A gaze independent spelling based on rapid serial visual presentation

Laura Acqualagna and Benjamin Blankertz

Abstract—**An event-related potential (ERP) speller is a brain computer interface (BCI) based on the detection on ERPs that can be used as spelling device for those people deprived of other means of communication. In the present online study we investigated in twelve participants the performance of an ERP speller based on the rapid serial visual presentation paradigm (RSVP). Three variants of the RSVP speller have been investigated regarding chromaticism and speed of stimulus presentation. All the subjects were able to successfully operate the RSVP speller and high mean symbol selection accuracies were reached in all conditions, (93.6% to 94.8%). Offline analysis revealed a possible mean spelling speed of about 2 symb/min for an optimized number of stimulus sequences. The RSVP speller is intuitive to use and it is gaze independent, which makes it suitable for patients with deterioration of oculomotor control.**

I. Introduction

Brain Computer Interfaces (BCIs) establish a direct communication pathway between the brain and an external device. They are a realistic perspective to develop communication devices for patients suffering from, e.g., tetraplegia or Amyotrophic Lateral Sclerosis (ALS). A category of brain signals widely used in BCIs is the event-related potential (ERP). ERPs are time and phase locked to an internal or external event. Attention to an event can enhance the positive and negative peaks of ERP components. An ERP-based BCI can detect these modulations and use them to drive an external device. In spelling devices, users can select symbols by allocating attention to stimuli that are presented on a computer screen. The first ERP speller was developed by Farwell and Donchin [1]. This device is known as Matrix speller, since the symbols are arranged in rows and columns. Recently, two independent studies showed that the Matrix Speller has a high classification accuracy only when the users overtly fixate the target symbol [2,3]. This result constitutes the need to develop gaze independent spellers, since oculomotor control can deteriorate in late stages of diseases like ALS.

One such approach is to investigate non-visual ERP BCIs, using a paradigm based either on auditory [4] or on tactile [5] stimuli.

Another approach is to use visual stimuli which let the users drive the speller using the covert attention. An example is the Hex-o-Spell [2], where the symbols are arranged in six groups on the vertices of an invisible hexagon. The symbol selection is a two-stages process. This design overcomes some limitation of the Matrix used in the covert attention condition and leads to an increment of the classification performance. Recently, the authors have investigated three different variants of the two-stage visual speller, based on covert attention and non-spatial feature attention [6]. The stimulus onset asynchrony (SOA) was 200 ms. In an online experiment, they reach a mean symbol selection accuracy of 88%-90%-97% for the different spellers. Another study [7] presents a BCI speller in which the characters are presented sequentially as clusters in a central location with an SOA of 400 ms. Participants search and recognize the target letter with covert shifts of attention while fixating the center. The symbol selection accuracies in online spelling amounted to 96.3%, respectively presenting the target either in a random or in a fixed position. The authors derive a spelling-speed of 1.38 symbols/min for the optimal number of stimulus repetitions from offline cross-validation using the theoretical information transfer rate, which is a positively biased measure as it assumes optimal error correction.

In the present study we investigate a BCI-speller based the rapid serial visual presentation (RSVP) paradigm, in an online experiment. Our previous study [8] showed that the RSVP is a promising new BCI-paradigm. In that study, we investigated two different SOAs (83 ms and 133 ms), and two different color conditions: "no-color", where all the symbols were black, and "color", were the 30 symbols were divided into 3 color groups: red, green, blue. Offline analyses revealed robust early visual and P300 components time-locked to the presentation of the target. The mean offline classification accuracy was 70% for both the 83 ms conditions, 90% for the color-133 ms and 85% in the no-color-133 ms, for selecting the correct symbol out of 30 possibilities. The RSVP speller has a large vocabulary and realizes a BCI independent on eye movements, since it is based on non-spatial selective feature attention. In this study, we improved the speller's design: in order to enhance the differences among the shape of the letters, we used both upper and lower cases. Like in the offline study, we investigated two color conditions, named no-color and color. However, in this study, we used 5 color groups for the color condition: red, green, blue, black and white. The two SOAs were 83 ms and 116 ms.

[a] This work was partly supported by grants of the Deutsche Forschungsgemeinschaft (DFG) (MU 987/3-1) and Bundesministerium fuer Bildung und Forschung (BMBF) (FKZ 01IB001A, 01GQ0850) and by the FP7-ICT Programme of the European Community, under the PASCAL2 Network of Excellence, ICT- 216886. This publication only reflects the authors' views. Funding agencies are not liable for any use that may be made of the information contained herein.

Laura Acqualagna (corresponding author, lauraacqualagna@libero.it) and Benjamin Blankertz (benjamin.blankertz@tu-berlin.de) are with the Berlin Institute of Technology, Machine Learning Laboratory, Berlin, Germany.

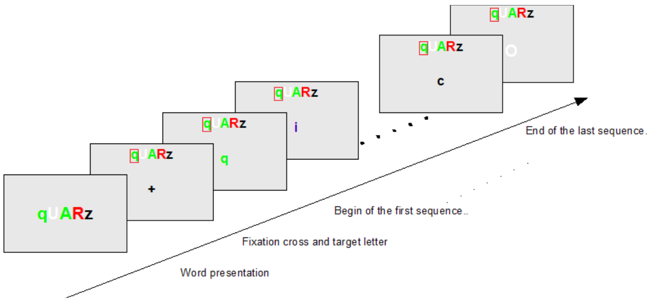

Fig. 1. Presentation of the stimuli based on the RSVP paradigm. Here an example of the word in the color condition is depicted. One sequence consisted in 30 symbols chosen either lowercase or uppercase and divided in 5 color groups: red, green, blue, black and white. After repeating the sequence 10 times, in the copy-spelling and free-spelling the output of the classifier was displayed and the next trial started.

II. Methods

A. Participants

12 subjects (6 males, 6 females, aged 25-52) participated in the experiment. All had normal or corrected-to-normal visual acuity. Normal color vision was confirmed using the Nishihara color vision test. None of them had a history of a neurological disease or injury. The study was performed in accordance with the declaration of Helsinki and all participants gave written consent.

B. Apparatus

EEG was recorded at 1000 Hz using BrainAmp amplifiers and an actiCAP active electrode system (Brain Products, Munich, Germany) with 64 electrodes. All the electrodes were placed according to the 10-20 system and referenced to left mastoids. For off-line analysis, all the electrodes were re-referenced to linked mastoids. All impedances were kept below 10 kΩ. Stimuli were presented on a 24" TFT screen with a refresh rate of 60 Hz and a resolution of 1920 x 1200 px^2.

C. Design and procedure

The RSVP consisted in 30 different symbols, comprising the 26 letters of the English alphabet, 2 punctuation marks: '.', '!', the underscore, used as space symbol, and the backspace, which could be used to erase the previous symbol. In order to enhance the differences between the shapes of the different letters, some were uppercase and some lowercase. To compare the performances got using just the uppercase letters, please refer to [8]. They had an height of 3.5 cm (1° visual angle). We investigated 3 different conditions: 1) no-color-116 ms; 2)color-116 ms; 3) color-83ms. In the no-color condition all the letters were black, in the color they were divided into 5 color groups: red (fRyGk<), white (pJUX!E), blue (iSwc_N), green (TBMqAH), black (LdvOz.). For the color condition, two stimulus onset asynchronies (SOAs) have been used: 83 ms and 116 ms. The visual presentation was time-locked with the screen refresh.

Participants sat at a distance of approximately 80 cm from the screen. For each condition, they had to perform a calibration, phase, copy-spelling phase and free-spelling phase. The order of the conditions was counterbalanced across participants. In the calibration, they had to spell the sentence 'BRAiN_cOMpUTER_iNTERfAcE'. The sentence was shown on the screen prior the start of the trial. Then, it was displayed on the top of the screen with the current target letter highlighted. After 4 seconds, a fixation cross appeared in the center of the screen for 3 seconds and the RSVP started (Fig.1). Stimuli were presented on a gray background. The 30 symbols were randomly shuffled and presented 10 times. Between the 10 sequences there was a short break of 0.3 s. In the color condition, the sequences have been chosen such that the order of the colors was fixed in order to ease the allocation of attention to the target symbol. Participants had to silently count the number of occurrences of the target letter in the RSVP, always fixating the center of the screen. In this phase, there was no feedback and the recorded EEG was used to train the classifier.

After calibration and training of the classifier, the on-line phases began. In the copy-spelling, subjects had to spell another sentence, which was different in the three conditions: 'LET_yOUR_BRAiN_TALk' in the color-116, 'wiNTER_iS_dEpRESSiNG' in the no-color-116, and 'doNT_wORRy_BE_HAppy' in the color-83. In this phase, after the 10 sequences, a symbol was selected based on the classifier's output and presented on the screen. In case of incorrect classification, there was no option to repeat the trial, but they had to proceed to the next one. Except for this, the instructions were the same as in the previous calibration phase. In the free-spelling phase, participants were asked to conceive a sentence containing at least 15 symbols. They were instructed to use the backspace symbol if the classifier selected the wrong letter. During this phase, all the 30 symbols were displayed on the bottom of the screen in alphabetical order, in order to remind the subjects about the capitalization and the color of the letters.

The RSVP speller was implemented in the open-source framework Pyff [9] using VisionEgg [10].

D. Data analysis

For ERP analysis, EEG data were down-sampled to 200 Hz and lowpass filtered with a Chebiyshev filter using passbands of 40 Hz and stopbands of 49 Hz. They were divided into epochs ranging from -100 ms to 1200 ms relative to the onset of each stimulus. Baseline correction was performed on the pre-stimulus period of either 116 ms (for the 116ms SOA) or 83 ms (for the 83 ms SOA). Epochs containing eye movements were detected and rejected using a min-max criterion (75μV) on the channels F9, Fz, F10, AF3 and AF4. For the ERP analysis, contamination of the non-target epochs by target presentations was reduced by considering only those target epochs wherein the three preceding and the three following symbols were also non-targets. For classification, all epochs were used. For the grand average, the ERP curves were averaged across all trials and participants. To compare the ERP curves of two classes (target-non-target), sgn r^2-values based on the point-biserial correlation coefficient were calculated. Sgn r^2-values were averaged across participants by using the *z*-transform. Classification was based on linear discriminant analysis (LDA) with shrinkage of the covariance matrix [11]. The time intervals for calculating the spatio-temproal features were determined by a heuristic searching for peaks based on the sgn r^2 [11]. During the training of the on-line classifier, 5 different temporal windows were selected and occasionally adjusted by the experimenter. The spelling-speed was calculated considering in the selection duration of one symbol: the time of the fixation cross and target presentation (7 s), the SOAs, the time necessary to display the classifier's output (1 s), the inter-sequences intervals and the number of sequences.

III. Results

A. ERPs

The grand average ERPs for the three conditions are shown in Figure 2. The plots on the left side represent the time-course of the ERPs at channels CPz and PO7. In all conditions, two main components arise time-locked with the target onset: the N2 visual component in channel PO7, focused in the occipital cortex between 300 ms and 400 ms, and the P3 component in channel CPz. In particular, it can be distinguished the P3a, which originates in a frontal location and peaks around 300 ms, and the P3b, focused in the central-parietal cortex around 500 ms. The scalp-plots adjacent show the spatial distribution of the sign- r^2 and refer to shadowed areas in the ERPs plots. The locations of the N2 and P3 components are evident where there is a high difference between the target and non-target classes. P3 and N2 amplitudes were subjected to a two-way analysis of variance (ANOVA) with factors "Condition" (no-color 116 ms, color 116 ms, color 83 ms) and "Status" (target, non-target). The analysis of the P3 amplitude reveals a significant effect of the Status ($F=244.67$, $p<0.001$), with targets having higher amplitude, but not of the Condition ($p=0.298$). Regarding the N2 component, there is a significant effect of both the Status ($F=5.1$, $p<0.05$) and the Condition ($F=8.63$, $p<0.001$). Tukey-Kramer post-hoc test reveals that the N2 amplitude is lower in the color 83 ms condition than in the color 116 ms and in the no-color 116 ms. For both the P3 and N2 amplitudes the Status x Condition interaction is not significant ($p=0.42$, $p=0.63$).

B. Classification

The mean online classification accuracy approaches 94.8 % in the no-color-116 ms condition, 94.7% in the color-116 ms and 93.6% in the color-83 ms, after 10 sequences. One-way ANOVA analysis reveals that the differences among the conditions are not significant ($p=0.756$). In the offline analyses, the data of the calibration phase were used as training set, and then the classifier was applied to the copy-spelling and the free-spelling. The performance of the classifier increases significantly with the number of sequences ($F=103.58$, $p<0.001$). These results show that the modifications made on the speller's design highly improved the classification accuracy, especially in the 83 ms condition.

The evaluation is based on the accuracy of symbol selection, one out of 30. The chance level is 3,33%. Figure 3 depicts the spelling speed, expressed in number of symbols per minute, as a function of the number of sequences, in the three conditions. The colored lines refer to the different subjects and the black thick line represents the mean spelling speed over all the subjects.

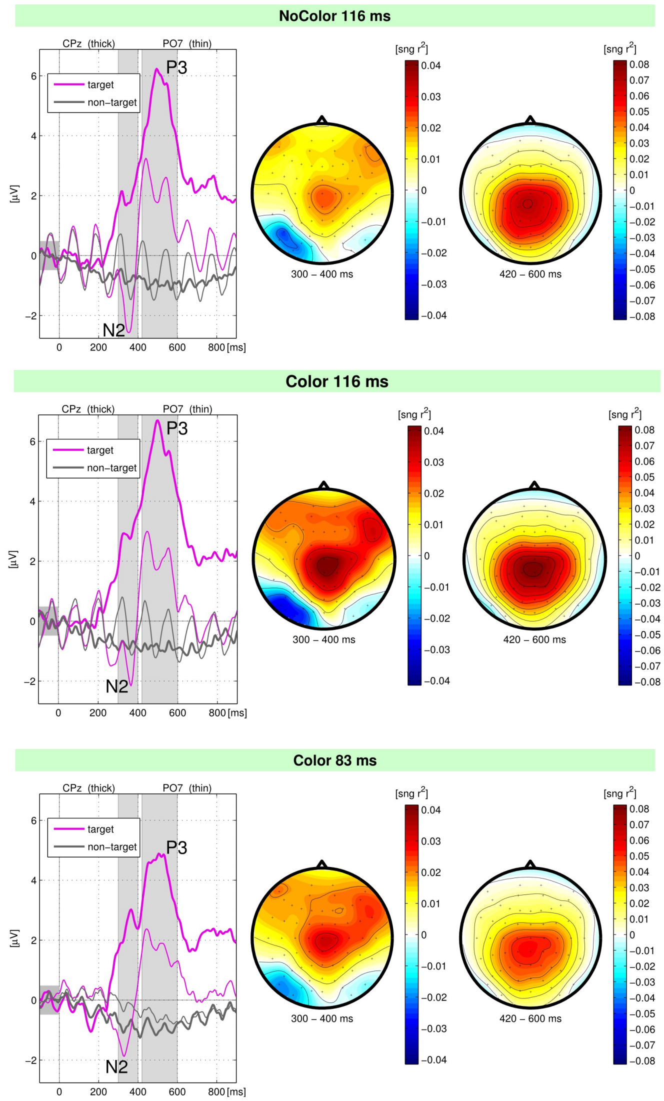

Fig. 2. Grand average of the ERP components in the three conditions. The plots on the right side show the time evolution of the ERPs for targets and non-targets. The onset of the target elicits strong P3 and N2 components. The scalp plots on the left side refer to the shaded areas of the plots adjacent and represent the signed correlation coefficient between the target and non-target classes.

The online spelling speed reaches 1.16 symb/min in the no-color 116ms (5.65 bits/min), 1.15 symb/min in the color 166ms (5.66 bits/min), and 1.43 symb/min in the color 83ms (7 bits/min). Offline analyses show that the maximum mean spelling speed is of 2 symb/min (10.5 bits/min) in the color-116 ms after 3 sequences (83.84 % of classification accuracy) and in the color 83 ms after 4 sequences (83.4% classification accuracy). In the no-color condition the maximum mean speed is about 1.75 symb/min (9.6 bits/min) after 3 sequences (79.2% classification accuracy).

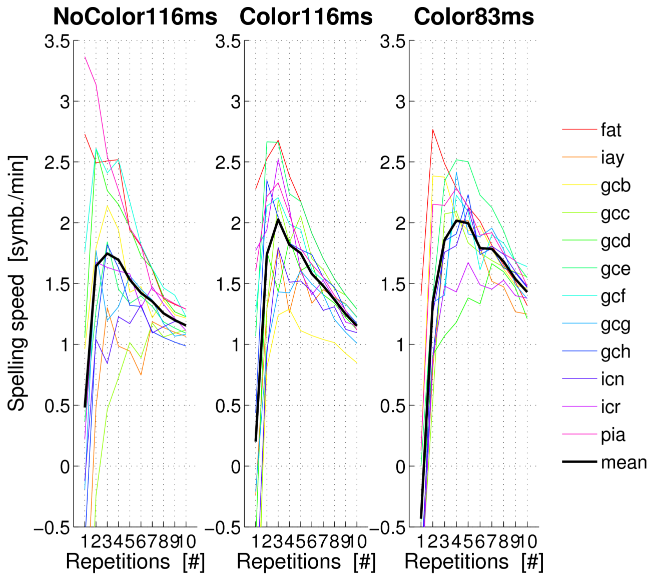

Fig. 3. Spelling speed expressed in number of symbols per minute as a function of the number of sequences. The symbol rate reaches its maximum mean value in the color 116 ms condition after 3 sequences and in the color 83 ms condition after 4 sequences. In the no-color condition 4 subjects achieve a spelling speed higher than 2.5 symb/min after 1 or 2 sequences.

IV. Discussion

Mean symbol selection accuracies of 94.8%, 94.7% and 93.6% were reached respectively in the conditions no-color 116 ms, color 116 ms and color 83 ms. The modifications made on the speller's design highly improved the classification accuracy compared to the previous offline study (90% for the no-color 133 ms, 85% for the color 133ms and 70% for the color 83ms) [8]. All the participants achieved an accurate control of the speller. The accuracies achieved are competitive in respect to the other gaze-independent visual ERP-spellers [6,7]. The RSVP speller has a considerably higher classification performance than the BCI system investigated in Zhang et al. [12], which uses the covert attention to modulate steady-state visual evoked potential (SSVEP), and than BCIs operated with auditory and tactile stimuli [4,5]. The maximum online spelling speed was 1.43 symb/min in the color 83 ms condition. This result was obtained considering the time necessary to show the target letter, the fixation cross, the speller's output. This calculation reflects the actual online selection rate. In some studies [13], authors calculate the number of selections per minute also excluding the time inter-selections. In order to compare our results with such studies, the spelling speed was recalculated offline, obtaining a maximum mean speed of 4 symb/min in all the 3 conditions after 2 sequences (classification accuracies: 72.3% no-color 116ms; 73.6% color 116ms; 66% color 83ms). This throughput is higher than the symbol rate reported in [7,4], but some improvements have to be introduced in order to make it competitive to the speed reached by visual spellers based on the gaze shifts [13,14]. The throughput of the RSVP speller could be ulteriorly increased introducing techniques of early stopping [15] and error detection [16]. To conclude, the present online study showed that the RSVP speller can be operated with high control by healthy subjects. It reaches high classification accuracy, it's fast-paced and has a large vocabulary. It is a gaze-independent BCI, so it can represent a valid choice for people with impaired oculomotor control.

Acknowledgment

We are grateful to Matthias Treder for stimulating discussions concerning the experimental design and Torsten Schmits for programming the stimulus presentation.

References

[1] Farwell, Donchin, “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials,” Electroencephalogr Clin Neurophysiol, vol. 70, pp. 510–523, 1988.

[2] Treder M. and Blankertz, B. “(C)overt attention and speller design in visual attention based brain-computer interfaces,” Behavioral & Brain Functions, 6:28, May 2010.

[3] P Brunner, S Joshi, S Briskin, J R Wolpaw, H Bischof, and G Schalk. Does the "P300" speller depend on eyegaze? J Neural Eng, 7:056013, 2010.

[4] Schreuder M., Blankertz B. and Tangermann M. A new auditory multi-class brain-computer interface paradigm: Spatial hearing as an informative cue. PLoS ONE, 5(4):e9813, 2010.

[5] Brouwer A and van Erp JB (2010) A tactile P300 brain-computer interface. Front. Neurosci. 4:19. doi:10.3389/fnins.2010.00019.

[6] Treder MS, Schmidt NM, and Benjamin Blankertz. Towards gaze-independent visual braincomputer interfaces. 2010. Conference Abstract: Bernstein Conference on Computational Neuroscience 2010.

[7] Y. Liu, Z. Zhou, D. Hu, Gaze independent brain–computer speller with covert visual search tasks, Clinical Neurophysiology, October 2010.

[8] Acqualagna L.,Treder MS.,Schreuder M., and Blankertz B.. A novel braincomputer interface based on the rapid serial visual presentation paradigm. In Proceedings of the 32nd Annual International IEEE EMBS Conference, 2010.

[9] Venthur B, Scholler S, Williamson J, Daehne S., Treder MS, Kramarek MT, Müller KR, and Blankertz B. Pyff- a pythonic framework for feedback applications and stimulus presentation in neuroscience. Front Neuroscience, 2010. in press.

[10] A. D. Straw. Vision Egg: an open-source library for realtime visual stimulus generation. Front Neuroinformatics, 2:4, 2008.

[11] B. Blankertz, S. Lemm, M. S. Treder, S. Haufe, K. R. Müller, Single-trial analysis and classification of ERP components - a tutorial, NeuroImage,2010,in press.

[12] D. Zhang, A. Maye, X. Gao, B. Hong, AK Enge and S. Gao, “An independent brain–computer interface using covert non-spatial visual selective attention”, J. Neural Eng. 7 (2010) 016010 (11pp)

[13] G. Townsend, B.K. LaPallo, C.B. Boulay, D.J. Krusienski, G.E. Frye, C.K. Hauser, N.E. Schwartz,T.M. Vaughan, J.R. Wolpaw, E.W. Sellers, A novel P300-based brain–computer interface stimulus presentation paradigm: Moving beyond rows and columns, Clinical Neurophysiology, 2010.

[14] E. Donchin, K. M. Spencer, and R. Wijesinghe. The mental prosthesis: Assessing the speed of a P300-based brain-computer interface. IEEE Trans Rehabil Eng, 8(2):174-179, June 2000.

[15] Alexander Lenhardt, Matthias Kaper, and Helge J. Ritter, An Adaptive P300-Based Online Brain–Computer Interface, IEEE Trans Rehabil Eng, 16, 8, 2008

[16] Schmidt N, Blankertz B, Treder M, Online detection of error potentials increases information throughput in a brain-computer interface, Neuroscience Letters (Conference Abstract: Meeting of the Society of Applied Neurosciences (SAN) 2011), in press.