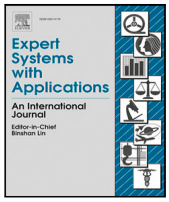
Expert Systems With Applications 232 (2023) 120815

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



Non-binary m-sequences for more comfortable brain–computer interfaces based on c-VEPs

Víctor Martínez-Cagigal [a,b,*], Eduardo Santamaría-Vázquez [a,b], Sergio Pérez-Velasco [a,b], Diego Marcos-Martínez [a,b], Selene Moreno-Calderón [a], Roberto Hornero [a,b]

[a] Biomedical Engineering Group, University of Valladolid, ETSIT, Paseo de Belén, 15, 47011, Valladolid, Spain
[b] Centro de Investigación Biomédica en Red de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Valladolid, Spain

ARTICLE INFO

Keywords:
Non-binary codes
Visual fatigue
Code-modulated visual evoked potential (c-VEP)
Brain–computer interface (BCI)
Electroencephalography (EEG)

ABSTRACT

Background and Objectives: Code-modulated visual evoked potentials (c-VEPs) have marked a milestone in the scientific literature due to their ability to achieve reliable, high-speed brain–computer interfaces (BCIs) for communication and control. Generally, these expert systems rely on encoding each command with shifted versions of binary pseudorandom sequences, i.e., flashing black and white targets according to the shifted code. Despite the excellent results in terms of accuracy and selection time, these high-contrast stimuli cause eyestrain for some users. In this work, we propose the use of non-binary *p*-ary m-sequences, whose levels are encoded with different shades of gray, as a more pleasant alternative than traditional binary codes. The performance and visual fatigue of these *p*-ary m-sequences, as well as their ability to provide reliable c-VEP-based BCIs, are analyzed for the first time.

Methods: Five different m-sequences were evaluated with 16 healthy participants, following the circular shifting paradigm: base 2 (63 bits), base 3 (80 bits), base 5 (124 bits), base 7 (48 bits), and base 11 (120 bits). Signal processing consisted of a 3-filter bank (1–60 Hz, 12–60 Hz and 30–60 Hz), followed by a canonical correlation analysis. Raster latency correction and artifact rejection approaches were also applied to compute command templates. For each m-sequence, users performed a 30-trial calibration stage, followed by an online spelling task of 32 trials. In addition, qualitative measures regarding visual fatigue and satisfaction were collected.

Results: Users were able to achieve an average accuracy of over 98% for all *p*-ary m-sequences. The differences between m-sequences were not significant in terms of accuracy, but they were in terms of visual fatigue. The higher the base, the less eyestrain perceived by users for both presentation rates of 60 Hz and 120 Hz. All *p*-ary m-sequences were also significantly less annoying when displayed at 120 Hz compared to 60 Hz.

Conclusion: Results suggest that all *p*-ary m-sequences are suitable for achieving high speed and high accuracy in c-VEP-based BCIs, reducing the visual fatigue as the base increases, without degrading system performance. It is concluded that the use of high presentation rates and non-binary m-sequences is a promising alternative to provide user-friendly c-VEP-based BCIs.

1. Introduction

Brain–computer interfaces (BCIs) have shown their ability to replace or even restore central nervous system outputs by providing a direct pathway between our brain activity and external devices (Wolpaw & Wolpaw, 2012). In particular, BCIs that monitor the electroencephalographic (EEG) signals of the user have traditionally been perceived as potential alternative and augmentative communication (AAC) technologies for populations with motor disabilities due to their non-invasiveness, ease of use, relative low cost and portability (Wolpaw & Wolpaw, 2012). From an expert system's point of view, BCIs employ signal processing and pattern recognition techniques to decode the user's intentions in real-time and translate them into commands that control applications or external devices (Wolpaw & Wolpaw, 2012).

Over the last decades, the efforts of many research groups have been devoted to gradually improving BCIs for communication and control by maximizing performance and user comfort, with the ultimate goal of becoming a plug-and-play technology. Control signals such as P300 potentials or steady-state visual evoked potentials (SSVEP)

* Corresponding author at: Biomedical Engineering Group, University of Valladolid, ETSIT, Paseo de Belén, 15, 47011, Valladolid, Spain.
E-mail addresses: victor.martinez.cagigal@uva.es (V. Martínez-Cagigal), eduardo.santamaria.vazquez@uva.es (E. Santamaría-Vázquez), sergio.perezv@uva.es (S. Pérez-Velasco), diego.marcos.martinez@uva.es (D. Marcos-Martínez), selene.moreno@uva.es (S. Moreno-Calderón), roberto.hornero@uva.es (R. Hornero).

https://doi.org/10.1016/j.eswa.2023.120815
Received 25 October 2022; Received in revised form 27 May 2023; Accepted 9 June 2023
Available online 15 June 2023

have been traditionally used to implement BCI spellers, reaching suitable performances in terms of accuracy and information transfer rate (ITR) (Wolpaw & Wolpaw, 2012).

P300-based BCIs rely on visual *oddball* paradigms to elicit P300 components as a response to unexpected target stimuli (Martínez-Cagigal et al., 2017, 2019). Despite having no restrictions regarding the limit of possible commands, P300-based BCIs require multiple repetitions of the same stimulus to obtain a reliable potential, as well as a calibration stage that may last around half an hour (Martínez-Cagigal et al., 2017, 2019). SSVEP-based BCIs, on the other hand, do not necessarily require a calibration stage as they rely on the generation of an oscillatory response in the EEG that mimics the frequency of the target command, which flickers at a constant frequency (Vialatte et al., 2010). In this case, the number of commands is limited by the monitor's refresh rate, and the decoding over certain bands such as beta can be challenging (Volosyak et al., 2011). Despite their limitations, these spellers have repeatedly shown their feasibility for healthy users (HU, P300: >90%, >25 bpm (Martínez-Cagigal et al., 2017, 2019; Santamaría-Vázquez et al., 2019), SSVEP: >90%, > 40 bpm (Bin et al., 2009; Gembler, Stawicki, Saboor, & Volosyak, 2019)) and motor-disabled users (MDU, P300: >80%, 10–25 bpm (Martínez-Cagigal et al., 2017, 2019; Santamaría-Vázquez et al., 2020), SSVEP: >80%, 10–40 bpm (Combaz et al., 2013; Peters et al., 2020)).

In this context, code-modulated visual evoked potentials (c-VEP) were recently proposed as an alternative control signal to the traditional P300 or SSVEP approaches (Bin et al., 2009; Sutter, 1992). In the circular shifting paradigm, commands flicker following uncorrelated shifted versions of a binary pseudorandom code. The identification of the desired command is determined in real-time by analyzing the correlation between the EEG response and these shifted templates (Martínez-Cagigal et al., 2021). BCI spellers based on c-VEP have reached excellent performances for both HU (e.g., 91%, 92.8 bpm (Bin et al., 2009), 94%, 92.7 bpm (Gembler, Rezeika, et al., 2020)) and MDU (79.3%, 20.3 bpm (Verbaarschot et al., 2021)) with reduced calibration times (1–3 min Gembler, Rezeika, et al. (2020), Verbaarschot et al. (2021)) or even without calibration (Thielen et al., 2021). The advantages of c-VEP-based BCIs over P300-based BCIs are clear: less calibration time, faster command selection and similar or even better accuracy (Martínez-Cagigal et al., 2021). Although accuracy and speed are comparable to SSVEP-based spellers, c-VEP are less sensitive to non-related basal EEG activity and usually less restrictive in terms of the number of possible commands than SSVEP (Martínez-Cagigal et al., 2021). For these reasons, it is not surprising than c-VEP are establishing themselves as a suitable control signal for reliable, high-speed state-of-the-art BCIs.

Future plug-and-play BCI systems should ideally be non-invasive, reliable and comfortable for users. Although most BCI studies have focused on optimizing the performance of EEG-based BCIs, the scientific community has paid less attention to ease of use and user comfort. In fact, some researchers believe that state-of-the-art c-VEP-based BCIs have already reached such reliability and robustness to make the leap from laboratories to real environments, but there is still room for improvement in terms of user convenience (Martínez-Cagigal et al., 2021). In this regard, it has been reported that c-VEP-based BCIs using binary pseudorandom codes (i.e., encoding commands using black and white flashes) may cause eyestrain for some users, especially for low refresh rates (Martínez-Cagigal et al., 2021). Therefore, making more efforts to reduce user discomfort caused by the flickering is crucial for the advancement of real-life BCI applications based on c-VEPs.

User convenience for c-VEP-based BCIs has been studied from several perspectives (Martínez-Cagigal et al., 2021). On the one hand, the scientific literature agrees that higher stimulation rates are less fatiguing for the user (Başaklar et al., 2019; Gembler et al., 2018; Gembler, Stawicki, Rezeika, & Volosyak, 2019; Nezamfar et al., 2011, 2016). Although most of the c-VEP studies use presentation rates of 60 Hz (i.e., the standard refresh rate for monitors), many studies suggest using 120 Hz as it provides shorter selection times while maintaining an excellent accuracy (Başaklar et al., 2019; Gembler et al., 2018; Wittevrongel et al., 2017). However, increasing the presentation rate is not a definitive solution, as several studies reported significant decreases in performance for rates above 120 Hz (Başaklar et al., 2019; Gembler et al., 2018), where templates start to become less orthogonal to each other. This could be explained because the cones are less responsive to visual stimuli and because the visual system is nonlinear, so EEG responses maintain the same spectral distribution even if the presentation rate increases (Martínez-Cagigal et al., 2021). On the other hand, some studies proposed alternative hand-crafted binary codes that confine spectral density to high-frequency bands under the premise of being more comfortable for the user, such as chaotic codes (Shirzhiyan et al., 2019), 6-target optimum sequences (Behboodi et al., 2020), or superposition optimized pulses (Yasinzai & Ider, 2020). More recently, Gembler, Rezeika, et al. (2020) proposed abandoning the high-contrast stimuli required by binary codes and using a quintary m-sequence (i.e., an m-sequence with 5 different values) encoded with different shades of gray. Binary and quintary m-sequences reached similar performances (99.4% vs. 98.5% at 60 Hz; 97.6% vs. 97.5% at 120 Hz, 97.9% vs. 97.6% at 240 Hz), but the latter was significantly less annoying for users, especially for the 60 Hz presentation rate (Gembler, Rezeika, et al., 2020). Nevertheless, to the best of our knowledge, there are no previous studies that have analyzed how different p-ary bases (apart from quintary, i.e. $p = 5$) affect user comfort and system performance. The influence of the base on users' c-VEP responses has also not yet been studied, nor if there are differences between visual fatigue perceived by different p-ary m-sequences.

The objective of this study is to quantitatively and qualitatively analyze the influence of p-ary m-sequences on (1) the performance of c-VEP-based BCIs and (2) perceived visual fatigue by users. For the first time, five different p-ary m-sequences will be used at a presentation rate of 120 Hz: base 2 (i.e., binary, 63-bit), base 3 (80-bit), base 5 (124-bit), base 7 (48-bit), and base 11 (120-bit). How do the new event levels (bases) affect the orthogonality of templates? Are p-ary m-sequences adequate to achieve reliable BCIs? Do users perceive different eyestrain by varying the base of the m-sequence? This work aims to answer these questions, which we consider necessary to advance the current state-of-the-art towards more reliable and comfortable c-VEP-based BCIs for the end user.

2. Subjects and signals

A total of 16 healthy users participated in this study (mean age: 28.8 ± 5.0 years old, 11 males, 5 females). All of them gave their informed consent to participate in this study, previously approved by the local ethics committee. A *g.USBamp* amplifier (g.Tec, *Guger Technologies*, Austria) with 16 channels and a sampling rate of 256 Hz was used to record EEG signals. According to the International System 10-10 (Oostenveld & Praamstra, 2001), electrodes were placed on F3, Fz, F4, C3, Cz, C4, CPz, P3, Pz, P4, PO7, PO8, Oz, I1 and I2; grounded on AFz and referenced to the earlobe. The equipment was attached to a PC Intel Core i7-7700 @ 3.6 GHz, 32 RAM. MEDUSA© (www.medusabci.com), a Python-based general-purpose software ecosystem to develop BCIs and neuroscience experiments, was used to monitor the data via lab streaming layer (LSL), display the paradigm and process the stimuli in real-time (Santamaría-Vázquez et al., 2023). The proposed "*P*-ary c-VEP Speller" was developed in an Unity-based app that communicates with MEDUSA© via TCP/IP sockets to ensure exact synchronization between stimuli onsets and EEG registering. This app was displayed in a LED FullHD @ 144 Hz monitor, capable of running the experiment at a presentation rate of 120 Hz. Moreover, the refresh rate was monitored using an external phototransistor. To foster open science, the "*P*-ary c-VEP Speller" app for MEDUSA© is publicly available in the app market (www.medusabci.com/market/pary_cvep) (Santamaría-Vázquez et al., 2023).

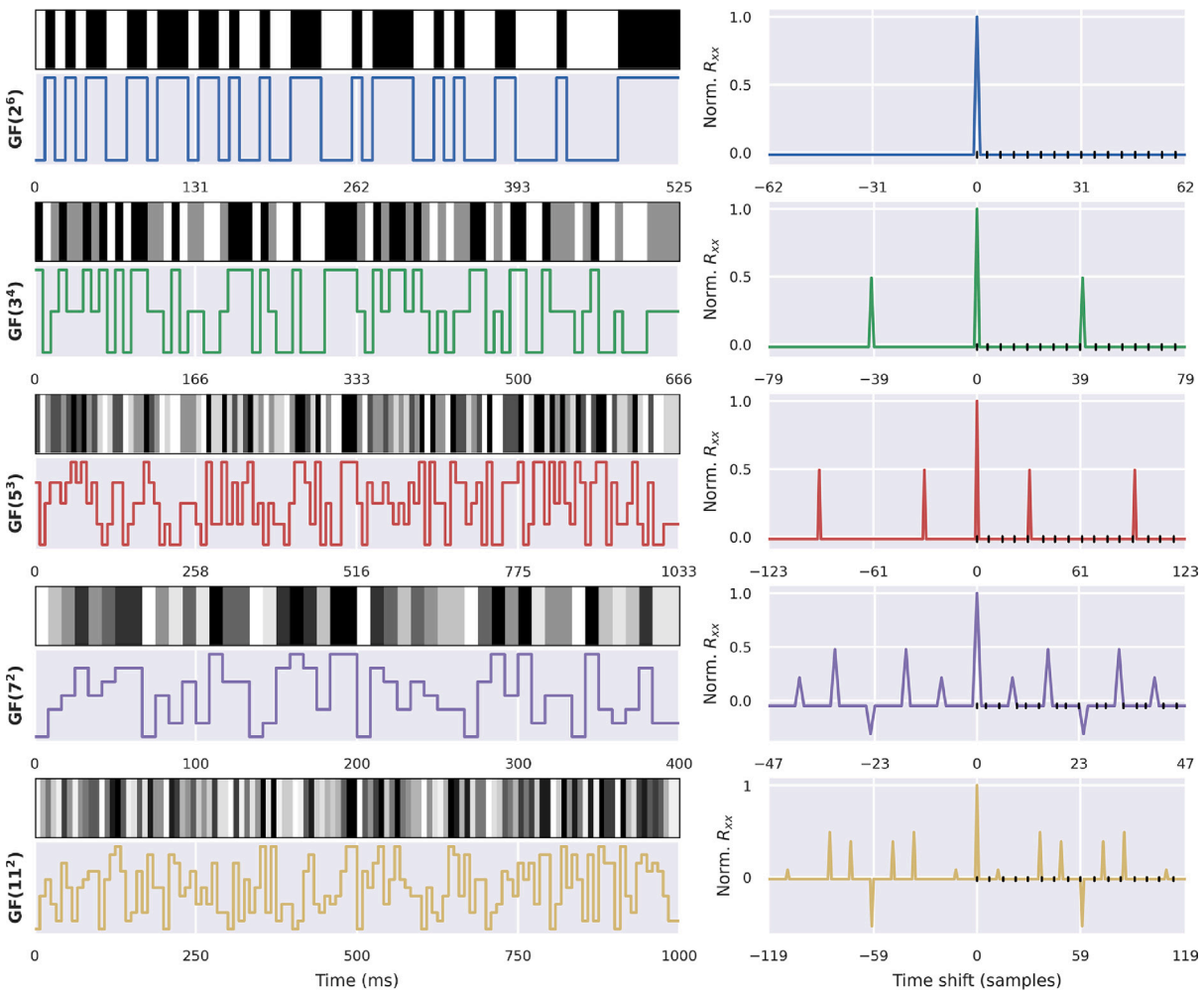


Fig. 1. *P*-ary m-sequences used in this study, from top to bottom: 63-bit GF(2^6), 80-bit GF(3^4), 124-bit GF(5^3), 48-bit GF(7^2), and 120-bit GF(11^2). For each one, visual encoding (top left), original sequence (bottom left) and autocorrelation function (right) are shown for a single cycle at a presentation rate of 120 Hz. The periodic autocorrelation function includes the selected relative shift positions (black points) to encode the 16 different commands. Occasional local minima and maxima correlations were avoided using the deterministic algorithm detailed in Appendix A.

3. Methods

3.1. P-ary m-sequence generation

Maximal length sequences (i.e., m-sequences) are pseudorandom periodic time series that are almost orthogonal to circularly shifted versions of themselves, generated through linear-feedback shift registers (LFSRs) (Martínez-Cagigal et al., 2021). LFSRs are shift registers that compute new values using a linear function of the immediate previous state (Martínez-Cagigal et al., 2021). They are determined by: (1) the base p, i.e. number of different levels or events (e.g., for binary m-sequences, $p = 2$); (2) the order r, i.e. the number of LFSR taps; and (3) the arrangement of taps, which can be expressed as a polynomial whose coefficients are bounded on a Galois Field with p elements, GF(p) (e.g., for $p = 2$, coefficients can be either 0 or 1) (Buračas & Boynton, 2002). Thus, the next bit is calculated using matrix product between the current state and the polynomial, and subsequently applying the modulo p operation (Buračas & Boynton, 2002). Of note, a necessary and sufficient condition for generating m-sequences is that the polynomial must have maximum degree r and be primitive of GF(p^r) (Buračas & Boynton, 2002).

By definition, m-sequences have several restrictions with respect to their length and base. On the one hand, an m-sequence cannot be of arbitrary length, but instead consists of exactly $N = p^r - 1$ bits, and it is repeated cyclically (Martínez-Cagigal et al., 2021). On the other hand, only finite fields in which p is a prime number satisfy the operations of multiplication, addition, subtraction, and division are eligible to generate m-sequences. This implies that (1) the base p must be a prime number, and (2) the p-ary m-sequence can be generated only if a primitive polynomial of degree r over GF(p^r) exists (Buračas & Boynton, 2002). Although primitive polynomials for bases 2, 3 or 5 are common in the literature (Mullen & Panario, 2013), finding polynomials for higher bases is not a trivial task, often requiring the use of complex linear numerical algebra algorithms (Di Porto et al., 1992).

Table 1 shows the generation details of the p-ary m-sequences employed in this study: GF(2^6), GF(3^4), GF(5^3), GF(7^2), and GF(11^2). The computed p-ary m-sequences are shown in Fig. 1, together with the color encoding and their circular autocorrelation functions. As shown, although binary m-sequences always deviate from orthogonality by $-1/N$, non-binary m-sequences (i.e., $p > 2$) present occasional phase shifts in which they are highly anti-/correlated (Buračas & Boynton, 2002). As BCI commands are encoded with shifted versions of the m-sequences, these phase points have been avoided using a deterministic algorithm, whose pseudocode is detailed in the Appendix A. The relative lags associated for the 16 commands are also shown in Fig. 1. Another interesting property of these codes is that combinations of event subsequences are nearly perfectly counterbalanced up to the length equal to the order r; i.e., a single cycle contains all possible

Table 1
Generation details of the p-ary m-sequences.

	Base	Order	Length (bits)	Polynomial	Duration (s/cycle)	
					120 Hz	60 Hz
GF(2^6)	2	6	63	$x^6 + x^5 + 1$	0.525	1.050
GF(3^4)	3	4	80	$x^4 + 2x^3 + 1$	0.667	1.333
GF(5^3)	5	3	124	$3x^3 + 2x^2 + 1$	1.033	2.067
GF(7^2)	7	2	48	$4x^2 + 1$	0.408	0.800
GF(11^2)	11	2	120	$3x^2 + x + 1$	1.000	2.000

combination of elements except the full-zero one (Buračas & Boynton, 2002).

3.2. Signal processing

The signal processing stage is mainly based on the reference processing pipeline for the c-VEP circular shifting paradigm (refer to Martínez-Cagigal et al. (2021) for further details). EEG signals were pre-processed in real-time using a notch filtering at 50 Hz to remove power line interference; followed by a filter bank composed of 3 bandpass filters between 1–60 Hz, 12–60 Hz and 30–60 Hz (Martínez-Cagigal et al., 2021). All of them were 7-th order infinite impulse response (IIR) Butterworth filters to ensure real-time processing. The underlying concept of this approach is to enhance the separation between natural brain activity (e.g., alpha band activity, associated with tiredness) and stimuli-induced responses (Gembler, Benda, et al., 2020). Thus, the initial filter (1–60 Hz) encompasses all frequency bands, whereas the second filter (12–60 Hz) targets the beta and gamma bands, and the third filter (30–60 Hz) exclusively focuses on the gamma band. The highest cutoff of 60 Hz is established owing to the maximum frequency that can be reflected in the EEG by a 120 Hz monitor refresh rate (i.e., equivalent to the encoding 10101010 …) (Martínez-Cagigal et al., 2021). Given that none of our m-sequences exhibit transitions from 0 to 1 (or vice versa) within a single bit, 60 Hz would be an unreachable upper limit for our m-sequences. The lower cutoff of 1 Hz in the first filter is set to take into account delta and theta bands, which also contain information regarding the stimuli (e.g., the constant repetition of GF(2^6) cycles produces a frequency of 1.90 Hz and its harmonics) (Martínez-Cagigal et al., 2021). Afterward, canonical correlation analysis (CCA) is used in each trial to decode the target command the user wants to select in real-time.

During the calibration stage, the user must look to a command encoded with the original p-ary m-sequence (i.e., without any lag) for k cycles (i.e., repetitions of the m-sequence), obtaining the pre-processed EEG signal $\boldsymbol{X} \in \mathbb{R}^{N_f,k,N_s,N_c}$, where $N_f = 3$ is the number of filters in the filter bank, N_s is the number of samples, and N_c is the number of channels. Of note, we have $N_s = \lceil f_s * N/f_m \rceil$ samples per cycle, where $f_s = 256$ Hz is the EEG sampling rate, N is the length of the m-sequence, and $f_m = 120$ Hz is the presentation rate. For each filter f in the bank, a multi-channel response $\hat{\boldsymbol{X}}_f \in \mathbb{R}^{N_s,N_c}$ is obtained by averaging across cycles. Then, CCA is applied to find the linear projections $\boldsymbol{W}_a, \boldsymbol{W}_b$ that maximize the correlation between projected versions of two signals $\boldsymbol{A}$ and $\boldsymbol{B}$, by optimizing:

$$\max_{\boldsymbol{W}_a,\boldsymbol{W}_b} \frac{\boldsymbol{W}_a^T \boldsymbol{A}\boldsymbol{B}^T \boldsymbol{W}_b^T}{\sqrt{\boldsymbol{W}_a^T \boldsymbol{A}\boldsymbol{A}^T \boldsymbol{W}_a \cdot \boldsymbol{W}_b^T \boldsymbol{B}\boldsymbol{B}^T \boldsymbol{W}_b}}. \quad (1)$$

In this case, $\boldsymbol{A} \in \mathbb{R}^{kN_s,N_c}$ is the concatenated version of $\boldsymbol{X}$ for a given filter, whereas $\boldsymbol{B} \in \mathbb{R}^{kN_s,N_c}$ is $\hat{\boldsymbol{X}}_f$ duplicated k times to match the dimensions. After training CCA, spatial filters $\boldsymbol{W}_a \in \mathbb{R}^{N_c,N_c}$ and $\boldsymbol{W}_b \in \mathbb{R}^{N_c,N_c}$ are obtained. However, only $\boldsymbol{w}_b$, the first column of $\boldsymbol{W}_b$, is used afterward to project the averaged response (i.e., $\hat{\boldsymbol{X}}_f \cdot \boldsymbol{w}_b$) and get the main template $\tilde{\boldsymbol{x}}_{f0} \in \mathbb{R}^{N_s,1}$. Templates for the rest of the commands $\tilde{\boldsymbol{x}}_{fi}$ are computed by circularly shifting $\tilde{\boldsymbol{x}}_{f0}$ according to the lag associated to each command θ_i. This procedure is repeated for each filter f, so we end up with $N_f = 3$ different sets of templates.

To maximize performance, we have also corrected the 'raster latencies', as suggested by Nagel et al. (2018). Since pixel lines are refreshed from top to bottom, a command placed at the bottom of the screen will have additional latency compared to a command placed at the top. We have measured the delay between the first top line and the last bottom line, obtaining 7.92 ms (approximately 95% of the refresh rate) (Nagel et al., 2018). In order to enhance the estimation of the calibrated templates, we have delayed each according to the raster latency of the commands they belong to. Furthermore, we have also taken care of powerful non-stationary artifacts that can worsen model performance, such as blinking or electrode-pops. During calibration, the standard deviation of $\boldsymbol{A}$ for each channel is computed, $\sigma_{\boldsymbol{A}}$. Artifacts are detected in a cycle if the standard deviation of that epoch is 3 times greater than $\sigma_{\boldsymbol{A}}$. Only epochs that do not have artifacts on any channel are selected to be used in model training.

In online mode, whenever a trial ends, the same pre-processing stage is applied to obtain $\boldsymbol{Z}_{test} \in \mathbb{R}^{N_f,k_t,N_s,N_c}$, where k_t is the number of cycles of the test trial. For each filter f, the signal is averaged across cycles and projected using the trained $\boldsymbol{w}_b$, obtaining the test response $\tilde{\boldsymbol{z}}_f \in \mathbb{R}^{N_s,1}$. The response is compared with all the templates of the filter f, returning a vector $\boldsymbol{\rho}_f$ containing the Pearson's correlation coefficients for each command. Finally, correlations are averaged across the filter bank and the selected command is the one that reached the maximum coefficient, i.e., $y = \arg\max_i \sum_f \boldsymbol{\rho}_f$. The code of this signal processing pipeline is publicly available in MEDUSA© Kernel as a PyPI package (pypi.org/project/medusa-kernel) (Santamaría-Vázquez et al., 2023).

3.3. Evaluation protocol

Participants carried out a single evaluation session, composed by EEG recordings and subjective questionnaires. First, users were asked to rate the subjective eyestrain produced by the p-ary m-sequences. A total of 5 commands were displayed on the screen, each one encoded with a different p-ary m-sequence. Users were asked to pay attention to each of the commands and rate from 0 to 10 the level of visual fatigue produced by each stimulus. This process was repeated twice, using presentation rates of 60 Hz and 120 Hz.

Afterward, users performed 5 blocks of 32 trials each with the c-VEP speller using a presentation rate of 120 Hz. In each block, a different m-sequence was calibrated and used to complete spelling tasks. Fig. 2 shows two snapshots of the calibration and evaluation stages. Trials were made up of 10 stimulation cycles, the duration of which varied as a function of the m-sequence (see Table 1). In the calibration stage, users were asked to pay attention to a single command encoded with the selected p-ary m-sequence (without lag) during 6 runs of 5 trials each, so 300 cycles were recorded. After training the decoding model, users were asked to perform an online spelling task that consisted of selecting each of the 16 commands in alphabetical order twice. Thus, a total of 320 test cycles were recorded. To avoid bias, the order of the blocks was randomized across participants. Furthermore, after each block, users were asked to re-rate the relative eyestrain of the p-ary m-sequence. Of note, users were unaware of which specific p-ary m-sequence was displayed to avoid unintentional biases. The analysis prior to the EEG recording, in which all p-ary m-sequences are shown together, along with the eyestrain rating after each block, will allow us to discern between the perception of eyestrain produced by the stimuli and the fatigue produced as the session progresses.

After the EEG recordings, users were asked to complete a satisfaction questionnaire, in which they had to assess aspects such as the speed of the MEDUSA© app, interface, accessibility, motivation, duration of the session or their expectation, among others. The survey consisted of 10 items following a 5-point Likert scale. Positive and negative statements were shuffled to avoid acquiescence bias (Martínez-Cagigal et al., 2019). Moreover, an open question was added at the end of the survey to collect personal suggestions for future improvements.

Table 2
Grand-averaged online results across users.

	No. cycles	1	2	3	4	5	6	7	8	9	10	Mean ± STD
GF(2^6)	Accuracy (%)	65.43	90.43	96.68	98.05	**99.41**	**99.41**	**99.61**	**99.61**	**99.41**	**99.61**	94.77 ± 10.14
	ITR (bpm)	208.06	188.03	141.04	108.91	90.14	75.11	64.65	56.57	50.08	45.25	102.78 ± 55.19
	Duration (s)	0.53	1.05	1.58	2.10	2.62	3.15	3.68	4.20	4.73	5.25	2.89 ± 1.51
GF(3^4)	Accuracy (%)	75.78	92.77	97.27	97.66	98.44	98.83	99.02	99.02	99.02	99.22	95.70 ± 6.89
	ITR (bpm)	216.59	154.70	113.27	85.68	69.67	58.54	50.31	44.02	39.13	35.36	86.73 ± 55.97
	Duration (s)	0.67	1.33	2.00	2.67	3.33	4.00	4.67	5.33	6.00	6.67	3.67 ± 1.91
GF(5^3)	Accuracy (%)	83.20	94.14	**98.05**	**98.83**	99.02	99.22	99.41	99.41	99.22	99.22	96.97 ± 4.83
	ITR (bpm)	163.72	103.05	74.01	56.54	45.47	38.02	32.76	28.66	25.39	22.85	59.05 ± 42.20
	Duration (s)	1.03	2.07	3.10	4.13	5.17	6.20	7.23	8.27	9.30	10.33	5.68 ± 2.97
GF(7^2)	Accuracy (%)	49.80	82.23	93.36	95.90	97.27	97.46	98.05	97.66	98.24	98.05	90.80 ± 14.42
	ITR (bpm)	168.14	205.30	174.38	137.45	113.43	95.18	82.46	71.65	64.31	57.72	117.00 ± 49.15
	Duration (s)	0.40	0.80	1.20	1.60	2.00	2.40	2.80	3.20	3.60	4.00	2.20 ± 1.15
GF(11^2)	Accuracy (%)	**86.72**	**97.46**	**98.05**	98.63	98.83	98.63	98.63	98.63	98.63	98.44	**97.27 ± 3.54**
	ITR (bpm)	179.64	113.79	76.79	58.36	46.93	38.99	33.42	29.24	25.99	23.33	62.65 ± 47.10
	Duration (s)	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	5.50 ± 2.87

Note: Unfolded results for each user are available in the supplementary material. STD: standard deviation.

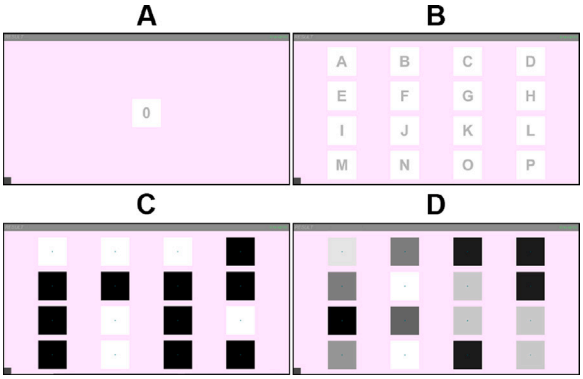


Fig. 2. Snapshots of the "*P*-ary c-VEP Speller" app. (A) Calibration stage, in which users are asked to pay attention to the central command, which flashes according to the original *p*-ary m-sequence. (B) Online arrangement, in which users pay attention to the target command to perform spelling tasks. (C) During online mode, command labels are replaced with a small centered dot that helps users focus on the desired command. In this snapshot, GF(2^6) was displayed. (D) Snapshot example of a stimulation using GF(11^2).

4. Results

Fig. 3 depicts the obtained participants' brain responses to each of the evaluated *p*-ary m-sequences, as well as their normalized circular autocorrelation functions. The number of mistakes made for each encoded lags are also shown to detect conflicting time shifts. The average standard deviation of the autocorrelation functions was 15.3%, 16.8%, 12.8%, 20.7%, and 11.9%; for GF(2^6), GF(3^4), GF(5^3), GF(7^2), and GF(11^2), respectively. Table 2 shows the grand-averaged online results across users, detailing the reached accuracy, ITR and trial duration as a function of the number of cycles per trial. The ITR, measured in bits per minute, was calculated using the original formula proposed by Wolpaw et al. (2002):

$$\text{ITR (bpm)} = Q \cdot \left[\log_2(S) + P\log_2(P) + (1-P)\log_2\left(\frac{1-P}{S-1}\right)\right], \qquad (2)$$

where Q denotes the number of selections per minute, S represents the number of commands (here $S = 16$), and P is the classification accuracy. Users reached more than 98% accuracy for all the *p*-ary m-sequences when using 10 cycles, being GF(2^6) the one that obtained highest accuracy, whereas GF(11^2) reached the highest average accuracy across cycles. Unfolded results for each participant can be found in the supplementary material.

To provide a fair comparison between *p*-ary m-sequences, Fig. 4 shows an interpolated accuracy progression as a function of the trial duration, rather than number of cycles. Wilcoxon signed-rank tests were used to perform the statistical analysis, then the *p*-values were corrected for false discovery rate (FDR) using the Benjamini–Hochberg procedure. Only the comparisons GF(5^3) vs. GF(2^6) and GF(5^3) vs. GF(7^2) yielded significant differences (p-value < 0.05) for early time windows, which are shown in Fig. 4(B). In these time windows, GF(2^6) and GF(7^2) reached significantly higher accuracy than GF(5^3). No significant differences were found between the accuracy progression of the rest of the comparisons.

Results of the qualitative analyses are summarized in Fig. 5. Again, Wilcoxon signed-rank tests and FDR corrections were used to perform the statistical analysis. On the one hand, normalized ratings (between 0 and 10) of the questionnaire are grouped in relevant aspects such as session motivation and tiredness; ease of use, utility, and speed; see Fig. 5(A). Users highlighted the speed (9.4) and ease of use (8.0) of the application, and reported that they felt motivated (9.3) to carry out the evaluation session, as well as similar studies in the future. They also perceived the usefulness (8.4) of this app as an adequate AAC for the daily life of MDUs. Although above average, the lowest valued aspect (6.2) was the tiredness produced throughout the evaluation, either by visual stimuli or by the duration of the session (i.e., 90 min). Unfolded ratings for each participant are also available in the supplementary material.

On the other hand, Fig. 5(B, C) analyze the subjective eyestrain perceived by the users as a function of the *p*-ary m-sequence and the condition: pre-session (PRE 60 Hz, and PRE 120 Hz) and during the evaluation (TASK 120 Hz), as explained in Section 3.3. As shown in Fig. 5(B), regarding pre-session conditions, GF(11^2) was perceived as significantly less fatiguing at PRE 120 Hz compared to PRE 60 Hz. Moreover, GF(2^6), GF(3^4) and GF(7^2) m-sequences at PRE 60 Hz were significantly more annoying than TASK 120 Hz. Fig. 5(C) shows the distribution of the eyestrain rating and the *p*-values of the comparisons within conditions. For both pre-session conditions (PRE 60 and PRE 120 Hz), most comparisons (15 out of 20) between *p*-ary m-sequences yielded significant differences. As expected, the number of significant differences decreases (5 out of 10) for the ratings taken during the evaluation (i.e., between EEG blocks). Anyway, for all conditions, the ordered *p*-ary m-sequences from least to greatest fatigue were GF(11^2), GF(7^2), GF(5^3), GF(3^4); and GF(2^6).

5. Discussion

5.1. The orthogonality of c-VEP templates

The first research question asks to analyze how the bases of different *p*-ary m-sequences affect the orthogonality of the templates. Fig. 3 showed the averaged templates across users for each *p*-ary m-sequence.

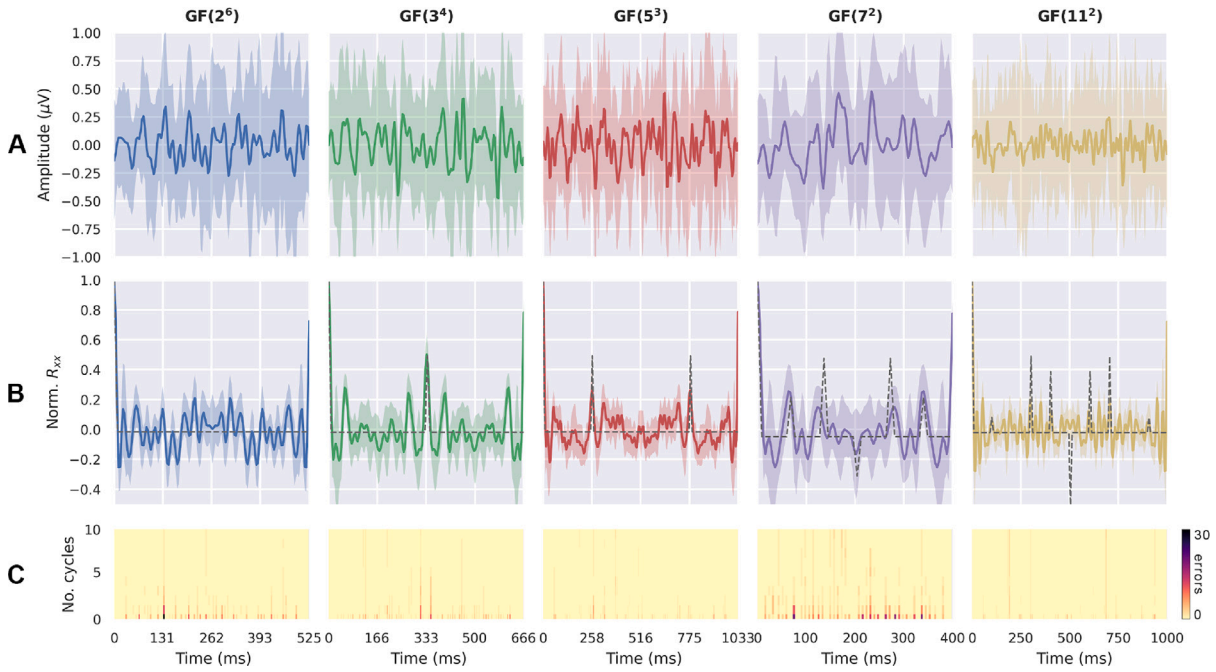


Fig. 3. Users' brain responses to the p-ary m-sequences. Mean values are depicted with solid lines, whereas shaded areas indicate ± standard deviation. (A) Grand-averaged CCA templates across users (summed along the filter banks). (B) Circular autocorrelation function of the templates, superimposed on the circular autocorrelation of the m-sequence (gray dashed line). (C) Number of misclassifications and their temporal displacement from the target command as a function of the number of cycles.

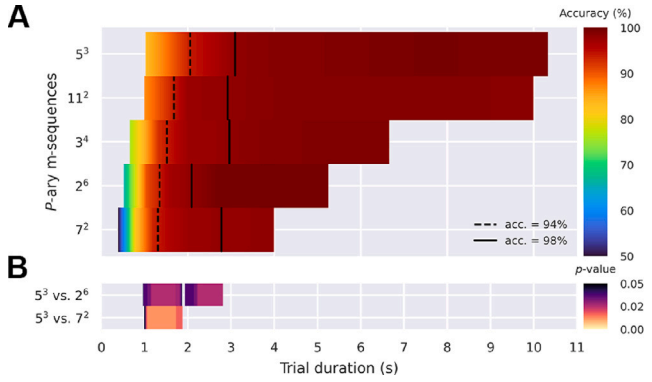


Fig. 4. Fair accuracy comparison between p-ary m-sequences considering trial duration. (A) Interpolated accuracy of each m-sequence, averaged across users. The dashed and solid lines indicate the points where the accuracy exceeds 94% and 98%, respectively. (B) FDR-corrected p-values for comparisons that yielded significant differences.

Although they may seem chaotic at first glance, the low standard deviation of their autocorrelation functions indicates that they are repetitive exogenous brain responses generated each time a new cycle of the m-sequence is shown; i.e., 15.3% for GF(2^6), 16.8% for GF(3^4), 12.8% for GF(5^3), 20.7% for GF(7^2), and 11.9% for GF(11^2). The most variable response is obtained for GF(7^2), which is also the shortest m-sequence (48 bits). This is in agreement with the scientific literature, where it has typically been found that longer codes generate flatter autocorrelation profiles (Martínez-Cagigal et al., 2021).

The brain is a nonlinear dynamic system (e.g., bifurcation, period-doubling, responsiveness of cones, etc.), so templates are not expected to be orthogonal, even though m-sequences are (Gembler, Rezeika, et al., 2020; Martínez-Cagigal et al., 2021). However, it is interesting to see how local minima and maxima of the autocorrelation of the original m-sequences are somewhat correlated with the steepest peaks of the autocorrelation of the obtained templates. Besides the nonlinearity of the brain, templates seem to share some of the autocorrelation properties of the m-sequences, which allows us to encode the commands with shifted versions of the m-sequences and still obtain high decoding accuracy (Martínez-Cagigal et al., 2021). This highlights the usefulness of m-sequences over other codes, since they mathematically offer the flattest autocorrelation profiles.

It is noteworthy that we have not observed a correlation between online selection error made by users and local minima or maxima of the template autocorrelation functions; see Fig. 3(C). The misclassifications that are present for all number of cycles were infrequent, and the majority of them did not coincide with local maxima/minima in the circular autocorrelation functions. Out of the 28 identified misclassifications, approximately 6 were attributable to local maxima/minima. Therefore, we contend that most of the errors were not caused by spurious correlations of the calibrated templates. Even though it is true that the number of errors made is very low (especially for the second cycle onwards) and thus the statistical power is limited, the primary cause of most errors was more related to external artifacts or user-related factors, such as fatigue or loss of concentration.

5.2. On the reliability of the p-ary m-sequences

The second research question prompts to determine whether p-ary m-sequences are adequate to achieve reliable BCIs. As can be seen in Table 2 and Fig. 4, the quality of these templates has allowed to achieve high online decoding accuracy (i.e., >98%) for all the p-ary m-sequences. For 10 cycles per trial, GF(2^6), GF(3^4) and GF(5^3) exceeded 99% grand-averaged accuracy; whereas GF(7^2) and GF(11^2) exceeded 98%. As expected, no significant differences were found when comparing these accuracy results for 10 cycles between m-sequences, which shows that all of them are perfectly adequate to provide a reliable BCI control. In fact, unfolded results (available in the supplementary material) show that almost all users reached a 100% accuracy for most m-sequences: 15 for GF(5^3) and GF(11^2), 14 for GF(2^6) and GF(3^4), and 13 for GF(7^2), out of 16 participants. However, it would not be fair to compare the accuracy results without taking into account the time required to make a selection.

Fig. 4 allows us to analyze the progression of accuracy as the duration of the time window required to make a command selection

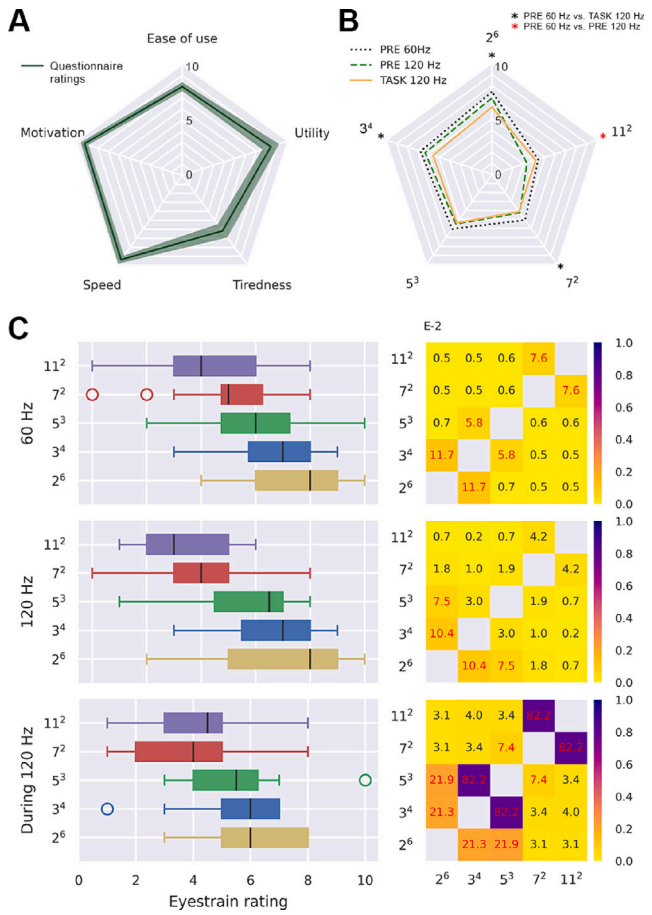


Fig. 5. Summary of qualitative results. (A) Normalized ratings of the questionnaire grouped in relevant aspects of the study. Unfolded ratings for each user are available in the supplementary material. The average is indicated by the solid line, whereas the shaded area represent the 95% confidence interval of the distribution. (B) Averaged perceived eyestrain of each m-sequence as a function of the condition: pre-session at 60 Hz, pre-session at 120 Hz and during tasks at 120 Hz. Significant differences (p-value < 0.05) between conditions are indicated with asterisks. (C) Left: detailed eyestrain rating distributions of each m-sequence within conditions. The vertical black lines denote the median of the distribution, while the length of the box indicates the interquartile range (i.e., Q3-Q1). Right: Corrected p-values from the eyestrain comparison between m-sequences (in scientific notation E-2). Black fonts indicate significant differences (p-value < 0.05).

increases. Obviously, as the duration of the trial increases, the accuracy increases for all p-ary m-sequences. Nevertheless, the slope of this increase varies between m-sequences. While the 94% landmark is reached between 1.3–2.1 s trial durations; the 98% point is reached between 2.7–3.1 s, except for GF(2^6), which achieves it at 2.1 s. Although the increment of GF(3^4), GF(5^3), GF(7^2) and GF(11^2) behaves similarly; GF(2^6) presents a steeper slope. Less evident patterns have been uncovered by the statistical analysis, though. When comparing these accuracy progressions as a function of the trial duration between all p-ary m-sequence combinations, only two yielded significant differences: GF(5^3) vs. GF(2^6) between 1–3 s, and GF(5^3) vs. GF(7^2) between 1–2 s. We can see that these differences were produced because GF(5^3) achieved significantly lower accuracy than GF(2^6) and GF(7^2) in those time windows. For instance, GF(5^3) obtained 94.14% with 2.06 s, while GF(2^6) reached 98.05% with 2.1 s, and GF(7^2) reached 97.27% with 2.00 s. These differences highlight the importance of analyzing performance while simultaneously considering accuracy and selection time. Since the rest of comparisons were not significantly different for any time window, we can consider that all m-sequences except GF(5^3) are similar in terms of accuracy and trial duration.

Given that all p-ary m-sequences have demonstrated their ability to provide reliable control, it would be worthwhile to briefly analyze other properties of our system, such as fault tolerance and resilience. Fault tolerance refers to a system's ability to continue operating correctly despite component faults (Koren & Krishna, 2020). Resilience can be defined as the system's ability to adapt to unexpected events (Woods, 2018). Additionally, some authors interpret this concept as a post-damage attribute of a system, characterizing it as the ability to operate at a desired level despite experiencing partial damage (Zhang & Lin, 2010). In this context, our proposal demonstrates the capability to address jitter (i.e., slight latency variations) in stimulus updating and/or EEG sampling. Additionally, the brain's broadband response to a c-VEP-encoded stimulus and the filter bank approach makes the system capable to handle narrow-band interferences like alpha waves or user blinks. In the presence of such events, the system retains its ability to accurately decode user intentions. However, it is expected that an increase in the required stimulation cycles will occur, consequently reducing the overall ITR of the system. The multi-channel approach also facilitates the handling of electrode failures, as long as they do not occur on the primary visual cortex range (i.e., occipital cortex). On the other hand, other faults like Bluetooth transceiver failures or software disruptions exceed the system's capability to handle.

5.3. Eyestrain differences

After demonstrating the usefulness of all the proposed p-ary m-sequences to provide high-speed, reliable c-VEP-based BCIs, the last research question must be answered. This prompts us to find out if users' eyestrain is related to the base of the p-ary m-sequence.

Firstly, it is interesting to examine the differences between presentation rates; i.e., PRE 60 Hz and PRE 120 Hz before starting the session, and TASK 120 Hz between EEG recordings. There is a consensus in the scientific literature that states that the higher the presentation rate, the less visual fatigue (Başaklar et al., 2019; Gembler et al., 2018; Gembler, Stawicki, Rezeika, & Volosyak, 2019; Nezamfar et al., 2011, 2016). Similarly, it seems that the higher the presentation rate, the less orthogonal the templates become (Başaklar et al., 2019; Gembler et al., 2018); therefore there is a trade-off between performance and user comfort, as far as the rate of presentation is concerned. As shown in Fig. 5, our results point in the same direction. Eyestrain ratings were higher for the 60 Hz presentation rate than for 120 Hz (both pre- and during session). In particular, these differences were significant for GF(2^6), GF(3^4) and GF(7^2) in PRE 60 Hz vs. TASK 120 Hz; and for GF(11^2) in PRE 60 Hz vs. PRE 120 Hz. These results show that, in general, the presentation rate of 60 Hz is significantly more annoying than 120 Hz for most p-ary m-sequences, probably because in the latter the contrast changes are faster and less noticeable.

The available studies on c-VEP (Başaklar et al., 2019; Gembler et al., 2018; Gembler, Stawicki, Rezeika, & Volosyak, 2019; Nezamfar et al., 2011, 2016) and SSVEP (Chen et al., 2023; Ladouce et al., 2022) provide evidence that a higher stimulation frequency results in less visual discomfort. However, none of these studies have provided an explanation for this phenomenon in terms of the neural mechanisms involved in visual perception. A plausible explanation for this phenomenon may be related to the critical flicker fusion frequency (CFFF), which is the minimum frequency at which a flickering light source appears steady (Ladouce et al., 2022). Young adults' CFFF averages at 60 Hz, ranging from 40 to 80 Hz (Ladouce et al., 2022). When a visual stimulus flickers at a constant rate of 120 Hz, it generates an SSVEP tone at 60 Hz and its harmonics. However, for many individuals, this flicker is not perceptible as it falls outside the CFFF threshold. Similarly, an m-sequence operating at 120 Hz would generate a broadband response that is limited to the range up to 60 Hz, and the spectral distribution of the same m-sequence at a rate of 60 Hz would produce a response up to 30 Hz (Martínez-Cagigal et al., 2021). Due to the proximity of the former to the CFFF value, visual stimuli presented at

a higher refresh rate are expected to produce more continuous visual stimulation, which in turn would limit the perceived differences in contrast between adjacent stimuli. This perspective is in line with the literature which posits that visual stimuli are perceived as unpleasant when their spatial and temporal characteristics deviate from the statistics of natural environment (Gentile & Aguirre, 2020). Hence, we hypothesize that a lower presentation rate would produce visual stimulation that is less natural and therefore more likely to cause visual discomfort.

The phenomenon may also be explained by the widely accepted theory of "visual stress", which posits that visual discomfort arises from the overactivation of the visual cortex (Wilkins, 1995). This phenomenon is particularly pronounced when stimulus power is concentrated at frequencies falling within the midrange of human perception (Fernandez & Wilkins, 2008); e.g., frequencies between 12–18 Hz, which are thought to generate the strongest SSVEPs (Kuś et al., 2013). From this perspective, a higher refresh rate would reduce the power of these low frequencies responsible for visual discomfort. However, further research is necessary to determine the precise neural mechanisms involved in this phenomenon, which remain incompletely understood.

Interestingly, eyestrain ratings were lower in PRE 120 Hz vs. TASK 120 Hz except for GF(5^3), although not significant. In the PRE condition, all m-sequences were displayed at once, so users could make relative comparisons among them, even though they were unaware of the duration of each m-sequence (no spelling task had been performed yet). In contrast, the order of presentation of each m-sequence in TASK condition was random, so users were unaware of which m-sequence encoding they were paying attention to. For these reasons, a slight variation between these ratings was expected. Of course, it is important to analyze whether the order of presentation of each of the m-sequences biases the results. Since this order was counter-balanced and randomized across participants, we took the users' ratings for each block regardless of the m-sequence shown; i.e., ratings of the first block, for the second block, and so on. Then, a Friedman test for repeated samples was computed to check if the distributions between blocks are significantly similar. No significant differences were found (p-value = 0.34), so we can state that the order of display of the m-sequences did not bias our results. This implies that the distribution of ratings for both 120 Hz conditions were consistent across users, and that the perception of visual fatigue is not influenced by the evaluation protocol, but that the ratings reliably represent users' comfort among p-ary m-sequences.

Since the order of presentation does not influence the perception of visual fatigue, it is time to analyze the differences between the proposed p-ary m-sequences. The detailed p-values of the comparisons between m-sequences of Fig. 5(C) let us infer some interesting conclusions. On the one hand, for all conditions, users generally reported that the higher the base, the less visual fatigue, especially for prior 60 Hz and 120 Hz conditions. On the other hand, significant differences between these ratings were generally found for nonadjacent m-sequences. For instance, GF(2^6) and GF(3^4) were not significantly different for any condition, but GF(2^6) vs. GF(7^2) or GF(11^2) yielded significant differences for all conditions. This shows that the ability to discern between m-sequences with a slightly different number of shades of gray is not trivial, and that differences between perceived eyestrain are enhanced as the variation between the number of levels (i.e., base) increases. This phenomenon was expected, since several users reported that they were not able to discern between different sequences; e.g., user U16 between GF(7^2) and GF(11^2), or user U12 between GF(5^3) and GF(7^2). With respect to the common binary sequences of the scientific literature, we have found that GF(7^2) and GF(11^2) were significantly less annoying for all conditions, and GF(5^3) for the 60 Hz condition.

The direct relation between the base increment with the reduction of visual fatigue may be also explained by the theory of "visual stress" (Wilkins, 1995). In our case, shades of gray are mainly processed by the LMS (luminance) postreceptoral pathway, which is though to elicit stronger VEP responses than the other pathways, i.e., L-M (red-green) and S (blue-yellow) (Gentile & Aguirre, 2020). On the other hand, primary visual cortex is more sensitive to high-contrast stimuli (Wandell et al., 2007), which activate a large number of neurons and lead to a stronger response in the foveal region of the retina (Baseler et al., 1994). It is then hypothesized that high-contrast stimuli that stimulate the LMS pathway (e.g., black/white changes) may result in overstimulation of the visual cortex, leading to an increased experience of visual discomfort. This is consistent with the findings of Ladouce et al. (2022) and Gembler, Rezeika, et al. (2020) who observed increased visual fatigue for high-contrast stimuli in SSVEP and c-VEP paradigms, respectively.

The contrast variations in our shades of gray can be measured using depth amplitude, which is the difference in luminance between the stimulus's maximum and minimum values (Ladouce et al., 2022). Maximal depth amplitude of 100% is caused by a change from black to white or vice versa. The binary m-sequence GF(2^6) presents maximal depth amplitude for all stimulus changes, whereas the percentage of maximal stimulus changes in the p-ary m-sequences decreases as the base increases: 33.96% for GF(3^4), 10.10% for GF(5^3), 4.88% for GF(7^2), and 1.83% for GF(11^2). It is suggested that higher base m-sequences may lead to reduced visual discomfort as a result of the visual cortex being less stimulated by the greater number of non-maximal amplitude depths present in such p-ary m-sequences. Once more, it is our belief that additional research is necessary to evaluate this hypothesis and provide a deeper understanding of the specific mechanisms at play.

Finally, we would like to analyze the satisfaction of the participants with the final c-VEP-based BCI concerning the questionnaire. All positive statements were rated above the neutral rating, and all negative statements were rated below it. Particularly, users were positively surprised by the speed of command selection and the reliability of the BCI. Some of them reported to be familiarized with P300-based BCIs, which require longer calibration times (around half an hour), as well more time to perform a selection (approx. 10–30 s depending on the number of repetitions) (Martínez-Cagigal et al., 2017, 2019; Santamaría-Vázquez et al., 2019). Users were also satisfied with the ease of use of the app, they reported to be motivated to carry out the evaluation, and that they felt that this app could be an AAC for MDUs. Noteworthy, they also reported slight tiredness throughout the assessment, a fact that must be considered in future studies to design shorter evaluation sessions.

5.4. Comparison with other studies

Only Gembler, Rezeika, et al. (2020) tested a c-VEP-based BCI speller with a GF(5^3) apart from the common binary m-sequences. Similarly to our results, Gembler, Rezeika, et al. (2020) found that GF(5^3) was significantly less annoying to users than GF(2^6), especially for a presentation rate of 60 Hz. They did not find significant differences between GF(2^5) and GF(5^3) in terms of accuracy: 99.4% and 98.5% (60 Hz), 97.6% and 97.5% (120 Hz), and 97.9% and 97.6% (240 Hz), respectively (Gembler, Rezeika, et al., 2020). A direct comparison between our accuracy and theirs is not appropriate, since there are substantial differences in experiment design. For instance, they used an 8-target speller, 6 evaluation sessions with 18 HU, and a fixed trial duration of 2.06 s. However, what is comparable are the differences between sequences in terms of precision and visual fatigue. Recalling our results, we found that GF(5^3) was significantly less fatiguing than GF(2^6) for the 60 Hz condition, and less fatiguing but not significant for 120 Hz conditions. Similarly, differences in accuracy were generally not significant, except for trial duration of 1–3 s, where we found that GF(5^3) had significantly lower accuracy than GF(2^6). We cannot compare the results for other m-sequences with bases other than 2 or 5, since their performance was not evaluated in their study (Gembler, Rezeika, et al., 2020). In general, we claim that the results of Gembler,

Rezeika, et al. (2020) are similar to ours, showing that non-binary sequences are also able to achieve high accuracy and are less obtrusive to the end user.

The rest of studies aimed at improving user comfort in c-VEP-based BCIs did not propose the use of p-ary m-sequences, but instead focused on other aspects. On the one hand, several studies proposed to increase the presentation rate from 60 Hz to 120 Hz, stating that the latter presentation rate was more comfortable for users (Başaklar et al., 2019; Gembler et al., 2018; Wittevrongel et al., 2017). In this regard, our comparison between 60 Hz and 120 Hz conditions has also shown that 120 Hz is significantly less annoying to users than 60 Hz for most of the proposed p-ary m-sequences. On the other hand, some studies proposed to use hand-crafted binary codes that confine spectral density to high-frequency bands (Behboodi et al., 2020; Gembler, Rezeika, et al., 2020; Shirzhiyan et al., 2019; Yasinzai & Ider, 2020). It would be interesting to apply the same concept to the p-ary m-sequences in the future and evaluate whether differences between eyestrain ratings are found.

In addition to supporting the conclusions of previous studies, in this work we have also analyzed, for the first time, the influence of different p-ary m-sequences on system performance and user comfort. We have also analyzed how the base affects the orthogonality of the c-VEP templates, which are responsible for achieving high decoding accuracy in real-time. Finally, the qualitative analysis between the sequences and the presentation frequencies of 60 Hz and 120 Hz has given insight into the perceived visual fatigue in different conditions, an interesting dissection that could drive the development of more user-friendly c-VEP systems.

5.5. Limitations and future work

After this discussion, we consider we have answered all the research questions posed in the introduction section, showing that all p-ary m-sequences are viable to provide high-speed, high-accuracy c-VEP-based BCIs. Nevertheless, there are aspects that can be improved in future studies.

First, the p-ary m-sequences have been tested with HUs. However, it would be interesting to increase the sample size with MDUs, the typical target users of BCIs. Since MDUs tend to achieve lower performance than HUs, it would be appropriate to analyze system performance with this population (Martínez-Cagigal et al., 2021). Second, command lags have been placed using the autocorrelation function of the original m-sequences. Instead, it would be also interesting to place command lags based on the correlation between brain responses (Thielen et al., 2015). Third, some users reported having difficulty keeping focus on target commands and that blurring their eyes helped them focus on the target command and avoid distractions from surrounding stimuli. Some of them also reported that they were aware of the cyclic repetition of commands and saw vertical or horizontal patterns between stimuli, which distracted them. This is probably because the shifts used to encode each command were placed sequentially on the command grid. An improved version of the c-VEP speller could be achieved by further separating the commands and/or optimizing the lags to prevent cross-talk between neighboring cells (Thielen et al., 2015).

The success of this study also suggests several future lines of research. For instance, analyzing the effect of using color encodings instead of gray tones would be an interesting research line. In fact, color stimuli have been shown to affect EEG responses in both c-VEP and SSVEP-based BCIs (Martínez-Cagigal et al., 2021). Studying how to implement an asynchronous stage (i.e., non-control detection) and/or early stopping algorithms (i.e., to dynamically detect the required number of cycles for each epoch) for these p-ary m-sequences would be interesting as well. Also, the extra dimension provided by the base (i.e., number of levels) opens up new possibilities for further analyses. How these levels affect the individual VEP response? Could EEG templates be predicted by overlapping individual events or using regression algorithms? Could these p-ary m-sequences be encoded with variations in stimuli sizes, apparent motions or changing images? These exciting open questions still remain unanswered in the current literature on c-VEPs.

6. Conclusion

In this study, we analyzed the ability of non-binary (i.e., p-ary, $p > 2$) m-sequences to provide less obtrusive c-VEP-based BCIs to end users, while maintaining high speed and accuracy. We generated for the first time five different p-ary m-sequences to encode a 16-target BCI speller under the circular shifting paradigm: base 2, 63-bit GF(2^6); base 3, 80-bit GF(3^4); base 5, 124-bit GF(5^3); base 7, 48-bit GF(7^2); and base 11, 120-bit GF(11^2). For non-binary bases, levels were encoded using flashing commands with different shades of gray. A quantitative and qualitative evaluation was performed with 16 healthy participants in a single session.

Users were able to reach more than 98% grand-averaged accuracy with 10 cycles using all the proposed p-ary m-sequences, and even 94% with 3 s of trial duration. We found that the higher the base (i.e., number of levels), the less eyestrain users perceive. The differences in eyestrain were significant between most of p-ary m-sequences, and between presentation rates of 60 Hz and 120 Hz, the latter being the most comfortable condition. In general, the differences in accuracy were not significant between the p-ary m-sequences, although they presented different slopes of progression when the number of cycles increased. Based on these results, we encourage the use of p-ary m-sequences to encode c-VEP-based BCI commands, as they achieve performance similar to common binary codes, but are significantly more comfortable to the user.

CRediT authorship contribution statement

Víctor Martínez-Cagigal: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Eduardo Santamaría-Vázquez:** Software, Writing – review & editing. **Sergio Pérez-Velasco:** Writing – review & editing. **Diego Marcos-Martínez:** Writing – review & editing. **Selene Moreno-Calderón:** Writing – review & editing. **Roberto Hornero:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: R. Hornero reports financial support was provided by Spain Ministry of Science and Innovation. V. Martínez-Cagigal, E. Santamaría-Vázquez, S. Pérez-Velasco, D. Marcos-Martínez, and R. Hornero report administrative support was provided by Biomedical Research Network Centre of Bioengineering Biomaterials and Nanomedicine. E. Santamaría-Vázquez, S. Pérez-Velasco, and D. Marcos-Martínez report financial support was provided by Junta de Castilla y León Consejería Educación.

Data availability

Data will be made available on request.

Acknowledgments

This study was partially funded by 'Ministerio de Ciencia e Innovación/Agencia Estatal de Investigación' and European Regional Development Fund (ERDF) [TED2021-129915B-I00, RTC2019-007350-1, and PID2020-115468RB-I00]; and 'CIBER en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN)' through 'Instituto de Salud Carlos III'. E. S.-V., S. P.-V. and D. M.-M. were in receipt of a PIF grant by the 'Consejería de Educación de la Junta de Castilla y León'.

Appendix A. Avoiding local maximum correlations

Given a p-ary m-sequence $\boldsymbol{x} \in \mathbb{R}^{N,1}$ with $p > 2$ and the number of commands to be encoded, m (where $N > m$), the objective of the algorithm is to avoid local minima and maxima in the autocorrelation function when assigning lags to different commands:

1. Compose a vector containing the circular autocorrelation values of the m-sequence, i.e. $\boldsymbol{r} = [R_{xx}(0), \dots, R_{xx}(N-1)]$, where $R_{xx}(\tau)$ denotes the circular autocorrelation value for a relative lag τ in samples.
2. Determine the indexes $\boldsymbol{i}$ in which $\boldsymbol{r}$ shows a local maximum or minimum:
   2.1. Calculate the standard autocorrelation value using $\gamma = \text{mode}(\boldsymbol{r})$, or $\gamma = -\min(|\boldsymbol{r}|)$.
   2.2. Find indexes whose value in $\boldsymbol{r}$ is different from γ, i.e. $\boldsymbol{i} = \arg_{j \neq \gamma}(r_j)$
3. Create a zero-vector $\boldsymbol{z}$ of length $N - b$, where b is the length of $\boldsymbol{i}$.
4. Put ones in evenly sampled positions, $\boldsymbol{z}(j) = 1$ for $j = \lfloor u(N-1)/m \rceil$, where $u = 1, \dots, m$ and $\lfloor \cdot \rceil$ denotes rounding.
5. Extend the length of $\boldsymbol{z}$ to N by inserting zeros at the positions indexed in $\boldsymbol{i}$.
6. Lags for the encoding commands $\boldsymbol{l}$ are found in the indexes where the vector $\boldsymbol{z}$ equals one, i.e. $l_j = j$ if $z_j = 1$ for $j = 1, \dots, N$.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.eswa.2023.120815.

References

Başaklar, T., Tuncel, Y., & Ider, Y. Z. (2019). Effects of high stimulus presentation rate on EEG template characteristics and performance of c-VEP based BCIs. *Biomedical Physics and Engineering Express*, *5*(3), http://dx.doi.org/10.1088/2057-1976/ab0cee.

Baseler, H. A., Sutter, E. E., Klein, S. A., & Carney, T. (1994). The topography of visual evoked response properties across the visual field. *Electroencephalography and Clinical Neurophysiology*, *90*(1), 65–81. http://dx.doi.org/10.1016/0013-4694(94)90114-7.

Behboodi, M., Mahnam, A., Marateb, H., & Rabbani, H. (2020). Optimization of visual stimulus sequence in a brain-computer interface based on code modulated visual evoked potentials. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(12), 2762–2772. http://dx.doi.org/10.1109/TNSRE.2020.3044947.

Bin, G., Gao, X., Wang, Y., Hong, B., & Gao, S. (2009). VEP-based brain-computer interfaces: Time, frequency, and code modulations. *IEEE Computational Intelligence Magazine*, *4*(4), 22–26. http://dx.doi.org/10.1109/MCI.2009.934562.

Buračas, G. T., & Boynton, G. M. (2002). Efficient design of event-related fMRI experiments using m-sequences. *NeuroImage*, *16*(3 I), 801–813. http://dx.doi.org/10.1006/nimg.2002.1116.

Chen, X., Liu, B., Wang, Y., Cui, H., Dong, J., Ma, R., Li, N., & Gao, X. (2023). Optimizing stimulus frequency ranges for building a high-rate high frequency SSVEP-BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *31*, 1277–1286. http://dx.doi.org/10.1109/TNSRE.2023.3243786.

Combaz, A., Chatelle, C., Robben, A., Vanhoof, G., Goeleven, A., Thijs, V., Van Hulle, M. M., & Laureys, S. (2013). A comparison of two spelling brain-computer interfaces based on visual P3 and SSVEP in locked-in syndrome. *PLoS One*, *8*(9), http://dx.doi.org/10.1371/journal.pone.0073691.

Di Porto, A., Guida, F., & Montolivo, E. (1992). Fast algorithm for finding primitive polynomials over GF (q). *Electronics Letters*, *28*(2), 118–120.

Fernandez, D., & Wilkins, A. J. (2008). Uncomfortable images in art and nature. *Perception*, *37*(7), 1098–1113. http://dx.doi.org/10.1068/p5814.

Gembler, F. W., Benda, M., Rezeika, A., Stawicki, P. R., & Volosyak, I. (2020). Asynchronous c-VEP communication tools—efficiency comparison of low-target, multi-target and dictionary-assisted BCI spellers. *Scientific Reports*, *10*(1), 1–13. http://dx.doi.org/10.1038/s41598-020-74143-4.

Gembler, F. W., Rezeika, A., Benda, M., & Volosyak, I. (2020). Five shades of grey: Exploring quintary m -sequences for more user-friendly c-VEP-based BCIs. *Computational Intelligence and Neuroscience*, *2020*, http://dx.doi.org/10.1155/2020/7985010.

Gembler, F., Stawicki, P., Rezeika, A., Saboor, A., Benda, M., & Volosyak, I. (2018). Effects of monitor refresh rates on c-VEP BCIs. In *International workshop on symbiotic interaction* (pp. 53–62). Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-91593-7, URL http://link.springer.com/10.1007/978-3-319-91593-7.

Gembler, F., Stawicki, P., Rezeika, A., & Volosyak, I. (2019). A comparison of cVEP-based BCI-performance between different age groups. In *Proceedings of the international work-conference on artificial neural networks, vol. 11506 LNCS* (pp. 394–405). Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-20521-8_33.

Gembler, F., Stawicki, P., Saboor, A., & Volosyak, I. (2019). Dynamic time window mechanism for time synchronous VEP-based BCIs-performance evaluation with a dictionary-supported BCI speller employing SSVEP and c-VEP. *PLoS One*, *14*(6), 1–18. http://dx.doi.org/10.1371/journal.pone.0218177.

Gentile, C. P., & Aguirre, G. K. (2020). A neural correlate of visual discomfort from flicker. *Journal of Vision*, *20*(7), 11. http://dx.doi.org/10.1167/jov.20.7.11.

Koren, I., & Krishna, C. M. (2020). *Fault-Tolerant Systems*. Morgan Kaufmann.

Kuś, R., Duszyk, A., Milanowski, P., Łabecki, M., Bierzyńska, M., Radzikowska, Z., Michalska, M., Zygierewicz, J., Suffczyński, P., & Durka, P. J. (2013). On the quantification of SSVEP frequency responses in human EEG in realistic BCI conditions. *PLoS One*, *8*(10), http://dx.doi.org/10.1371/journal.pone.0077536.

Ladouce, S., Darmet, L., Torre Tresols, J. J., Velut, S., Ferraro, G., & Dehais, F. (2022). Improving user experience of SSVEP BCI through low amplitude depth and high frequency stimuli design. *Scientific Reports*, *12*(1), 1–12. http://dx.doi.org/10.1038/s41598-022-12733-0.

Martínez-Cagigal, V., Gomez-Pilar, J., Álvarez, D., & Hornero, R. (2017). An asynchronous P300-based brain–computer interface web browser for severely disabled people. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *25*(8), 1332–1342. http://dx.doi.org/10.1109/TNSRE.2016.2623381.

Martínez-Cagigal, V., Santamaría-Vázquez, E., Gomez-Pilar, J., & Hornero, R. (2019). Towards an accessible use of smartphone-based social networks through brain-computer interfaces. *Expert Systems with Applications*, *120*, 155–166. http://dx.doi.org/10.1016/j.eswa.2018.11.026.

Martínez-Cagigal, V., Thielen, J., Santamaría-Vázquez, E., Pérez-Velasco, S., Desain, P., & Hornero, R. (2021). Brain-computer interfaces based on code-modulated visual evoked potentials (c-VEP): a literature review. *Journal of Neural Engineering*, *18*(6), Article 061002. http://dx.doi.org/10.1088/1741-2552/ac38cf.

Mullen, G. L., & Panario, D. (2013). *Handbook of finite fields, vol. 17*. CRC Press Boca Raton.

Nagel, S., Dreher, W., Rosenstiel, W., & Spüler, M. (2018). The effect of monitor raster latency on VEPs, ERPs and brain–computer interface performance. *Journal of Neuroscience Methods*, *295*, 45–50. http://dx.doi.org/10.1016/j.jneumeth.2017.11.018.

Nezamfar, H., Orhan, U., Purwar, S., Hild, K., Oken, B., & Erdogmus, D. (2011). Decoding of multichannel EEG activity from the visual cortex in response to pseudorandom binary sequences of visual stimuli. *International Journal of Imaging Systems and Technology*, *21*(2), 139–147. http://dx.doi.org/10.1002/ima.20288.

Nezamfar, H., Salehi, S. S. M., & Erdogmus, D. (2016). Stimuli with opponent colors and higher bit rate enable higher accuracy for C-VEP BCI. In *2015 IEEE signal processing in medicine and biology symposium - proceedings* (pp. 1–6). IEEE, http://dx.doi.org/10.1109/SPMB.2015.7405476.

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, *112*(4), 713–719. http://dx.doi.org/10.1016/S1388-2457(00)00527-7.

Peters, B., Bedrick, S., Dudy, S., Eddy, B., Higger, M., Kinsella, M., McLaughlin, D., Memmott, T., Oken, B., Quivira, F., Spaulding, S., Erdogmus, D., & Fried-Oken, M. (2020). SSVEP BCI and eye tracking use by individuals with late-stage ALS and visual impairments. *Frontiers in Human Neuroscience*, *14*(November), 1–15. http://dx.doi.org/10.3389/fnhum.2020.595890.

Santamaría-Vázquez, E., Martínez-Cagigal, V., Gomez-Pilar, J., & Hornero, R. (2019). Asynchronous control of ERP-based BCI spellers using steady-state visual evoked potentials elicited by peripheral stimuli. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*(9), 1883–1892.

Santamaría-Vázquez, E., Martínez-Cagigal, V., Marcos-Martínez, D., Rodríguez-González, V., Pérez-Velasco, S., Moreno-Calderón, S., & Hornero, R. (2023). MEDUSA©: A novel python-based software ecosystem to accelerate brain-computer interface and cognitive neuroscience research. *Computer Methods and Programs in Biomedicine*, *230*, Article 107357. http://dx.doi.org/10.1016/j.cmpb.2023.107357.

Santamaría-Vázquez, E., Martínez-Cagigal, V., Vaquerizo-Villar, F., & Hornero, R. (2020). EEG-inception: A novel deep convolutional neural network for assistive ERP-based brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(12), 2773–2782. http://dx.doi.org/10.1109/TNSRE.2020.3048106.

Shirzhiyan, Z., Keihani, A., Farahi, M., Shamsi, E., GolMohammadi, M., Mahnam, A., Haidari, M. R., & Jafari, A. H. (2019). Introducing chaotic codes for the modulation of code modulated visual evoked potentials (c-VEP) in normal adults for visual fatigue reduction. *PLoS One*, *14*(3), 1–29. http://dx.doi.org/10.1371/journal.pone.0213197.

Sutter, E. E. (1992). The brain response interface: communication through visually-induced electrical brain responses. *Journal of Microcomputer Applications*, *15*(1), 31–45. http://dx.doi.org/10.1016/0745-7138(92)90045-7.

Thielen, J., Marsman, P., Farquhar, J., & Desain, P. (2021). From full calibration to zero training for a code-modulated visual evoked potentials brain computer interface. *Journal of Neural Engineering*, *18*(5), 56007. http://dx.doi.org/10.1088/1741-2552/abecef.

Thielen, J., Van Den Broek, P., Farquhar, J., & Desain, P. (2015). Broad-band visually evoked potentials: Re (con)volution in brain-computer interfacing. *PLoS One*, *10*(7), 1–22. http://dx.doi.org/10.1371/journal.pone.0133797.

Verbaarschot, C., Tump, D., Lutu, A., Borhanazad, M., Thielen, J., Van, P. D. B., Farquhar, J., Weikamp, J., Raaphorst, J., Groothuis, J. T., & Desain, P. (2021). A visual brain-computer interface as communication aid for patients with amyotrophic lateral sclerosis. *Clinical Neurophysiology*, http://dx.doi.org/10.1016/j.clinph.2021.07.012.

Vialatte, F. B., Maurice, M., Dauwels, J., & Cichocki, A. (2010). Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. *Progress in Neurobiology*, *90*(4), 418–438. http://dx.doi.org/10.1016/j.pneurobio.2009.11.005.

Volosyak, I., Valbuena, D., Lüth, T., Malechka, T., & Gräser, A. (2011). BCI demographics II: how many (and what kinds of) people can use a high-frequency SSVEP BCI? *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *19*(3), 232–239. http://dx.doi.org/10.1109/TNSRE.2011.2121919.

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366–383. http://dx.doi.org/10.1016/j.neuron.2007.10.012.

Wilkins, A. J. (1995). *Visual stress, vol. 24*. OUP Oxford.

Wittevrongel, B., Van Wolputte, E., & Van Hulle, M. M. (2017). Code-modulated visual evoked potentials using fast stimulus presentation and spatiotemporal beamformer decoding. *Scientific Reports*, *7*(1), 1–10. http://dx.doi.org/10.1038/s41598-017-15373-x.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, *113*(6), 767–791. http://dx.doi.org/10.1016/S1388-2457(02)00057-3.

Wolpaw, J., & Wolpaw, E. W. (2012). *Brain-computer interfaces: Principles and practice*. OUP USA.

Woods, D. D. (2018). Resilience is a verb. *Domains of Resilience for Complex Interconnected Systems*, 167.

Yasinzai, M. N., & Ider, Y. Z. (2020). New approach for designing cVEP BCI stimuli based on superposition of edge responses. *Biomedical Physics and Engineering Express*, *6*(4), 45018. http://dx.doi.org/10.1088/2057-1976/ab98e7.

Zhang, W. J., & Lin, Y. (2010). On the principle of design of resilient systems - application to enterprise information systems. *Enterprise Information Systems*, *4*(2), 99–110. http://dx.doi.org/10.1080/17517571003763380.