity comparing the flicker and motion stimuli would correspond to the differences in motion energy: motion contains more motion energy than flicker stimuli. This explanation can also partially account for the surprisingly low accuracy of the FD stimulus types compared with other publica-

tions: changes happening at large symbols should result in more motion energy than changes happening at smaller ones like the ones we used.
Although the classification performance was better in the CnB condition than in the CB condition, both conditions resulted in similar relations between bitrate and target frequency: the bitrate decreased with number of stimulus presentations. The more stimulus events per trial, the better was the signal to noise ratio, but the worse became the bitrate. The former statement was in perfect agreement with recent results in the literature, although the effect was not as strong in our data as in other publications. The latter however was not. One important aspect concerning the bitrates is, that there are only few publications that tried to compare dense codes with single or in general sparser codes. In particular, codes denser than the classical row-column code, have been rather neglected in the literature. Moreover, the direct comparison of row-column codes and sparse codes is not always possible, because the different codes have not been presented under the exact same conditions. For instance Guan et al. varied the ISI's such that each trial had the same duration in the two code conditions they tested [3]. The two stimulus codes could not be directly compared in those experiments, because the experimental conditions changed with the code. In addition to that the scientific community has not yet agreed on a formula to compute the bitrates of a BCI (see [11]), so classification accuracy is still the most common benchmark. This is sensible from the patients point of view, because for them, incorrectly spelled symbols are much worse than being a few seconds slower or not. Our results show, that sparser codes indeed improve accuracy but not to an extent that denser codes could not be considered as feasable alternative.
In other words, even if the dense code condition was not as accurate as the RC condition (that holds true for the single code condition, too, see fig. 5) and more errors occurred, the bitrates were still higher in that condition.

5 Conclusion

The assumption *"the lower the target stimulus frequency, the better the accuracy"* was confirmed for two stimulus types (CB/CnB) and for the other two at least partially. However, when using dense stimulus codes, the signal to noise ratio of the relevant ERP's (fig. 5, 6, 3) is at least good enough to yield higher bitrates compared with sparser code conditions, when the faster presentation time is taken into account.
Of course even if the bitrates are higher in the dense code condition, our hypothesis that enhancing the *perceptual* impact of a target stimulus (relative to non-targets) does automatically increase the signal to noise ratio of the associated ERP was not supported. In particular the control condition (CnB) was better compared to the CB condition. We attributed this superiority mainly to the considerably longer resting phase in the CnB condition and the low signal to noise ratio in the CB condition due to the continously flashing grey screen. Another discouragement was the low mean accuracy and bitrates compared to other publications under the FD condition. This could also be explained by the short resting phases and, in addition to that, the small size of the symbols used. Reducing the resting phase or in general the ISI is a very efficient means to de-

crease the trial duration and thus also the total bitrate. On the other hand we found that there is a limit value for the resting phase: when reducing the resting phase down to 50 ms as in the case of the CB condition, we achieved only very poor results (5). Future experiments will have to confirm whether and to what an extent these variables, the resting phase or ISI, influence the quality of the EEG signal.
To use the new stimuli and especially dense codes in an online BCI, we would have to improve the classification accuracy of our speller. Given that we could use for a 36 letter grid instead of 36 stimulus events (in the commonly assumed best single code condition) only $log_2(36) \approx 6$ stimulus events (using a dense code in combination with one of our new stimulus types), the increase in bitrate for a alphanumeric grid would be even higher. In addition to that, we could incorporate some more stimulus events and would have enough codewords left when using a dense code to incorporate some error correction. Additionally, we could come up with a more intelligent design of the flipgrid. Symbols that are very often used in spellings or that are very often cooccurring should be placed far apart from each other to decrease distractive stimulus events in the focus of attention. This is the goal of upcoming experiments.

References

[1] Bergen JR Adelson EH. Spatiotemporal energy models for the perception of motion. *J Opt Soc Am*, 2(2):284–99, 1985. 18

[2] Thomas JA Cover, TM. Elements of information theory. *John Wiley and Sons, New York*, 1991. 12

[3] M. Jiankang Wu. Cuntai Guan, Thulasidas. High performance p300 speller for brain-computer interface. *Biomedical Circuits and Systems, 2004 IEEE International Workshop on*, 2004. 5, 15, 18

[4] 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*, 70(S2):510523, 1988. 4, 5, 6

[5] A Goldstein, K Spencer, and E Donchin. The influence of stimulus deviance and noveltz on the p300 and novelty p3. *Psychophysiology*, 39:781–790, 2002. 4, 5

[6] NJ Hill, TN Lal, M Schröder, T Hinterberger, B Wilhelm, F Nijboer, U Mochty, G Widman, C Elger, B Schölkopf, A Kübler, and N Birbaumer. Classifying eeg and ecog signals without subject training for fast bci implementation: comparison of nonparalyzed and completely paralyzed subjects. *IEEE Trans Neural Syst Rehabil Eng*, 14(2):183–6, 2006. 4

[7] Jaakko Särelä Hugo Gävert, Jarmo Hurri and Aapo Hyvärinen. Fastica for matlab 7.x, version 2.5,. 2005. 9

[8] D. J. McFarland J. R. Wolpaw, H. Ramoser and G. Pfurtscheller. Eeg-based communication: improved accuracy by response verification. *IEEE Trans. on Rehab. Eng.*, 6(3):326–333, 1998. 12

[9] M Kaper, P Meinicke, U Grossekathoefer, T Lingner, and H Ritter. Bci competition 2003–data set iib: support vector machines for the p300 speller paradigm. *IEEE Trans Biomed Eng*, 51(6):1073–6, 2004. 6, 11, 12, 13, 15, 17, 18

[10] M Kaper and H Ritter. Progress in p300-based brain-computer interfacing. *Biomedical Circuits and Systems, 2004 IEEE International Workshop on*, 2004. 6

[11] Voloshynovskiy S. Pun T. Kronegg, J. Analysis of bit-rate definitions for brain-computer interfaces. *Int. Conf. on Human-computer Interaction (HCI'05), Las Vegas, Nevada, USA*, 2005. 12, 18

[12] T Nykopp. Statistical modelling issues for the adaptive brain interface. *M.Sc. Thesis*, 2001. 12

[13] Azzam PN. Patel SH. Characterization of n200 and p300: selected studies of the event-related potential. *Int J Med Sci.*, 2(4):147–54, 2005. 4, 5, 12

[14] GF. Potts. An erp index of task relevance evaluation of visual stimuli. *Brain and Cognition.*, (56):5–13, 2004. 4, 5

[15] RA Rensink. Change detection. *Annu Rev Psychol*, 53((null)):245–77, 2002. 7

[16] Cavanagh P. Seiffert AE. Position displacement, not velocity, is the cue to motion detection of second-order stimuli. *Vision Res.*, 38(22):3569–82, 1998. 8, 9

[17] Cavanagh P. Seiffert AE. Position-based motion perception for color and texture stimuli: effects of contrast and speed. *Vision Res.*, 39(25):4172–85, 1999. 9

[18] EW Sellers and E Donchin. A p300-based brain-computer interface: initial tests by als patients. *Clin Neurophysiol*, 117(3):538–48, 2006. 4

[19] ME Silverman and A Mack. Change blindness and priming: When it does and does not occur. *Conscious Cogn*, 2005. 7

[20] DJ Simons and RA Rensink. Change blindness: past, present, and future. *Trends Cogn Sci*, 9(1):16–20, 2005. 7

[21] Sperling G van Santen JP. Elaborated reichardt detectors. *J Opt Soc Am*, 2(2):300–21, 1985. 18

[22] V.N. Vapnik. Statistical learning theory. *John Wiley and Sons, New York*, 1998. 10

[23] JR Wolpaw, N Birbaumer, DJ McFarland, G Pfurtscheller, and TM Vaughan. Brain-computer interfaces for communication and control. *Clin Neurophysiol*, 113(6):767–91, 2002. 15

[24] N Xu, X Gao, B Hong, X Miao, S Gao, and F Yang. Bci competition 2003–data set iib: enhancing p300 wave detection using ica-based subspace projections for bci applications. *IEEE Trans Biomed Eng*, 51(6):1067–72, 2004. 15