

Comparison of sequence, stimulus and decoding algorithm combinations in c-VEP-based BCIs

A. Martín-Fernández^{1,2}, V. Martínez-Cagigal^{1,2,3}, E. Santamaría-Vázquez^{1,2,3}, R. Hornero^{1,2,3}

¹Biomedical Engineering Group, University of Valladolid, ETSIT, Paseo de Belen, 15, 47011, Valladolid, Spain

²Centro de Investigacion Biomedica en Red de Bioingenieria, Biomateriales y Nanomedicina (CIBER-BBN), Valladolid, Spain

³Valladolid Health Research Institute (IBioVALL), C. Rondilla Sta. Teresa, 47010, Valladolid, Spain

E-mail: ana.martin.fernandez23@uva.es

ABSTRACT: Brain-computer interfaces (BCIs) enable users to control external devices using brain activity. Among exogenous paradigms, code-modulated visual evoked potentials (c-VEP) stand out. These are neural responses elicited by flickering visual stimuli encoded with pseudorandom sequences, allowing the system to identify the target attended by the user. Optimizing c-VEP performance requires understanding the combined effects of encoding sequences, stimulus patterns, and decoding algorithms. In this study, 26 healthy participants evaluated a 16-target speller with 6 c-VEP configurations, combining 2 sequences (m-sequences, burst codes) with 3 stimulus patterns (plain, grating, checkerboard) using 4 decoding algorithms. Results showed that burst codes with a checkerboard pattern and a bit-wise reconstruction decoder offered the best balance between accuracy (99.76%) and comfort (eyestrain score 4.81/10). However, several combinations also yielded reliable performance, highlighting that optimal design may depend on the specific application needs. This work emphasizes the importance of jointly considering sequence type, stimulus pattern, and decoding algorithm in c-VEP research.

INTRODUCTION

Brain-computer interfaces (BCIs) can be described as systems that enable individuals to interact with the environment relying solely on their brain activity [1]. By establishing a direct communication pathway between the brain and an external application, BCIs translate recorded neural signals into executable commands through signal processing and intention decoding algorithms. Over the past decades, substantial research efforts have advanced the development of these systems, demonstrating their feasibility across a wide range of applications [1].

Despite relying on a common operational framework, BCIs may exhibit considerable variability depending on the intended use. One aspect common to most BCI systems is the method used to acquire brain activity, which is typically electroencephalography (EEG), due to its non-invasive nature, relatively low cost, and high temporal

resolution compared to other techniques [2]. In contrast, the paradigm employed to decode user intentions is highly application-dependent. Rehabilitation-oriented systems typically rely on endogenous paradigms, in which users learn to voluntarily modulate their brain activity [1]. Conversely, communication and control systems more often employ exogenous paradigms, where external stimuli are presented to elicit specific, stimulus-driven neural responses [1].

Among the exogenous approaches that have gained considerable attention in the literature are code-modulated visual evoked potentials (c-VEP) [3]. These signals arise as natural brain responses to visual stimulation in which brief light flashes are presented according to a pseudorandom sequence. By analyzing the EEG activity the system can determine which target the user is focusing on and execute the task associated with it [3]. Although impressive performance has been demonstrated, with reported accuracies exceeding 90%, the main challenges of c-VEP-based BCIs lie in maintaining such performance while reducing calibration and selection times [4], as well as ensuring an adequate user experience by minimizing visual fatigue induced by prolonged stimulation [5, 6]. Research efforts aimed at optimizing these systems have primarily focused on three main aspects: encoding sequence, stimulus pattern, and decoding algorithms.

Regarding the encoding sequences used to codify the visual stimulation, it is necessary to employ sets of codes with minimal cross-correlation, as greater dissimilarity between sequences facilitates accurate target classification. Several families of codes have been proposed, such as m-sequences [3] or burst codes [5]. Concerning stimulus patterns, the classical binary approach represents the 0-bit as black and the 1-bit as white. However, due to the visual discomfort induced by high-contrast flashes, alternative designs have been proposed, incorporating edges and contours instead of plain patterns. These designs include checkerboard [6] and grating-like stimuli [5]. Finally, with respect to decoding algorithms, two main processing pipelines are commonly followed: template matching (TM) [3, 7] and bit-wise reconstruction (BWR)

[5, 8]. The TM approach averages and spatially filters the user's EEG responses to each stimulation code to generate representative templates. During online operation, the incoming EEG signal is compared against these templates. On the other hand, the BWR approach trains a classifier to predict whether a given EEG epoch corresponds to a 0-bit or a 1-bit stimulation. The reconstructed bit sequence is then compared with the original codes to identify the intended command.

The increasing number of alternatives in the design of c-VEP-based BCI systems highlights the need to understand not only which encoding sequences, stimulus patterns, or decoding algorithms perform best individually under specific conditions, but also which combinations of these three elements yield the best overall performance. Previous studies have primarily focused on pairwise comparisons, typically evaluating a novel proposal against a standard approach. However, no study has analyzed the performance of multiple configurations while jointly considering encoding sequences, stimulus patterns, and decoding algorithms within a unified comparison framework. The main objective of the present study is to examine both performance and visual fatigue across all possible combinations derived from a selected subset of sequences, stimulus patterns, and decoding algorithms. This work represents a novel contribution by introducing previously unexplored c-VEP configurations and by providing a comprehensive, multi-perspective evaluation framework for c-VEP-based BCI systems.

MATERIALS AND METHODS

Participants and equipment: Twenty-six healthy volunteers (15 males and 11 females; mean age 24.88 ± 4.96 years) with normal or corrected-to-normal vision participated in the study. EEG signals were recorded using a g.USBamp amplifier (g.tec Medical Engineering GmbH, Austria) at a sampling frequency of 256 Hz. Sixteen active electrodes (F3, Fz, F4, C3, Cz, C4, CPz, P3, Pz, P4, PO7, POz, PO8, Oz, I1, and I2) were positioned according to the International 10-10 System. The ground electrode was placed at AFz, and the reference electrode at the right earlobe (A2). Data acquisition and signal processing were performed on a PC equipped with an Intel Core i7-13700 processor, 32 GB RAM, and an NVIDIA GeForce RTX 3050 GPU. Visual stimulation was presented on a Full HD LED monitor with a refresh rate of 120 Hz.

System configuration: The study was implemented using MEDUSA[®], a software ecosystem for the development and deployment of BCI experiments [9]. The different combinations of encoding sequences, stimulus patterns, and decoding algorithms were configured and evaluated using the c-VEP Speller application (version 5.0.0), which is publicly available at medusabci.com.

The configuration parameters that remained constant throughout the experiment included the number of targets, fixed at 16; the layout, defined as a 4×4 matrix;

and the background, which consisted of a uniform bluish color. EEG signals were consistently preprocessed with a 1-60 Hz bandpass filter and a 50 Hz notch filter. In contrast, the encoding sequence, stimulus pattern, and decoding algorithm were varied according to the experimental condition. To conduct the comparative analysis, a subset of configurations for these three elements was selected based on their relevance in the literature and practical feasibility.

The selected sequences were binary m-sequences and burst codes (see Fig. 1). M-sequences represent the standard approach in c-VEP, whereas burst codes have recently emerged as an alternative to increase VEP amplitude and reduce visual fatigue [5]. M-sequences are characterized by excellent autocorrelation properties, exhibiting a maximum autocorrelation at zero lag and near-zero autocorrelation for shifted versions of the sequence. This property allows the system to be trained using a single base sequence, while additional commands are generated through temporal shifts of that sequence [3]. In this study, a 63-bit m-sequence was generated using a 6th-order base-2 polynomial implemented through a linear feedback shift register. For the 16-target configuration, each command was generated by circularly shifting the base sequence, resulting in an average lag of 3.94 bits between commands. Burst codes consist of short aperiodic visual flashes presented at a slow stimulation rate. Burst onsets derived from a master sequence were pseudo-randomly assigned to the different codes in batches, with the assignment order reshuffled after each distribution cycle until all onsets were allocated [5]. In this study, a set of 16 burst codes was generated, each with a fixed length of 80 bits and containing 3 single-bit bursts. This resulted in a minimum inter-burst interval of 50 ms and a maximum interval of 480 ms, with no temporal overlap between bursts of different codes.

The evaluated stimulus patterns comprised plain, grating, and checkerboard (see Fig. 1). As with the encoding sequences, the plain pattern represents the classical approach, whereas the grating and checkerboard designs correspond to more recent proposals aimed at improving user experience [3]. Plain stimulation encodes each bit of the binary sequence using a single color: bit-0 is represented by black and bit-1 by white, thereby producing maximum contrast [3]. Grating stimulation is implemented using 75 small, randomly oriented Gabor patches consisting of a central white bar surrounded by two parallel black bars on a mid-gray background. In this case, bit-0 corresponds to “off” state in which only the mid-gray background is displayed, whereas bit-1 corresponds to “on” state in which the Gabor patches appear superimposed on the background [5]. Checkerboard stimulation follows a similar on/off logic. However, bit-0 is represented by black background, and bit-1 by the appearance of a 16×16 black-and-white checkerboard pattern [6].

The decoding approaches considered in this study included two TM and two BWR pipelines. The first TM approach is the standard one, here referred as event-related

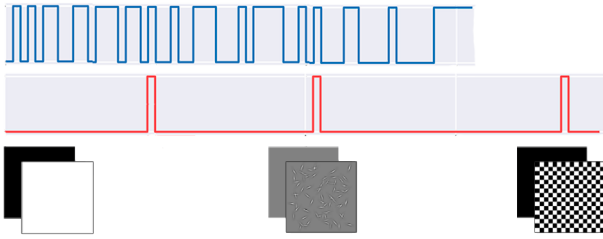


Figure 1: Subset of sequences and stimuli. Sequences (top to bottom): m-sequence and burst code. Stimuli (right to left): plain, grating, checkerboard.

potential template matching (TM-ERP) [3]. For each command, a template of the EEG response is generated. To obtain this template, trials of stimulation are averaged. Then, a canonical correlation analysis (CCA) is performed to maximize the correlation between the averaged and the single-trial signals. The first CCA component is used as a spatial filter [3]. The second TM approach is Reconvolution (TM-Reconvolution) [7]. Here, instead of averaging across trials, EEG response is first decomposed into individual event responses of 300 ms, which capture the brain’s response to elementary visual events such as contrast changes. Then, full responses for each command are reconstructed by convolving these event responses according to the temporal structure of the sequences. The CCA step is subsequently applied to compute a spatial filter, which, when applied to the signals, produces the final template for each command [7]. For the BWR approaches, the first method is Riemann bit-wise reconstruction (BWR-Riemann) [5]. This is a machine-learning approach based on Riemannian geometry. First, xDAWN spatial filtering is applied to enhance the evoked response, after that for each EEG epoch covariance matrices are computed. These matrices are projected into the tangent space of the Riemannian manifold to yield vector features. The features are input to a linear discriminant analysis (LDA) classifier to estimate the probability that a given EEG epoch belongs to a bit-0 or bit-1. The predicted sequence is then correlated to all sequences [5]. The remaining BWR algorithm is the EEG-Inception bit-wise reconstruction (BWR-EEGInception) [8]. A deep learning pipeline based on the EEG-Inception model, a convolutional neural network (CNN) whose effectiveness in c-VEP decoding has been previously demonstrated [8]. EEG signals are downsampled and segmented into short epochs time-locked to each bit. These epochs are processed by the EEG-Inception network, which extracts spatiotemporal features to classify each epoch as a bit-0 or bit-1. The predicted sequence is then correlated to all the sequences [8].

Experimental protocol: To conduct the comparative analyses addressed in this study, an experimental protocol was designed to evaluate 6 distinct c-VEP configurations. These configurations resulted from combining each of the two proposed pseudorandom sequences (m-sequences and burst codes) with each of the three stimulus patterns (plain, grating, and checkerboard). Regarding

the decoding algorithms, all conditions were evaluated online using a single method (TM-ERP for m-sequence conditions and BWR-Riemann for burst code conditions). The remaining decoding algorithms were evaluated offline to avoid excessively increasing the duration of the experimental session.

Each participant completed a single experimental session in which all proposed c-VEP conditions were evaluated, with the order of condition presentation randomized to minimize potential order effects. For each condition, participants completed two calibration runs followed by one online run. Each calibration run consisted of 80 cycles (i.e., complete sequence repetitions). In the m-sequence conditions, the same target was presented throughout, as a single base sequence and its shifted versions were used to encode all commands. In the burst code conditions, the target letter changed every 10 cycles to ensure that all 16 targets were presented, since each command is associated with a distinct code. The online run consisted of 16 selections, one for each target, with 8 cycles per selection. After completing each condition, participants rated their perceived visual fatigue using a 0-10 Likert scale, where 0 indicated no fatigue and 10 indicated extreme fatigue.

Evaluation metrics: The evaluation of the experimental conditions in this study was based on three key aspects: system performance, EEG responses, and user experience. System performance was assessed in terms of accuracy, calculated as the number of correct selections divided by the total number of selections. To characterize EEG responses, calibration data were analyzed. For the m-sequence conditions, the average EEG response across complete cycle presentations was computed. In contrast, for the burst code conditions, each of the 16 commands is associated with a distinct sequence, and responses to individual events exhibit considerably less temporal overlap. Therefore, ERPs were computed time-locked to each burst onset (i.e., bit-1 onset). Finally, user experience was evaluated through the subjective visual fatigue ratings provided by participants in the questionnaires completed during the experimental protocol.

To assess differences between experimental conditions, statistical analyses were conducted on both accuracy and visual fatigue metrics. Since the data were paired and did not follow a normal distribution, pairwise comparisons between conditions were performed using the Wilcoxon signed-rank test. To account for multiple comparisons and control the false discovery rate (FDR), the p -values were adjusted using the Benjamini-Hochberg procedure.

RESULTS

The system performance was computed separately for each of the decoding algorithms selected for comparison. Results were averaged across participants, yielding an accuracy curve for each condition and algorithm. These curves are shown in Fig. 2. The time point closest to 4.2 s was used for statistical comparisons, as this corresponds to the maximum duration of data shared across all condi-

Table 1: Subjective visual fatigue ratings reported by participants for each experimental condition.

Condition	Mean	SD	CLD
burst-grat	2.35	1.96	a
mseq-grat	3.19	1.94	b
burst-check	4.81	1.90	c
burst-plain	4.96	1.99	c
mseq-check	6.00	2.00	d
mseq-plain	7.42	1.96	e

SD: Standard Deviation

CLD: Compact Letter Display

tions, limited by the length of the m-sequence conditions. The corrected p -values from these statistical analyses are also shown in Fig. 2, where the central matrices display the results of all pairwise comparisons.

EEG responses computed for each condition were grand-averaged across participants at Oz, and the resulting mean signals are shown in Fig. 3.

Finally, regarding user experience, the results are presented in Tab. 1. Visual fatigue scores provided by the participants were averaged for each condition and are reported together with their standard deviation (SD). The last column of the table shows the results of the pairwise statistical comparisons. These results are presented using the Compact Letter Display (CLD) format, with a significance level of corrected p -value < 0.05 . According to this format, any two conditions assigned the same CLD letter do not differ statistically significantly, whereas any two conditions labeled with different letters exhibit statistically significant differences [10].

DISCUSSION

System performance: Looking at the accuracy curves in Fig. 2, a high degree of variability can be observed depending on the combination of the three factors considered (sequence, stimulus pattern, and decoding algorithm). The statistical analysis provides great insight into this variability, as numerous statistically significant differences are found among different sequence and stimulus combinations evaluated with the same decoding algorithm. Additionally, certain conditions defined by specific sequences and stimulus patterns are not consistent across different decoding algorithms. In other words, its performance depends strongly on the decoding method applied. These results suggest that the choice of all three factors has a considerable impact on system performance and should therefore be carefully considered. Importantly, not only do these factors influence performance independently, but their interaction must also be jointly taken into account to optimize the configuration of c-VEP-based BCIs.

Focusing on the influence of the pseudorandom sequence choice, it cannot be clearly stated whether m-sequences or burst codes exhibit superior performance, as the results also depend on the stimulus pattern. However, m-sequences tend to achieve higher accuracies overall. Except for the checkerboard stimulus pattern, m-sequences

show higher accuracy curves than their corresponding burst-code conditions across all decoding algorithms.

Regarding stimulus patterns, no single design can be considered universally superior, as performance also depends on the sequence and decoding algorithm. Checkerboard patterns achieve the best overall results when paired with burst codes, whereas plain patterns may outperform checkerboard in combination with m-sequences. One consistent finding across all decoding algorithms is that grating patterns produce the lowest accuracies, regardless of the sequence type. This outcome may be influenced by the monitor refresh rate; previous studies analyzed the performance at 60 Hz [5], whereas the present study used 120 Hz.

Regarding the decoding algorithms, BWR-Riemann achieved the highest and most consistent performance, outperforming the other methods. It was the only algorithm to reach accuracies above 90% for both plain and checkerboard conditions. Moreover, despite the lower performance observed for the grating conditions, BWR-Riemann still outperformed the other algorithms. The remaining decoding methods showed effective performance but just for specific combinations of sequence and stimulus pattern.

Overall, the highest performance was achieved with the combination of burst codes and the checkerboard pattern, which consistently outperformed other conditions across all decoding algorithms, reaching nearly 100% grand-averaged accuracy in less than 4 s of decoding. However, it is important to note that several other combinations also yielded effective performance. The key takeaway is that system optimization should consider the joint effects of all elements, designing the configuration coherently based on the combined impact of sequence, stimulus pattern, and decoding algorithm.

EEG responses: To analyze EEG responses, the m-sequence and burst code conditions need to be considered separately, due to the inherent differences in the temporal structure and characteristics of the two sequence types. For the m-sequence conditions, EEG responses obtained from the calibration data (upper plot in Fig. 3) show that the grating and checkerboard patterns elicit similar neural responses, both in amplitude and waveform morphology, compared to the plain stimulation. In contrast, the plain stimulation produces a stronger neural response in terms of amplitude, but also shows greater variability across participants, as reflected by the wider confidence interval showed in the graph as the shaded region.

For the burst code conditions, EEG responses (lower plot in Fig. 3) resemble classical VEPs, due to the longer intervals between bursts and consequently less overlapping stimulation. In this context, the plain condition resembles a flash VEP, whereas the checkerboard and grating conditions resemble onset/offset VEPs [11]. In the plain condition, the positive peak around 120 ms likely corresponds to the P2 component of the flash VEP [11]. In the checkerboard condition, the negative peak near 120 ms and a positive peak around 150 ms may correspond to the

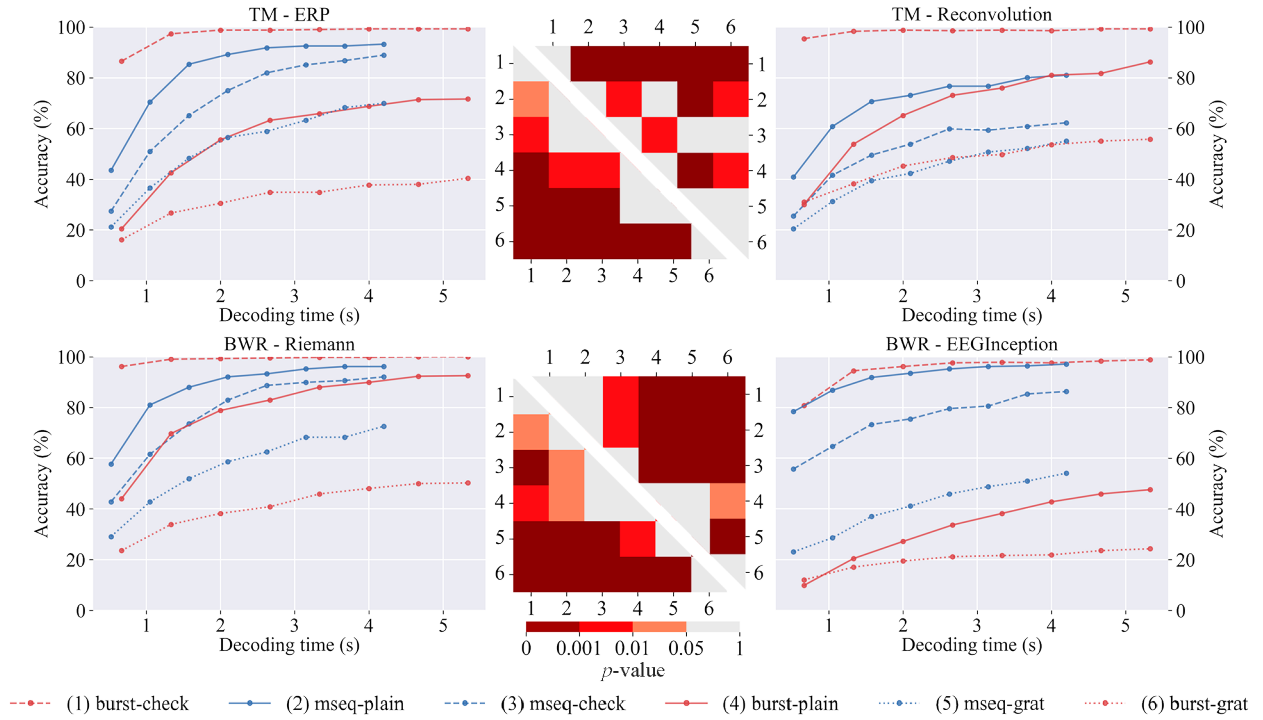


Figure 2: Grand-average accuracy curves over decoding time for each condition and decoding algorithm. The central matrices show corrected p -values from pairwise statistical comparisons. Each matrix is divided by the main diagonal, with the lower triangle corresponding the left-hand decoding algorithms and the upper triangle to those for the right-hand. Conditions are labeled from 1 to 6, where 1 denotes the condition with the highest grand-average accuracy across participants and decoding algorithms, and 6 the opposite.

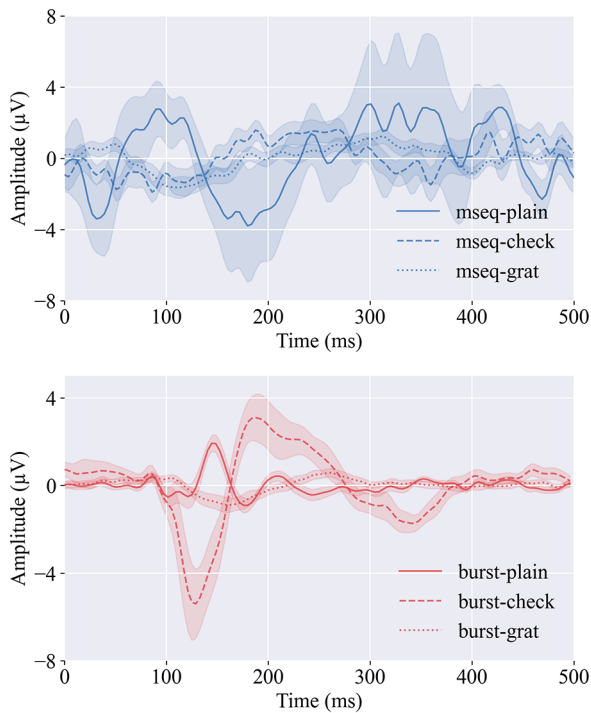


Figure 3: Grand-average EEG responses across users at Oz for each experimental condition using calibration data. Shaded areas represent the 95% confidence interval.

C2 and C3 components of the onset/offset VEP [11]. For the grating condition, these components are also present but they exhibit markedly lower amplitude and a latency delayed by approximately 20 ms. These characteristics could explain the observed differences in system performance: the delayed and weaker response in the grating condition may account for its lower accuracy, while the pronounced amplitude and consistent morphology of the checkerboard may contribute to its superior performance.

User experience: The subjective visual fatigue ratings reported by participants during the experimental protocol reveal a clear ranking of conditions, as shown in Tab. 1. Moreover, almost every step in this ranking corresponds to statistically significant differences. The only exception is the comparison between the plain and checkerboard burst conditions, for which no statistically significant differences were observed.

The conditions that induced the lowest levels of visual fatigue were those using the grating stimulus pattern, suggesting that the stimulus design may be the factor with the greatest impact on user experience. Intermediate fatigue levels were observed for the burst code conditions with plain and checkerboard patterns, whereas the most fatiguing conditions corresponded to the same stimulus patterns combined with m-sequences. Overall, these results indicate that burst codes are generally preferred over m-sequences in terms of sequence type. Regarding stimulus patterns, grating is the most preferred, followed by checkerboard, and finally by plain, which is perceived as

the most fatiguing.

When considering both system performance and user experience, a clear trade-off emerges. The conditions that yielded the lowest performance were those preferred by users in terms of visual fatigue. Therefore, achieving an optimal balance between these two key aspects of c-VEP-based BCIs becomes essential. In this context, the burst code condition with the checkerboard pattern appears to offer the best solution. It provides the highest system performance while remaining among the least fatiguing conditions, only behind the grating conditions, which, despite being more comfortable, showed poor performance.

Limitations and future research: This study offers important contributions to research on c-VEP-based BCIs; however, several limitations should be considered. A larger and more heterogeneous cohort would improve statistical power and generalization of the results. In addition, visual fatigue was assessed through subjective ratings; incorporating EEG biomarkers or eye-tracking metrics, would provide a more objective evaluation.

Additional limitations related to the comparative design must be acknowledged. Given the large number of possible combinations of sequence type, stimulus pattern, and decoding algorithm, only a subset of conditions was evaluated, potentially excluding others. Moreover, differences between sequences led to unequal amounts of online and calibration data; while decoding performance was analyzed as a function of decoding time, its dependence on calibration time remains unstudied.

CONCLUSION

The present research provides a comprehensive comparison of multiple c-VEP-based BCI configurations, combining different pseudorandom sequences, stimulus patterns, and decoding algorithms. Results indicate that the best balance between system performance and user experience is achieved with burst codes combined with the checkerboard pattern and the BWR-Riemann decoding algorithm, yielding a mean accuracy of 99.76% and a mean fatigue score of 4.81 ± 1.90 . However, findings also showed that a wide range of parameter combinations can yield reliable performance, and that the optimal choice may depend on the specific research or application requirements. Importantly, performance does not depend on any single factor alone but on the interaction between sequence type, stimulus pattern, and decoding algorithm. When a particular sequence is chosen, the most suitable stimulus pattern or compatible decoding algorithm may vary considerably. This interaction-driven behavior underscores the importance of considering these parameters jointly rather than independently when designing c-VEP-based BCI systems.

ACKNOWLEDGMENTS

This work was supported by 0124_EUROAGE_MAS_4_E project, co-financed

by the EU through the Interreg VI-A Spain-Portugal Programme (POCTEP) 2021-2027, and by the VA140P24 project, funded by the Regional Government of Castilla y León (Consejería de Educación) and EU-ERDF. Support has also been provided by the Centre for Biomedical Research in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN, CB19/01/00012) through the Carlos III Health Institute, co-financed with ERDF funds. Ana Martín-Fernández is a beneficiary of a PIF-UVa grant.

REFERENCES

- [1] Wolpaw J, Wolpaw EW. Brain-computer interfaces: Principles and practice. Oxford University Press (2012).
- [2] Min BK, Marzelli MJ, Yoo SS. Neuroimaging-based approaches in the brain-computer interface. Trends in Biotechnology. 2010;28(11):552–560.
- [3] Martínez-Cagigal V, Thielen J, Santamaría-Vázquez E, Pérez-Velasco S, Desain P, Hornero R. Brain-computer interfaces based on code-modulated visual evoked potentials (c-VEP): A literature review. Journal of Neural Engineering. 2021;18(6):061002.
- [4] Zheng L, Dong Y, Tian S, Pei W, Gao X, Wang Y. A calibration-free c-VEP based BCI employing narrow-band random sequences. Journal of Neural Engineering. 2024;21(2):026023.
- [5] Dehais F, Castillos KC, Ladouce S, Clisson P. Leveraging textured flickers: A leap toward practical, visually comfortable, and high-performance dry EEG code-VEP BCI. Journal of Neural Engineering. 2024;21(6):066023.
- [6] Fernández-Rodríguez Á, Martínez-Cagigal V, Santamaría-Vázquez E, Ron-Angevin R, Hornero R. Influence of spatial frequency in visual stimuli for cVEP-based BCIs: Evaluation of performance and user experience. Frontiers in Human Neuroscience. 2023;17:1288438.
- [7] Thielen J, Broek Pvd, Farquhar J, Desain P. Broad-band visually evoked potentials: Re(con)volution in brain-computer interfacing. PLOS ONE. 2015;10(7):e0133797.
- [8] Santamaría-Vázquez E, Martínez-Cagigal V, Hornero R. Bit-wise reconstruction of non-binary visual stimulation patterns from EEG using deep learning: A promising alternative for user-friendly high-speed c-VEP-based BCIs. In: Advances in Computational Intelligence. Switzerland, 2023, 603–614.
- [9] Santamaría-Vázquez E et al. MEDUSA©: A novel python-based software ecosystem to accelerate brain-computer interface and cognitive neuroscience research. Computer Methods and Programs in Biomedicine. 2023;230:107357.
- [10] Piepho HP. An algorithm for a letter-based representation of all-pairwise comparisons. Journal of Computational and Graphical Statistics. 2004;13(2):456–466.
- [11] Odom JV et al. ISCEV standard for clinical visual evoked potentials: (2016 update). Documenta Ophthalmologica. 2016;133(1):1–9.