Suppressing Flashes of Items Surrounding Targets during Calibration of a P300-Based Brain-Computer Interface Improves Performance

GE Frye[1], **CK Hauser**[1], **G Townsend**[2], and **EW Sellers**[1]

[1]East Tennessee State University, Johnson City, Tennessee 37601, USA

[2]Algoma University, Sault Ste. Marie, Ontario P6A 2G4, CAN

Abstract

Since the introduction of the P300 BCI speller by Farwell and Donchin in 1988 speed and accuracy of the system has been significantly improved. Larger electrode montages and various signal processing techniques are responsible for most of the improvement in performance. New presentation paradigms have also led to improvements in bit rate and accuracy (e.g. Townsend et al. 2010). In particular, the checkerboard paradigm for online P300 BCI-based spelling performs well, has started to document what makes for a successful paradigm, and is a good platform for further experimentation. The current paper further examines the checkerboard paradigm by suppressing items which surround the target from flashing during calibration (i.e., the suppression condition). In the online feedback mode the standard checkerboard paradigm is used with a stepwise linear discriminant classifier derived from the suppression condition and one classifier derived from the standard checkerboard condition, counter-balanced. The results of this research demonstrate that using suppression during calibration produces significantly more character selections/min ((6.46) time between selections included) than the standard checkerboard condition (5.55), and significantly fewer target flashes are needed per selections in the SUP condition (5.28) as compared to the RCP condition (6.17). Moreover, accuracy in the SUP and RCP conditions remained equivalent (~90%). Mean theoretical bit rate was 53.62 bits/min in the suppression condition and 46.36 bits/min in the standard checkerboard condition (ns). Waveform morphology also showed significant differences in amplitude and latency.

Keywords

Brain-Computer Interface; EEG; P300; Event-Related Potential; Communication and Control

1. Introduction

In 1988, Farwell and Donchin [1] introduced the first P300-based BCI paradigm in which a computer presents a 6x6 matrix of letters and commands on-screen and participants attend to the item they wish to select. In this first P300 BCI paradigm, and in most since, items are grouped as flashing rows and columns: hence, the nomenclature row/column paradigm, or RCP. The intersection of the row and the column that elicited the combination of the largest and most temporally consistent (or classifiable) response is identified as the attended item by the classification algorithm. With the target of achieving efficient and practical in-home use,

Corresponding author: Eric W. Sellers, Ph.D., Director, Brain Computer Interface Laboratory, Assistant Professor, Department of Psychology, East Tennessee State University, 807 University Parkway, Johnson City, TN 37614, Office Phone: 423.439.4476, Lab Phone: 423.439.5695, Fax: 423.439.5695, sellers@etsu.edu.

researchers have tested presentation paradigm design qualities such as inter-stimulus interval (ISI) and matrix size [2], electrode montages [3], and signal processing methods [4–7]. However, the RCP remains subject to design errors that slow communication. For instance, errors typically occur with the greatest frequency in locations adjacent to the attended item (i.e., the target item) and almost always in the same row or column [8–10]. Recently this has been referred to as ''adjacency distraction'' [10] and it occurs when an item surrounding the target inadvertently attracts attention, thereby creating a target response (i.e., P300). In contrast, when the target item flashes and is correctly attended to (i.e., produces a P300), all items flashing with the target, produce a target response due to temporal proximity with the target. This is especially problematic in the RCP because each time the target row or column produces a P300 all other items in the row or column include a P300 [11]. Thus, if adjacency distraction occurs on one or more trials, a nontarget can be identified as the target because the classifier applies coefficients to each row and each column, sums the scores, and the cell with the highest row and column score is identified as the target response.

Some have explored alternatives to the RCP paradigm. Guger et al. [12] compared the RCP to a paradigm which randomly flashed single items. Martens et al. [13] compared an RCP speller to a paradigm making use of apparent motion; however the motion occurred in rows and columns. Takano et al. [14] investigated RCP accuracy using three different luminance and chromatic flash patterns; the luminance/chromatic condition produced online accuracy higher than the luminance or chromatic conditions alone. Hong et al. [15] compared the RCP to an apparent motion and color onset paradigm that also presented in a row/column arrangement. Salvaris and Sepulveda [16] compared changes to the character size, distances between characters, and background/foreground colors. Others have designed paradigms that do not rely on variations of the RCP [17]. However, none of these manipulations have resulted in substantive improvements in performance.

The checkerboard paradigm or CBP, completely disassociates rows and columns [10]. In the CBP, the items of an 8x9 matrix are logically separated by superimposing the 72 matrix items into a "virtual" checkerboard (that is not seen by the participants). The items in "white" cells of the 8x9 matrix are logically segregated into one 6x6 matrix and the items in the "black" cells are segregated into another 6x6 matrix. Before each sequence of flashes, the designated items randomly populate the white or black matrix, respectively. The end result is that the participants see quasi-random groups of six items flashing. As vertically or horizontally (but not diagonally) adjacent items cannot be included in the same flash group, and thus cannot flash simultaneously, the virtual checkerboard layout partially controls for adjacency-distraction errors. The CBP also introduces a constraint that does not allow any matrix item to flash a second time for a minimum of six intervening flashes or a maximum of 18 flashes. This constraint avoids the problem of overlapping target epochs. The expansion to an 8x9 matrix was expected to produce larger P300 amplitudes for the target items by reducing the probability of the target stimulus occurring [2, 18, 19]. Townsend et al [10] showed that online accuracy was significantly higher in the CBP (92%), as compared to the RCP (77%), demonstrating that the CBP is superior to the RCP presentation method.

As a paradigm that performs well and has begun to document what makes a successful paradigm, the CBP is a good platform for further experimentation on a practical level, and as a means of exploring what is necessary in a robust paradigm. For example, the CBP still allows for paradigmatic errors. Item flashes diagonally adjacent to the target can potentially result in adjacency-distraction. As is well documented in spatial attention literature, in a standard flanker task, response time significantly increases when nearby items belong to a response class different from the target class [20]. In the RCP, when adjacency-distraction errors occur, the distractions typically cause another item in the same row or column as the target to be selected unintentionally [8–10]. That is, given that each time a target character

flashes all items in the target row (and/or) column produce P300 responses. Thus, adjacency-distraction (i.e., distraction from a row or column adjacent to the target) may change the target ERP in such a manner that responses to erroneous items more closely resemble the expected target ERP. In other words, the SWLDA classifier models target and non-target responses. Thus, when an error is made, one can assume that the erroneously selected character is more similar to the canonical target response than the desired target is to the canonical response. However, the current study was not designed to explicitly examine this hypothesis. It should also be noted that Townsend et al [10] showed that in the CBP 5% of the errors occurred in the target row or column, while the RCP produced 84% of the errors in the target row or column.

Building upon the results of [10], removing these errors should create performance improvements. In the case of the CBP, and likely for other presentation paradigms, the easiest way to remove the effects of adjacency-distraction is to remove the possibility of simultaneously flashing any of the eight immediately adjacent items during the calibration phase of the experiment. The current paper examines this hypothesis by using a completely suppressed CBP (SUP) where none of the eight surrounding items flash simultaneously with a target during calibration (figure 1(a)). Two competing hypotheses are as follows: 1) A stepwise linear discriminant (SWLDA) classifier derived from the SUP condition will perform better than a classifier derived from the CBP because the adjacency distraction has been removed, presumably producing a more reliable ERP; 2) the SUP classifier will not perform as well as the CBP classifier because it will not generalize to the standard CBP presentation during online testing.

2. Methods

2.1. Participants

Eighteen able-bodied adults (9 female, 9 male; age range 19–49) were recruited from the East Tennessee State University undergraduate psychology participant pool. Fourteen were completely naïve to BCI use and 4 had previous BCI experience. All had normal or corrected-to-normal vision and no known cognitive deficit. This work was approved by the East Tennessee State University Institutional Review Board and each participant provided informed consent.

2.2. Data Acquisition

Each participant sat in a comfortable chair approximately 1 m from a computer monitor that displayed the 8x9 matrix. EEG was recorded with a 32-channel tin electrode cap (Electro-Cap International, Inc.). All channels were referenced to the right mastoid and grounded to the left mastoid. Impedances were reduced to below 10.0 kΩ before recording. Signals were digitized at 256 Hz, and bandpass filtered from 0.5 Hz to 30 Hz using two g.tec (Guger Technologies) 16-channel USB biosignal (g.USBamp version 2) amplifiers. Electrodes Fz, Cz, Pz, Oz, P3, P4, PO7, and PO8 were used for BCI operation [3]. BCI2000 [21] was used for stimulus presentation, and data collection.

2.3. Experimental Paradigm

Each participant completed two experimental sessions within a one-week period. Classification coefficients were generated with data collected during a calibration phase and subsequently applied during an online test phase. In each phase, participants were provided with strings of items to select. The participant's task was to attend to the number of times the item in parentheses flashed (by counting or mentally repeating the target. During SUP calibration (as shown in figure 1(a; left)) the target item "D" is flashing and none of the surrounding items are flashing. During CBP calibration items diagonal to the "D" could

flash. One sequence of flashes included 24 flashes; each flash consisted of six items. For each of 36 item selections, five complete sequences (i.e., 10 flashes of each matrix item) occurred. Sessions were counterbalanced such that half of the participants began with the SUP session and the other half began with the CBP session. During the calibration phase of the SUP session, participants were presented with flash patterns which did not include any simultaneous flashes of adjacent items (figure 1(a)). During the calibration phase of the CBP, participants were presented with standard CBP flash patterns which included diagonally adjacent flashes, but not horizontal or vertical adjacent flashes. The event sequences (i.e., flash patters) and target to target intervals for the SUP and CBP were identical during calibration. In addition, during online testing, the event sequences for SUP and CBP were identical (i.e., no suppression). This was achieved by presenting subjects with identical pre-determined flash patterns that were produced within the constraints of the CBP. The online test phase was identical to the calibration phase except for two differences. First, the number of item flashes-per-selection was changed from 10 to a participant-specific number (described below). Second, item selections were classified using SWLDA feature weights generated from the calibration data and visual feedback of the selections was provided to the participant directly below the item to be selected. Both SUP and CBP operated in the standard CBP (i.e., no suppression) during online testing.

2.4. Classification

Independent SWLDA classifiers were used to determine the signal features that best discriminated between target and non-target flashes (MATLAB version 7.6 R2008a, stepwisefit function) [22]. Classifiers were derived separately for the SUP and CBP, as described in Krusienski et al. [3]. The SWLDA algorithm was then used for online classification. Epochs from each of the 72 stimulus items were averaged before applying the SWLDA classification coefficients. Coefficients were then applied to the specific spatiotemporal features of each of the 72 items of the matrix and summed. The item with the highest score was selected and presented to the participant as feedback.

2.5. Determining the optimal number of sequences

Due to the P300 response's relatively low signal-to-noise ratio, each item must be flashed multiple times and averaged [23]. During calibration, the number of target item flashes was constant across participants and presentation methods. Item sets were flashed in quasi-random sequences with two flashes of the target item per sequence, and thus 10 target item flashes in the five sequences used for each selection. During the online testing phase, we optimized the number of sequences by calculating each participant's maximum written symbol rate (WSR, or symbols/min; [24]). This metric represents the number of item selections a participant can correctly make in 1 min, taking into account error correction.

3. Results

3.1. Optimal Flashes/Selections, Selections/Minute, Bit Rate, and Accuracy

When participants were characterized with SUP, as opposed to CBP, their mean optimal flashes-per-selection was 5.28; mean flashes-per-selection in the CBP was 6.17. This totals to an average difference of 0.89 flashes-per-selection (or approximately one target flash; p=0.019). As shown in figure 1(b) the SUP condition also resulted in significantly higher mean selections-per-minute, with SUP at 6.46 and CBP at 5.55 (Δ= 0.91; p=0.008). This calculation includes time between selections, which is necessary in practical online use; in this study 3.5s were used, which is consistent to the time between selections used by [10]. We have also conducted other studies that used 5.0s between selections [2, 25]; however, given that the current study extends that reported in [10], we opted to use 3.5s. Moreover using a short inter-selection interval increases the practical utility of the system. Thus, this

value represents the mean number of online selections/min. Theoretical selections-per-minute were also higher for SUP at 10.95 and CBP at 8.95 (Δ= 2.0; *p*=0.047), this value is calculated with the 3.5 s between selections removed (for comparison to other studies that omit time between selections when reporting bit rate). These are the highest group mean online bit rates (SUP=31.66 (range of 16.70 to 54.42) and CBP=28.75 (range of 10.60 to 48.68; p=0.114) and group mean theoretical bit rates (SUP=53.62 (range of 20.87 to 108.79) and CBP=46.36 (range of 13.25 to 97.31); p=0.112) reported to date. There was not a significant difference in selection accuracy between the SUP and CBP, 87.7% and 89.8%, respectively. Given that the same algorithm (WSR) was used to optimize the number of target flashes per selection it is not surprising that accuracy was similar in both conditions. This is because the tradeoff between additional time per selection (i.e., number of flashes) diminishes as accuracy begins to asymptote.

3.2. Waveform Morphologies

The CBP and SUP conditions produced different waveforms in several respects. Our analyses focused on four electrodes, (Cz, Pz, PO7, and PO8) as these electrodes have previously been shown to typically contribute most to classification accuracy [4, 16]. The analyses were conducted using the calibration data to hold the amount of data per participant constant because variable numbers of flashes were used during the online test. Figure 1(c) shows target (top row) and non-target response grand means for all participants (N=18) at each of the four electrodes. Amplitude and latency differences were examined at each electrode location.

Positive peak amplitude and latency between 125 and 350 ms were measured at each electrode location. The peak at electrode Cz was significantly earlier for the SUP (242 ms) than the CBP (258 ms; p=0.049). Also, latency and amplitude differences in the positive peak were observed for electrode PO7. Peak latency was earlier for the SUP (249 ms) than for the CBP (283 ms; p=0.020 and SUP amplitude was lower (2.701 µV) than CBP amplitude (3.124 µV; p=0.009).

Negative peak amplitude and latency between 300 and 600 ms were also examined. At electrode Pz, SUP showed a significantly earlier peak (449 ms) than CBP (470 ms; p=0.036). The negative peak at electrode PO8 in the SUP was also significantly earlier (455 ms) than the CBP (494 ms; p=0.028).

4. Discussion

Calibrating with SUP increased data throughput: selections-per-minute, and theoretical selections-per-minute both increased, while the optimally required flashes-per-selection decreased. The SUP reduction in flashes-per-selection resulted in an additional 0.89 selections/min. The fact that SUP and CBP showed equal mean accuracy is explainable, in part, because the WSR was used to determine the optimal number of flashes to be used online. Each sequence takes three seconds; thus, as soon as accuracy begins to asymptote, the minimal tradeoff in improved accuracy is not worth the time it takes to present additional sequences of stimuli. On a practical level, SUP calibration increased online speed without reducing accuracy, thus improving performance of the CBP to the highest mean P300 BCI bit rate reported to date. Moreover, given that the two experimental paradigms were identical in every aspect (i.e., stimulus sequence and target to target interval) except for the fact that surrounding characters did not flash during SUP calibration; the current results indirectly suggests that SUP calibration provides a more stable SWLDA classification algorithm. This, in fact, is somewhat surprising given that during online testing items that can cause adjacency distraction were present.

These results suggest that suppressing adjacent items is an important consideration in paradigm design. Optimal signal quality is achieved during calibration as a function of the presentation paradigm: before filters, amplifiers, or classification algorithms. Optimal signal quality (e.g., spatial and temporal consistency of the participant's response) produced by manipulations at the paradigmatic level can be reasonably expected to translate into better performance at later processing stages as well. As experimenters have the greatest ability to manipulate ERP morphology through paradigm manipulations and innovations, it is practical to focus on extending basic psychophysiological paradigms that have been extensively studied for more than fifty years. More generally, BCI paradigm design could benefit from the knowledge provided by various paradigm designs that affect waveform morphology in different ways and also focus on the vast cognitive psychophysiology literature regarding comparisons between clinical populations and healthy controls.

5. Conclusions and Future Directions

The results of this research demonstrate the efficacy of SWLDA coefficients derived while suppressing all items that surround the target character during calibration and then applying those coefficients to an online version of the speller that *does not* suppress adjacent items. These results show that more character selections/min are produced by using fewer flashes for each selection, while accuracy remains equivalent in the SUP condition and in the standard CBP. An important next step is to replicate the present research using participants with severe neuromuscular diseases.

Acknowledgments

References

1. Farwell LA, Donchin E. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. Electroencephalogr Clin Neurophysiol. 1988; 70(6):510–523. [PubMed: 2461285]

2. Sellers EW, et al. A P300 event-related potential brain-computer interface (BCI): the effects of matrix size and inter stimulus interval on performance. Biol Psychol. 2006; 73(3):242–252. [PubMed: 16860920]

3. Krusienski DJ, et al. Toward enhanced P300 speller performance. J Neurosci Methods. 2008; 167(1):15–21. [PubMed: 17822777]

4. Lenhardt A, Kaper M, Ritter HJ. An adaptive P300-based online brain-computer interface. IEEE Trans Neural Syst Rehabil Eng. 2008; 16(2):121–130. [PubMed: 18403280]

5. Kaper M, et al. BCI Competition 2003--Data set IIb: support vector machines for the P300 speller paradigm. IEEE Trans Biomed Eng. 2004; 51(6):1073–1076. [PubMed: 15188881]

6. Kaper M, Ritter H. Generalizing to new subjects in brain-computer interfacing. Conf Proc IEEE Eng Med Biol Soc. 2004; 6:4363–4366. [PubMed: 17271271]

7. Serby H, Yom-Tov E, Inbar GF. An improved P300-based brain-computer interface. IEEE Trans Neural Syst Rehabil Eng. 2005; 13(1):89–98. [PubMed: 15813410]

8. Fazel-Rezai R. Human error in P300 speller paradigm for brain-computer interface. Conf Proc IEEE Eng Med Biol Soc. 2007; 2007:2516–2519. [PubMed: 18002506]

9. Donchin E, Spencer KM, Wijesinghe R. The mental prosthesis: assessing the speed of a P300-based brain-computer interface. IEEE Trans Rehabil Eng. 2000; 8(2):174–179. [PubMed: 10896179]

10. Townsend G, et al. A novel P300-based brain-computer interface stimulus presentation paradigm: moving beyond rows and columns. Clin Neurophysiol. 2010; 121(7):1109–1120. [PubMed: 20347387]

11. Sellers, EW., et al. Non-invasive brain-computer interface research at the Wadsworth Center. In: Dornhege, JMG.; Hinterberger, T.; McFarland, D.; Müller, K., editors. Towards Brain-Computer Interfacing. Cambridge, MA: MIT Press; 2007. p. 31-42.

12. Guger C, et al. How many people are able to control a P300-based brain-computer interface (BCI)? Neurosci Lett. 2009; 462(1):94–98. [PubMed: 19545601]

13. Martens SM, et al. Overlap and refractory effects in a brain-computer interface speller based on the visual P300 event-related potential. J Neural Eng. 2009; 6(2):026003. [PubMed: 19255462]

14. Takano K, et al. Visual stimuli for the P300 brain-computer interface: A comparison of white/gray and green/blue flicker matrices. Clin Neurophysiol. 2009

15. Hong B, et al. N200-speller using motion-onset visual response. Clin Neurophysiol. 2009; 120(9): 1658–1666. [PubMed: 19640783]

16. Salvaris M, Sepulveda F. Visual modifications on the P300 speller BCI paradigm. J Neural Eng. 2009; 6(4):046011. [PubMed: 19602731]

17. Hill, J., et al. Effects of stimulus type and of error-correcting code design on BCI speller performance. In: Koller, D., et al., editors. Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press; 2009. p. 665-672.

18. Allison BZ, Pineda JA. Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: implications for a BCI system. Int J Psychophysiol. 2006; 59(2):127–140. [PubMed: 16054256]

19. Allison BZ, Pineda JA. ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system. IEEE Trans Neural Syst Rehabil Eng. 2003; 11(2):110–113. [PubMed: 12899248]

20. Sanders AF, Lamers JM. The Eriksen flanker effect revisited. Acta Psychol (Amst). 2002; 109(1): 41–56. [PubMed: 11766139]

21. Schalk G, et al. BCI2000: a general-purpose brain-computer interface (BCI) system. IEEE Trans Biomed Eng. 2004; 51(6):1034–1043. [PubMed: 15188875]

22. Draper, NR.; Smith, H. Wiley series in probability and mathematical statistics. 2d ed.. New York: Wiley; 1981. Applied regression analysis; p. 709xiv

23. Cohen J, Polich J. On the number of trials needed for P300. Int J Psychophysiol. 1997; 25(3):249–255. [PubMed: 9105949]

24. Furdea A, et al. An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology. 2009; 46(3):617–625. [PubMed: 19170946]

25. Nijboer F, et al. A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. Clin Neurophysiol. 2008; 119(8):1909–1916. [PubMed: 18571984]
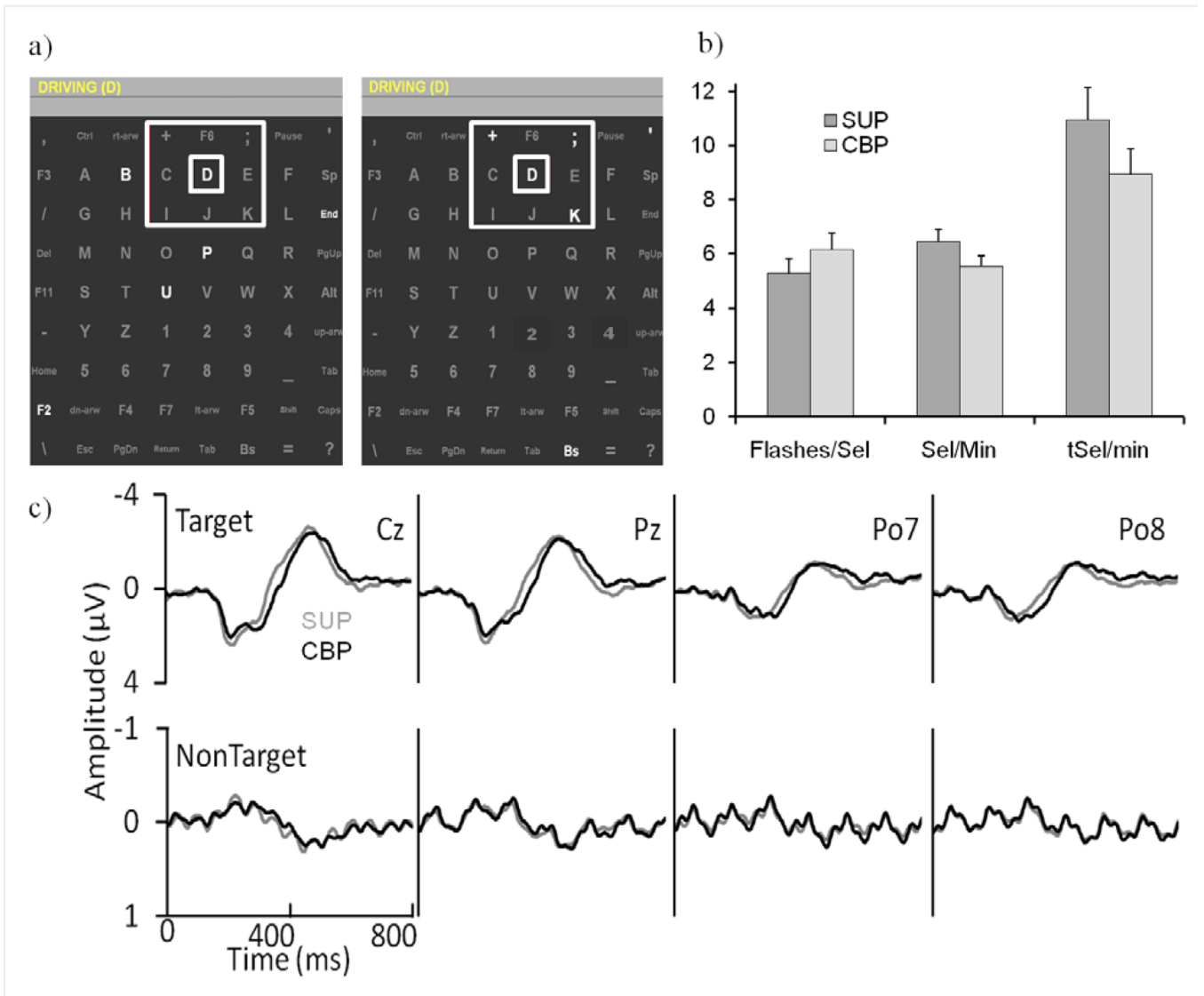
Figure 1.
Suppression condition presentation and data. (a) Left: Presentation of SUP (suppression condition) with flashes of all 8 adjacent diagonals restricted during a target flash. Right: A CBP (standard checkerboard) flash showing diagonal items flashing with the target. (b) Significant differences were found in flashes-per-selection, selections-per-minute and theoretical (t) selections-per-minute. (c) Grand mean waveforms for all 18 participants at electrode locations Cz, Pz, PO7, and PO8. The top row consists of target responses for both conditions, and the bottom row consists of non-target responses (SUP data is presented in grey and CBP data is presented in black).