



**HAL**  
open science

## Auto-calibration of c-VEP BCI by word prediction

Federica Turi, Nathalie T.H. Gayraud, Maureen Clerc

► **To cite this version:**

Federica Turi, Nathalie T.H. Gayraud, Maureen Clerc. Auto-calibration of c-VEP BCI by word prediction. 2020. hal-02844024

**HAL Id: hal-02844024**

**<https://hal.science/hal-02844024>**

Preprint submitted on 7 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Auto-calibration of c-VEP BCI by word prediction

F Turi<sup>1,2</sup>, N T H Gayraud<sup>1,2</sup>, M Clerc<sup>1,2</sup>

<sup>1</sup> Inria Sophia Antipolis-Méditerranée, France

<sup>2</sup> Université Côte d’Azur, France

E-mail: federica.turi@inria.fr

**Abstract.** A code-modulated Visual Evoked Potential Brain Computer Interface (c-VEP BCI) allows for spelling from a virtual keyboard of flashing characters. All characters flash simultaneously, and each character flashes according to a predefined pseudo-random binary sequence, circular-shifted by a different time lag. For a given character, the pseudo-random stimulus sequence evokes a VEP in the electroencephalogram (EEG) of the subject, which can be used as a template. This template is usually obtained during a calibration phase and it is applied for the target identification during the spelling phase. A downside of a c-VEP BCI system is that it needs a long calibration phase to reach good performance. This paper proposes an unsupervised method that avoids the calibration phase in a c-VEP BCI, by extracting relative lags from the VEP responses, between successive characters, and predicting the full word using a dictionary. We tested it in offline experiments on a public dataset. We simulated the spelling of four groups of words with a different total number of characters selected from an English dictionary. Each experiment is parameterized by the number of stimulus cycles. The obtained results show that a word-prediction-based auto-calibration method in c-VEP BCIs can be efficient and effective.

*Keywords:* Brain Computer Interface, c-VEP BCI, auto-calibration, word prediction

## 1. Introduction

A Brain Computer Interface (BCI) is a hardware and software communication system that converts electro-physiological input from the user into an output that allows to control external devices and to communicate [1]. One of the uses of BCI is to restore linguistic communication by word spelling. Among the BCIs based on electroencephalographic signals (EEG) there are several BCI spellers based on different paradigms, such as the P300 Speller [2], BCI spellers based on motor imagery (MI) [3] and BCI spellers that use the Visual Evoked Potential (VEP). Among VEP BCI spellers it is possible to distinguish different systems depending on the specific stimulus modulation design used [4]: frequency modulated VEP BCIs, such as the steady state evoked potential (SSVEP) BCI [5], and BCI systems using pseudo-random code-modulated VEP (c-VEP) [6].

Each type of BCI needs a calibration phase, in which the system sets the parameters to extract the relevant information from the EEG signal, in term of spatial or temporal

filters and classifiers. This phase is fundamental but it is really annoying, long and tedious for the subject. It is considered time consuming and one of the limitations for the diffusion of the BCI system in real applications [7]. A solution to this limitation is the development of a system that avoids the calibration phase. Many studies underline the importance of a calibration-free BCI and propose different strategies to reduce the calibration time duration. Kraudelat et al. [8] proposed a *Zero-training* method in which they showed how to learn good spatial filters and classifiers from data of previous sessions, which eliminates the necessity of going through a calibration during each session of a subject. This method can reduce the duration of the calibration phase or even avoid the calibration for “experienced” BCI users. Many works [3] proposed similar approaches based on transfer learning methods [9]. Lu et al. [10] proposed an adaptive online learning method for a P300 speller. The first step of the method consists in defining offline a generic subject-independent model from a EEG data set of several subjects, in order to find the specific structure of the P300 response. Then, with a new subject the system automatically captures subject-specific EEG characteristics during online operation without a supervised calibration. The final decision is made by labels predicted by either the subject-independent model or the adapted subject-specific model, depending on a confidence score. They demonstrated that after the spelling of 10–20 characters with the online adaptation, the accuracy of the adapted model reached the accuracy value of a trained supervised subject-specific model. Lotte et al. [11] proposed a method to reduce calibration time in which they increased the training set size by adding numerous artificial EEG signals from a few signals recorded from the user. They demonstrated that with this method the classification accuracy increased with respect to existing approaches, especially when the training trial size available is small. Kindermans et al. [12] proposed an unsupervised model incorporating a transfer learning method and a language model for a P300 speller BCI. They demonstrated that this unsupervised model can reach even better performance than the supervised model.

Many BCI systems exploit a language model to improve the performance of the systems [13]. Thanks to the prior language information it is possible to include features, such as word completion or automatic error correction that allow to improve accuracy and typing speed of the speller-BCIs [14]. Among the BCI spellers that include a language model, Gembler et al. [15] developed a dictionary-driven c-VEP BCI, and proved that the incorporation of word prediction allows to increase the performance of a c-VEP system.

In our work we propose an unsupervised classifier for a c-VEP BCI system exploiting only the language information. The proposed model extracts relative lags between successive characters from the VEP responses and then predicts the full word using a dictionary, eliminating the calibration phase. To the best of our knowledge, this is the first work proposing a word-prediction-based auto-calibration method for c-VEP BCIs.

This paper is organized as follows. Section 2 briefly presents the standard c-VEP BCI system and summarizes the main parts that characterize this type of BCI. Section 3 presents the proposed auto-calibration method. The experiments and results

are detailed, respectively, in Section 4 and Section 5, followed by the discussion and concluding remarks in Section 6 and Section 7.

## 2. The c-VEP BCI system

In a c-VEP BCI system, all characters flash simultaneously according to a predefined pseudo-random binary sequence circular-shifted by a different time lag [6] for each character. The sequences with maximum length, called m-sequences [16], are generally applied as stimulus sequences in c-VEP BCI. The m-sequence has specific properties of correlation: it is nearly orthogonal to its time shifted sequence and its auto-correlation function is close to a unit impulse function. Therefore, for a given character, the m-sequence generates a VEP response largely dependent on the characteristics of the stimulus code [4]. The VEP response recorded in the electroencephalogram (EEG) of the subject can be used as a template [17]. This template is obtained during a calibration phase at the beginning of each session.

During the calibration phase the user gazes at the reference character  $C_r$  and all the characters flash simultaneously for  $N$  stimulus cycles. The raw EEG data  $\mathbf{X}_r \in \mathbb{R}^{n \times c \times s}$  is recorded during  $n = 1, \dots, N$  stimulus cycles from  $c$  channels. The evoked response  $\mathbf{R} \in \mathbb{R}^{q \times s}$  of the reference character can be obtained averaging the time-windowed EEG data  $\mathbf{X}_r$  from  $N$  stimulus cycles from  $q$  channels. The number of channels  $q$  can be the same as  $c$  or a subset, if we choose to consider only the channels in which the evoked response is most prominent. To improve the signal to noise ratio (SNR) of the system the Canonical Correlation Analysis (CCA) is applied as spatial filter [18] to compute a reference template  $T_r(t)$  [6]. The goal of CCA is to find the two transformations  $W_X$  and  $W_S$  which maximize the correlation between the raw EEG data  $\mathbf{X}_r$  and the expected VEP response  $\mathbf{S}$ , that can be obtained by concatenating  $N$  times the evoked response of the reference character.

$$CCA(X, S) = \max_{W_X W_S} \frac{W_X^T \mathbf{X} \mathbf{S}^T W_S}{\sqrt{W_X^T \mathbf{X} \mathbf{X}^T W_X} \cdot \sqrt{W_S^T \mathbf{S} \mathbf{S}^T W_S}} \quad (1)$$

Then the spatially filtered EEG data  $x_n$  is obtained.

$$x_n = W_X \mathbf{X}_n \quad (2)$$

The reference template  $T_r$  is computed by averaging  $x_n$  over  $N$  stimulus cycles. The templates  $T_k$  of all other  $K$  characters in the virtual keyboard are obtained by shifting the reference template  $T_r$  by a specific time lag  $\tau_k = \tau \cdot k$ , where  $\tau$  is the time lag between the flashing sequences of two consecutive characters and  $k = 0, 1, 2, 3, \dots, K$  is the index of the corresponding character in the virtual keyboard.

$$T_k(t) = T_r(t - \tau_k) \quad (3)$$

The calibration phase is followed by a spelling phase. During the spelling phase the user can gaze at a character of his/her choice, called *target*, for a number of stimulus cycles  $N$ . Then, by template matching, it is possible to identify the target

that the subject is gazing at. The correlation  $\rho_k$  (4) between the spatially filtered EEG  $x = W_X \mathbf{X}$ , recorded during the online spelling phase, and each template  $T_k$ , obtained during the calibration phase, is computed. Hence, the system outputs the detected character  $k_{target}$  (5) by selecting the one that corresponds to the index of the template with the highest value of correlation.

$$\rho_k = \frac{x^T T_k}{\sqrt{xx^T \cdot T_k T_k^T}} \quad (4)$$

$$k_{target} = \arg \max_k \rho_k \quad (5)$$

### 3. Auto-calibration by word spelling and prediction

To tackle the limits of speller-BCIs caused by long and tedious calibration phases, we propose an auto-calibration method for a c-VEP BCI system, that allows to spell words and uses word prediction to avoid the traditional calibration phase. The framework of the method is shown in figure 1. The detailed explanation of the auto-calibration method (AC) is as follows.

- (i) The average response of the first character  $\mathbf{X}_{a1} \in \mathbb{R}^{c \times s}$  is computed over  $N$  stimulus cycles, where  $c$  is the number of channels and  $s$  the number of samples of the stimulus cycles.  $\mathbf{X}_{a1}$  is computed using the time-windowed EEG recorded while the user is gazing at the first character, but since the system has not been calibrated, the first character is unknown at this step. For the first character the relative position is 0.
- (ii) For the second character, the average response  $\mathbf{X}_{a2} \in \mathbb{R}^{c \times s}$  is computed over  $N$  stimulus cycles.  $\mathbf{X}_{a2}$  is shifted by  $(l \cdot \tau)$  time samples, where  $\tau$  is the time lag between two consecutive characters and  $l$  is the index of the corresponding character in the virtual keyboard. This produces  $L$  shifted averages  $\mathbf{X}_l$  (6),  $l = 0, \dots, L - 1$ , where  $L$  is the number of characters on the virtual keyboard.

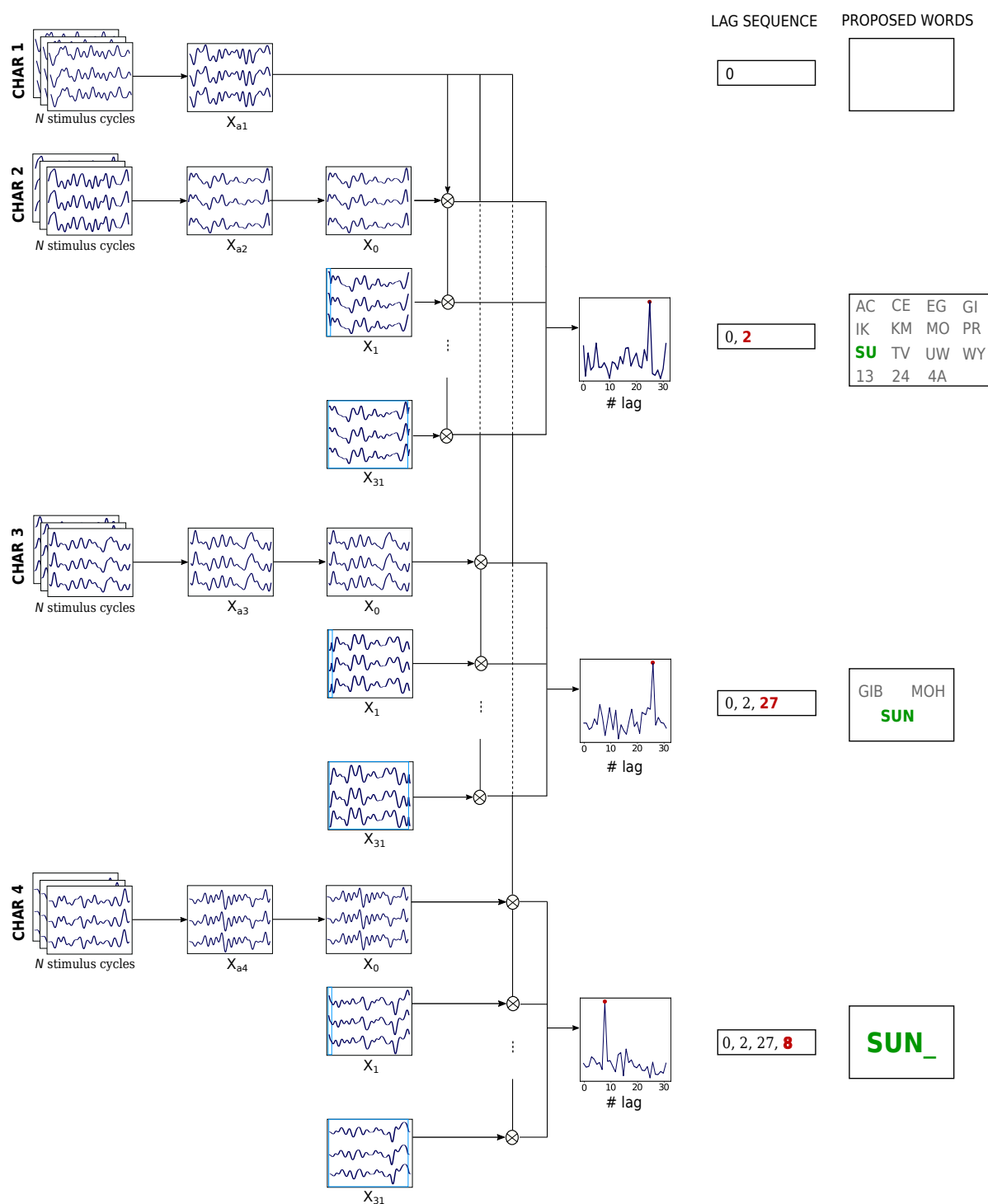
$$\mathbf{X}_l = \mathbf{X}_{a2}(t - (l \cdot \tau)) \quad (6)$$

- (iii) The correlation  $\rho_l$  between the initial average response  $\mathbf{X}_{a1}$  and the average response of the second character  $\mathbf{X}_l$  is computed.
- (iv) Using the lag  $l_{max}$  (7) which produces the maximum correlation between  $\mathbf{X}_{a2}$  and  $\mathbf{X}_{a1}$ , the relative position of the second character with respect to the first is computed.

$$l_{max} = \arg \max_l \text{corr}(\mathbf{X}_{a1}, \mathbf{X}_l) \quad (7)$$

- (v) Among all pairs of characters separated by  $l$  only retain those corresponding to the beginning of valid words within a dictionary.

This method is repeated for the following characters, until we are left with a single word. At that moment, we will have recovered the original letter, and the absolute position of  $\mathbf{X}_{a1}$  can thereafter be used during the computation of the time lag.



**Figure 1.** Outline of the auto-calibration method for a c-VEP BCI to spell the target word “SUN\_”.

In order to guarantee the uniqueness of each word belonging to the dictionary a character “\_” is added at the end of each word, in this way each word has a unique sequence of lags, independently of the type and size of the dictionary. Moreover it allows the system to understand when the spelling of a word is concluded.

The auto-calibration method can be improved considering not only the most correlated character, as explain at the step (iv), but also the second most correlated one. However the latter is taken into account only if its correlation value is higher than a specific percentage  $P_c$  of the correlation value of the most correlated character. The second most correlated character gives the lag  $l_{2,max}$ , as shown in figure 2.

The coefficient  $\varphi_{n,k}$  is introduced, for each character of the word excluding the first ( $n = 2, \dots, N_{chars}$ ) and for the  $k^{th}$  most correlated character.

$$\varphi_{n,k} = \frac{corr(\mathbf{X}_{a1}, \mathbf{X}_{l_{k,max}})}{corr(\mathbf{X}_{a1}, \mathbf{X}_{l_{max}})} \quad (8)$$

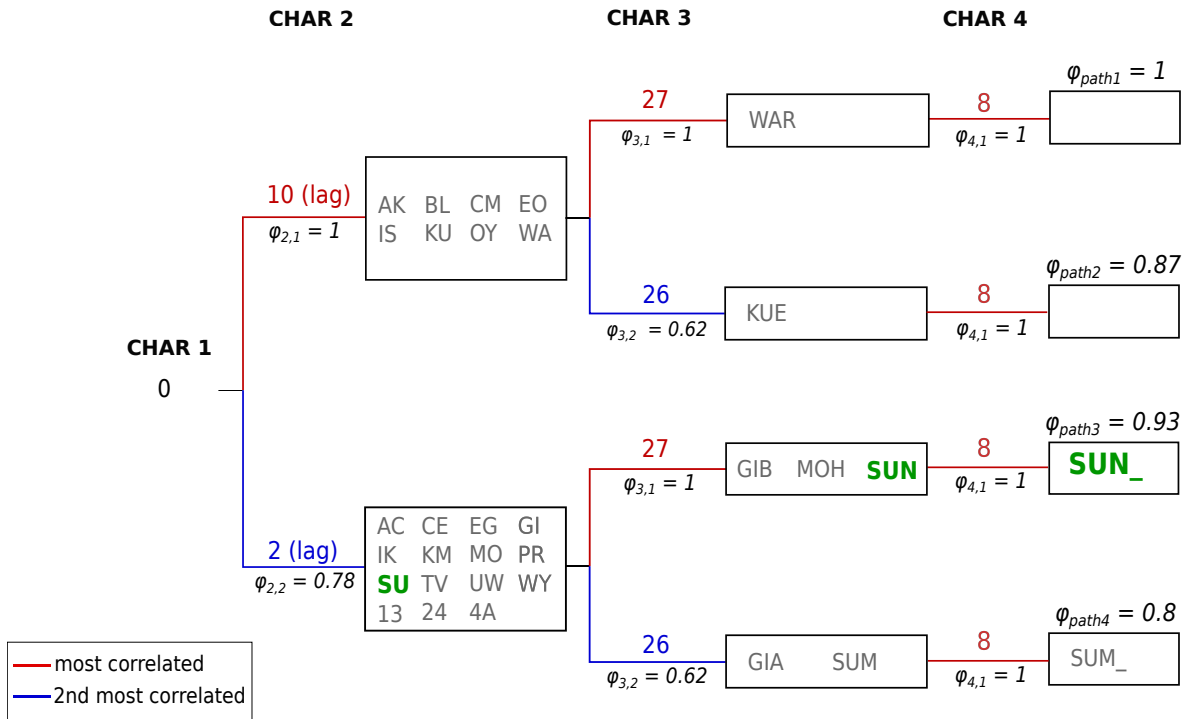
Clearly the coefficient  $\varphi_{n,1}$  is always equal to 1, because it corresponds to the character with highest correlation ( $l_{max} = l_{1,max}$ ). Then, we can use the coefficient  $\varphi_{n,2}$  of the character corresponding to  $l_{2,max}$  as a representation of the distance of its correlation with respect to the highest one, and we can apply the threshold  $P_c$  to choose whether it has to be considered as a spelling candidate or not. The system will give as output the word belonging to the dictionary with the highest value of  $\varphi_{path_i}$  (9), computed as the average of the  $\varphi_{n,1}$  and  $\varphi_{n,2}$  in the path  $i$  through the  $(N_{chars} - 1)$  characters. Each path  $i$  is obtained combining all the possible suggested lags ( $l_{max}$  and  $l_{2,max}$ ).

$$\varphi_{path_i} = \frac{\sum_{n \in path_i} \varphi_{n,1} + \sum_{n \in path_i} \varphi_{n,2}}{N_{chars} - 1} \quad (9)$$

#### 4. Data and experiments

We investigated data from the public dataset of Spüler et al. [17]. Nine healthy subjects participated in the c-VEP BCI experiment. Each subject took part in two identical sessions. For five subjects the session was performed in the same day, for the other four on different days. For a complete subject review refer to [17].

The EEG signal of the subjects was recorded from a Brainproducts Acticap system with 32 channels, with a g.tec g.USBamp (600 Hz sampling rate). They used 30 electrodes located at Fz, T7, C3, Cz, C4, T8, CP3, CPz, CP4, P5, P3, P1, Pz, P2, P4, P6, PO9, PO7, PO3, POz, PO4, PO8, PO10, O1, POO1, POO2, O2, OI1h, OI2h and Iz. The remaining two electrodes were used for electrooculography (EOG). The reference electrode was positioned at Oz and the ground electrode was positioned at FCz. Each session consisted of 10 stimulus cycles per target. In each run the subject had to spell 32 targets on a virtual keyboard twice, using the c-VEP BCI system, so in total one session consisted in 640 trials. The arrangement of the targets in the keyboard



**Figure 2.** Spelling of the word “SUN\_” applying the improved auto-calibration method (see Section 3). In red the lag  $l_{max}$  of the most correlated character and in blue the lag  $l_{2,max}$  of the second most correlated character, with their corresponding values of  $\varphi_{n,1}$  and  $\varphi_{n,2}$ , where  $n$  indicates the character position in the word. The figure shows that for characters 2 and 3 two lags are found, while for character 4 only the most correlated lag is considered. This is because the values of  $\varphi_{2,2}$  and  $\varphi_{3,2}$  are higher than the threshold  $P_c$ , while  $\varphi_{4,2}$  is lower. At the end, the system proposes the word belonging to the dictionary with the highest value of  $\varphi_{path}$ , that respects the lag sequences defined during the spelling.

was conformed to the principle of equivalent neighbors [6]. Each target was modulated by a 63-bit binary m-sequence with a stimulus presentation rate of 60 Hz, hence the length of a stimulus cycle was  $t_s = 63/60 = 1.05$  s. The modulation sequence of each target was shifted by 2 bits with respect to its preceding target, hence the time lag between two consecutive targets was  $\tau = 2/60 = 0.033$  s.

#### 4.1. Offline experiments

The signals were pre-processed using a Butterworth filter between 2 and 15 Hz. We simulated the spelling of four different groups of 5 words called *3-char*, *4-char*, *5-char* and *sentence*. Each group has a different number of characters per word and consequently a different total number of characters. The composition of the groups of words is detailed in table 1. The words belonging to *3-char*, *4-char* and *5-char* groups are randomly selected from a free corpus of ten thousand English words downloaded from the Wikipedia corpus, that was used as dictionary for the auto-calibration method. For the *sentence* group, the sentence chosen to be spelled was



**Table 1.** Composition of the groups of words applied in the experiments.

Word group	# total chars	# words	# chars per word
3-char	15	5	3
4-char	20	5	4
5-char	25	5	5
sentence	51	6	[5, 12]

“BRAIN\_COMPUTER\_INTERFACE\_LETS\_EVERYONE\_COMMUNICATE”. The dictionary was modified appending the character “\_” at the end of each word, as explained in Section 3.

The channels OI1h, OI2h and Iz are not considered for our experiments because the amplitude of the EEG signal is too low. We conducted experiments with different methods and we use the following convention to name them: standard calibration (C), explained in Section 2, the auto-calibration method considering only the most correlated character (AC1) and the auto-calibration method considering the two most correlated characters (AC2), explained in Section 3.

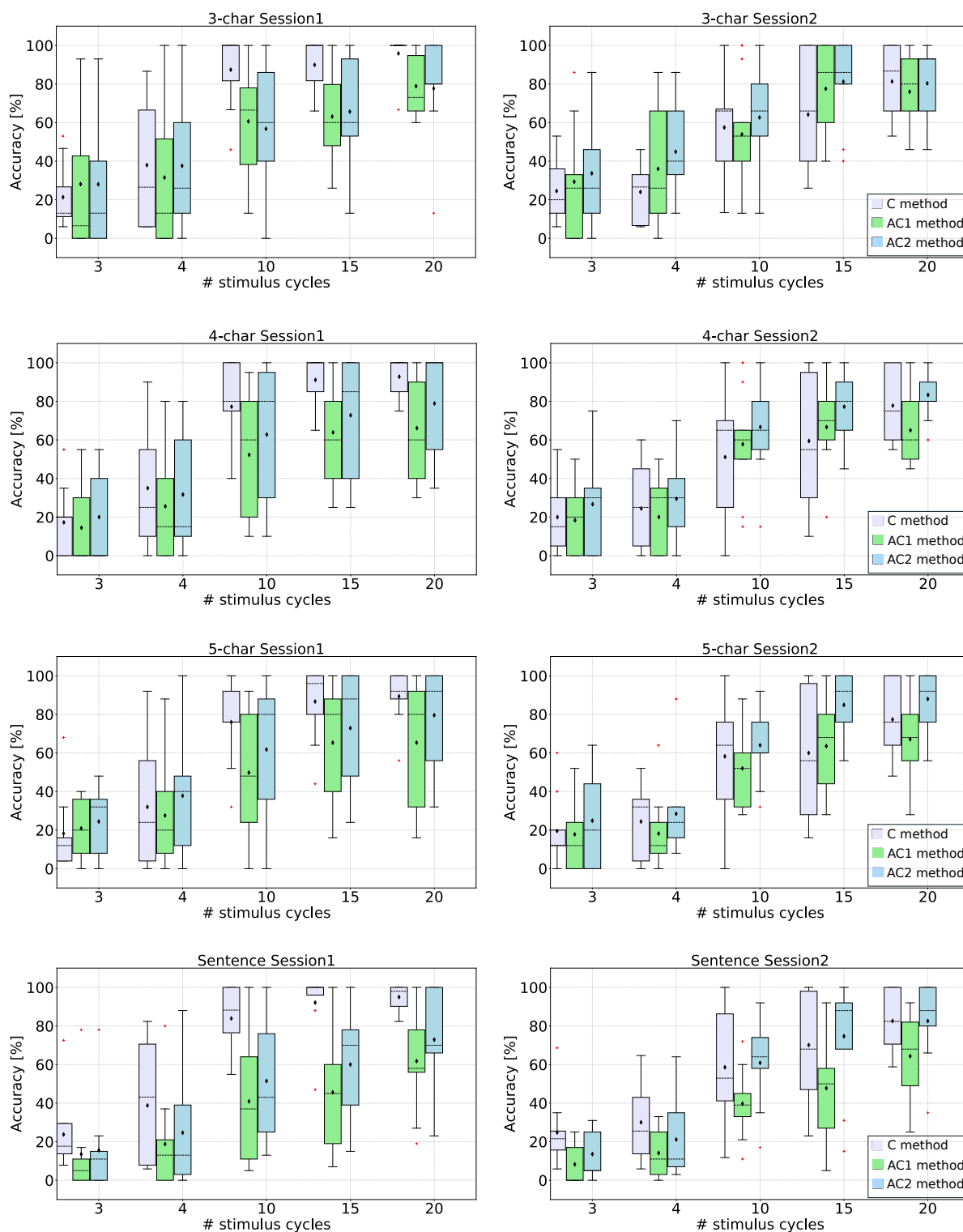
For the experiments with method C we performed the same pre-processing as applied in [17]. Then we applied the CCA spatial filter to  $N$  stimulus cycles of the character “A” to compute an average absolute response  $X_1$ , used as reference template. Finally, we followed the process described in Section 2. Each experiment was parameterized by the number of stimulus cycles, in two different ways. The former experiments evaluated the performance of methods C, AC1, AC2 with the same number of stimulus cycles for each group of words. The latter experiments evaluated the performance of AC2 method to spell the same groups of words with a number of stimulus cycles fixed to 10 for the first character and following the parameterized number of stimulus cycles for all the other characters belonging to each word. This can be considered as an option of the AC2 method, called AC2\* method.

## 5. Results

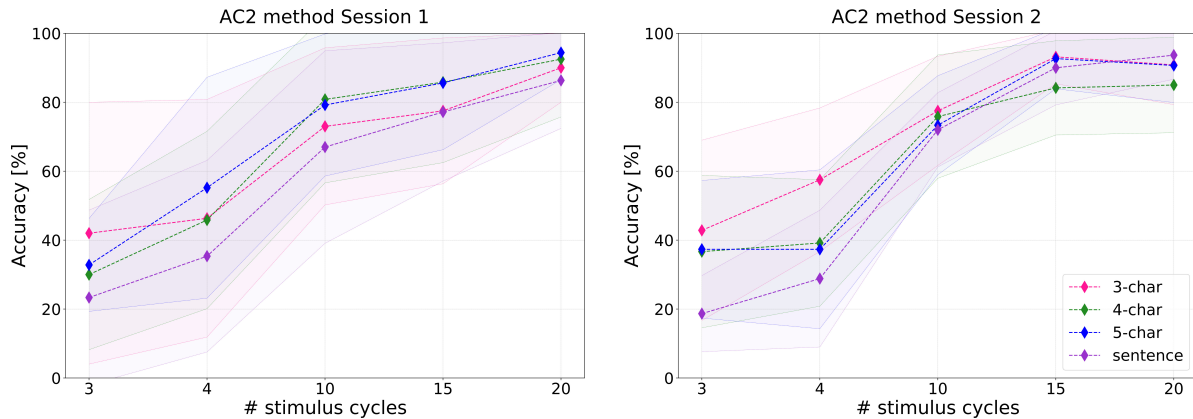
### 5.1. Comparison between methods

Figure 3 shows the accuracy reached by all subjects of the dataset to spell different groups of words applying the C method, explained in Section 2, the AC1 and AC2 methods explained in Section 3. The composition of each group of words is described in table 1.

The accuracy is computed as the ratio of characters well spelled parameterized by the number of stimulus cycles. The results reached in Session 1 show that the average accuracy achieved with the C method is higher than the average accuracy reached applying both the AC1 and AC2 methods, except for the *3-char*, *4-char* and *5-char* groups for which the average accuracy with the AC2 method is higher for 3 and 4



**Figure 3.** Boxplots of accuracy reached by all subjects for spelling four different groups of words parameterized by the number of stimulus cycles in Session 1 and Session 2, applying the standard calibration (C) in violet, the auto-calibration method considering only the most correlated character (AC1) in green, and the auto-calibration method considering the two most correlated characters (AC2) in light blue. The box edges represent quartiles, diamond markers represent the mean and dash lines denote the median. Outliers are marked with red crosses.



**Figure 4.** Evolution of AC2 method performance over sessions. Diamond markers represent the average accuracy reached across repetitions for a subset of subjects, with the *3-char* group (pink), the *4-char* group (green), the *5-char* group (blue) and the *sentence* group (purple). The transparent zones indicate the standard deviation values. The subset of subjects includes all subjects that reached an average accuracy higher than 30 % across stimulus cycles and groups of words.

stimulus cycles, even if the dispersion is larger with respect to C method.

On the contrary, in Session 2 the average accuracy attained with the C method is lower than the average accuracy reached with the AC2 method for all the tested groups of words and for all the numbers of stimulus cycles. The only exception is for the spelling of a sentence with 3 and 4 stimulus cycles, for which the average accuracy of the C method is the highest.

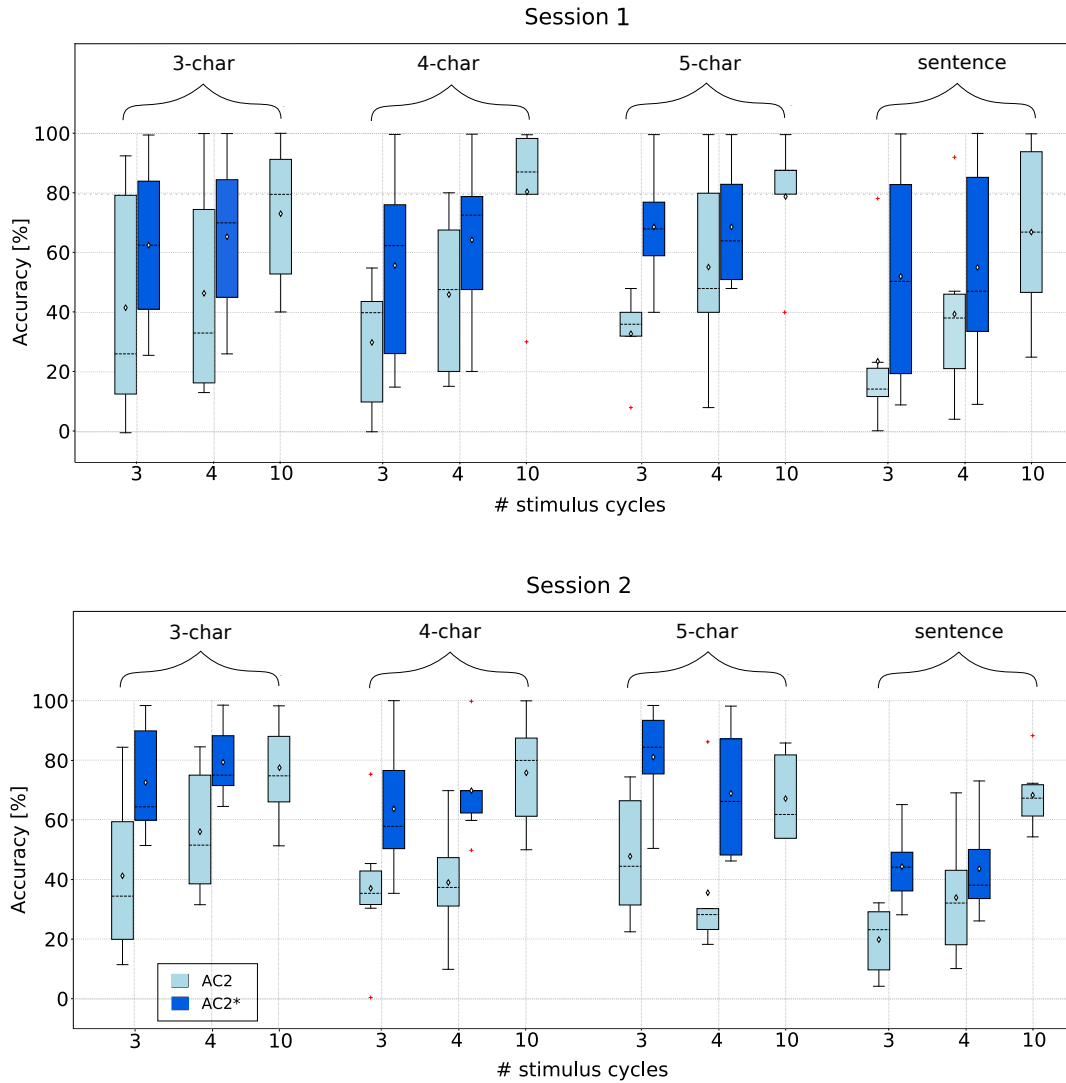
Generally, comparing the results reached across sessions, we can observe that the average accuracy improved increasing the number of stimulus cycles, for all methods, and the dispersion in Session 1 is larger with respect to Session 2, in particular for AC1 and AC2 methods.

Moreover it is evident that the AC2 method performed better than the AC1 method for all the tested cases reported in figure 3.

## 5.2. Method AC2

To evaluate the performance of the AC2 method we selected the subset of six subjects with good performance, reaching an average accuracy higher than 30 % across the number of stimulus cycles and the groups of words. The subjects not included in this subset are AE, AA, AH, for more details refer to [17].

In figure 4 is presented the performance reached with the AC2 method across sessions. The trends in figure 4 confirm the improvement of the performance increasing the number of stimulus cycles, except for the *3-char* and the *5-char* groups, for which the performance decreased in Session 2 from 15 to 20 stimulus cycles. Specifically for the *3-char* group the average accuracy is equal to 93 % with 15 stimulus cycles and equal to 85 % with 20 stimulus cycles, while for the *5-chars* group the corresponding values are respectively 95 % and 91 %. Nevertheless, it is evident that the results in



**Figure 5.** Boxplots of accuracy reached with the AC2 method (light blue) and the AC2\* method (dark blue) in Session 1 and Session 2. The box edges represent quarterlies, diamond markers represent the mean and dash lines denote the median. Outliers are marked with red crosses.

Session 2 are less spread than in Session 1, as we can also observe in figure 3.

In Session 1 the *4-char* and the *5-char* groups obtained better performance on average with respect to other groups. Instead in Session 2 the accuracy reached for the *3-char* group is higher than the other groups, especially for 3 and 4 stimulus cycles. On the contrary we can observe that for more than 4 stimulus cycles the performance for all groups is really close, except for the *4-char* group with 15 and 20 stimulus cycles. Comparing the two plots it is clear that the variability of the performance is larger for low numbers of repetitions (3 and 4 stimulus cycles) across groups and sessions.

In order to evaluate a possible improvement of the performance with a lower number of stimulus cycles and, at the same time, avoiding a very long stimulation time, the methods AC2 and AC2\* are compared in the boxplots shown in figure 5. The

performance of both methods is compared across sessions on the same subset of best subjects evaluated in figure 4. The average accuracy increases with the AC2\* method, reaching values always higher than the ones obtained with the AC2 method, usually close to the accuracy reached with 10 stimulus cycles. This is especially evident for the Session 2, with the exception of the *sentence* group, indicating a particular benefit for the performance with a lower number of stimulus cycles.

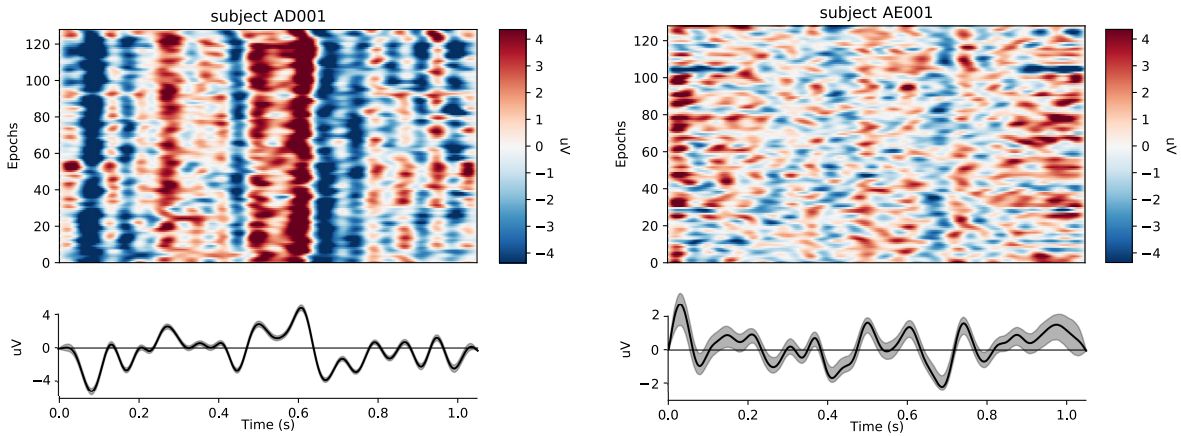
## 6. Discussion

The objective of our work is to develop a method for a c-VEP BCI with no calibration phase. The AC1 and AC2 methods proposed in our work were tested for four different groups of words to evaluate the stability of the method when varying the number of characters of a single word and the total number of characters to spell.

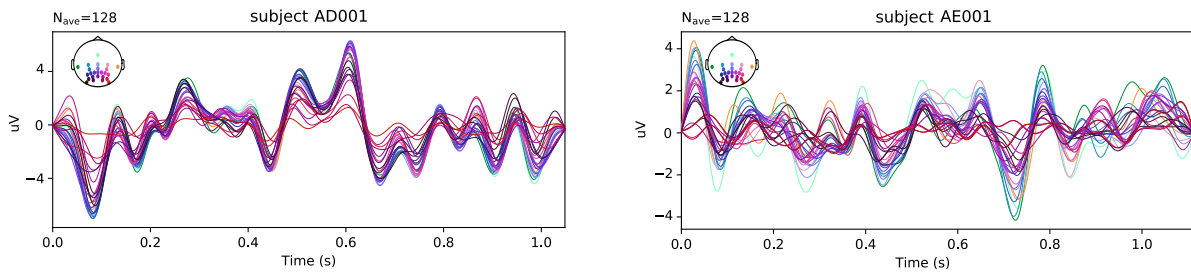
The results shown in figure 3 and, in particular, figure 4 demonstrate that there is no clear prevalence in terms of number of stimulus cycles with respect to the number of characters in the word. However, it appears that the performance is lower with a low number of stimulus cycles for a high number of characters, but when increasing the number of stimulus cycles, the performance is comparable for all the groups of words, except for *sentence* group. Indeed it is evident that the performance for the *sentence* group is lower with respect to other groups of words. In this case we simulated the spelling of a sentence with a total of 51 characters and 6 words with a maximum of 12 characters per word. We chose this sentence with long words in order to evaluate the efficacy of the method in a more complex scenario. In fact with longer words the risk of not reaching the correct spelling of the target word increases, because the probability of not correctly identifying a character increases and, consequently, the wrong identification of the word in the dictionary. This justifies the lower performance of the spelling of the *sentence* group compared to other groups of words.

This aspect may be improved including, for example, the error correction [14] in an online implementation of the method, to reduce the wrong target word identification caused by the erroneous selection of single characters. Furthermore, in online applications, the subject may initially spell shorter words in order to become familiar with the system, improving the ability to use it and consequently improving the repeatability of the VEP responses, thus limiting errors also when spelling of long words.

The repeatability of the VEP responses is fundamental for c-VEP BCI systems and in particular to achieve good performance with the proposed auto-calibration method. The results in figure 3 show a consistent dispersion with all methods, especially in Session 1, demonstrating that for some subjects it is possible to obtain very good results also with a lower number of stimulus cycles, but for others it is difficult to reach good results. Figure 6 shows the repetitions of each character realigned with respect to the unshifted m-sequence for one of the best and one of the worst performers in Session 1, in terms of accuracy (subject AD and subject AE, respectively). We can notice visible bands of activation (red zone) for all the repetitions for the subject AD



**Figure 6.** The first four stimulus cycles of each target character realigned with respect to the unshifted m-sequence and the average realigned response with the standard deviation in Session 1, for subject AD and subject AE, respectively.



**Figure 7.** Average response of each channel over the first four realigned stimulus cycles of each character in Session 1, for subject AD and subject AE, respectively.

and an average with a low standard deviation. On the other hand, for subject AE it is not possible to define specific zones of activation because there is a visible variability between the repetitions, also demonstrated by the standard deviation computed over all stimulus cycles, which is greater than for subject AD. Moreover figure 7 shows the average response over all the realigned repetitions for each channel for the same two subjects. We can notice that for subject AE some electrodes in the occipital part did not work well. For sure they influenced the low accuracy reached for the subject in the offline experiments. Instead for subject AD all the electrodes have the same response shape, with a difference in amplitude. So we can affirm that the performance of the auto-calibration method is influenced by the quality of the data. In particular it is important to have a good VEP response repeatability between the stimulus cycles for each character.

The quality of the data is fundamental to build a high performance BCI system. There are many factors that can influence the EEG variability. Some of them can be controlled, for example the design of the system and the hardware used for the acquisition of EEG signals, the application of specific algorithms and the control of some external factors like noise. But we can not control the state of subject, like the loss

of concentration, visual fatigue, distraction and motivation [19]. The EEG variability influences the performance of the BCI system from a session to another. With the auto-calibration method it is possible to face the problem of inter-session variability because for each session the system is not calibrated with signals acquired in a different time, but it is based on the data acquired during the specific session considering all the possible factors that can influence the it. This explains why with a lower number of stimulus cycles, for many subjects, the accuracy reached with the auto-calibration method is higher than the accuracy reached with the standard calibration method. However, we can notice that the performance with the calibration method increased with the number of stimulus cycles. Indeed, generally, in c-VEP BCI the performance increases with a longer calibration phase [6, 20].

Analyzing the results showed in figure 3 it is evident that with the auto-calibration method the accuracy increased when considering the two most correlated characters. A possible explanation is that if we analyze the stimulus modulation applied to collect the dataset [17], each character flashes according to a binary sequence of 63-bit and the sequence generated for each character is 2-bit circularly shifted with respect to the sequence of the consecutive characters [6]. This means that the time shift between two consecutive characters is  $2 \text{ bit}/60 \text{ Hz} = 0.033 \text{ s}$ . Therefore the timing precision of the VEP response is fundamental to distinguish the response of each character. In fact it is possible that some stimulation parameters can influence the VEP response, such as stimulus proximity and the lag between two adjacent stimuli [20]. The subject could be distracted by the interference caused by consecutive characters in the keyboard which adds the difficulty to focus the attention on the target character. This is also proved by the fact that when applying the AC2 method the couple of lags detected corresponds to two consecutive character, as shown for example in figure 2.

### 6.1. Performance evaluation

To compare the potential effectiveness of the method with the actual state of art of online c-VEP BCI, we compute the theoretical information transfer rate (ITR) [1], following equation (10). The ITR returns the amount of information contained in a selection in bits per minute (bpm).

$$ITR = \left( \log_2 N + P \log_2 P + (1 - P) \log_2 \left( \frac{1 - P}{N - 1} \right) \right) \cdot \left( \frac{60}{T} \right) \quad (10)$$

Where  $N$  is the number of classes, equal to 32, which is the number of characters in the virtual keyboard,  $P$  is the accuracy (the percentage of characters correctly spelled) and  $T$  is average time required for the prediction. We computed the theoretical ITR for the spelling of all word groups comparing the values reached with the AC2 method and the AC2\* method, across sessions, for the subset of best subjects.

The time  $T$  was computed as the average time required to select a character, taking in account the number of stimulus cycles for the total number of characters for each group of words. For instance for the *3-char* group the time  $T$  is computed as the number

**Table 2.** Theoretical average ITR (bpm), computed for AC2 method for 3, 4 and 10 stimulus cycles (s.c.) and AC2\* method for 3 and 4 stimulus cycles (s.c.).

	ITR (bpm) Session 1					ITR (bpm) Session 2				
	3 s.c.		4 s.c.		10 s.c.	3 s.c.		4 s.c.		10 s.c.
	AC2	AC2*	AC2	AC2*	AC2	AC2	AC2*	AC2	AC2*	AC2
<b>3-char</b>	25.1	27.6	25.4	27.1	17.3	26.6	33.1	26.6	35.7	18.5
<b>4-char</b>	16.7	23.3	21.8	26.4	20.7	21.1	25.9	21.1	28.5	18
<b>5-char</b>	15.8	28.9	15.1	28.2	20.1	20.3	29.6	20.3	29.7	16.7
<b>sentence</b>	12.7	22.1	16.4	20.3	15.8	6.7	13.3	6.7	11	16

of stimulus cycles multiplied by the duration of one stimulus cycle (1.05 s) for the AC2 method. Instead for the AC2\* method we computed an average  $T$ , taking in account that the first character is flashed for 10 stimulus cycles, thus for each group of words the specific time  $T$  was computed following the parameter reported in table 1.

The results listed in table 2 show that even if the duration increased with the AC2\* method with respect to the AC2 method with 3 and 4 repetitions, nevertheless the ITR increased thanks to the improvement of the average accuracy. Indeed the ITR is a parameter to evaluate BCI performance that takes in account both speed ( $1/T$ ) and accuracy, in this case the benefit reached in term of accuracy by the AC2\* method overcompensates the increase of time required to spell a character with respect to the AC2 method. On the other hand, we can notice that with 10 stimulus cycles the duration increased excessively and, although the average accuracy is higher, the ITR is lower with respect to the ITR computed for 3 and 4 repetitions.

The average theoretical ITR, listed in table 2, demonstrates that both proposed methods, AC2 and AC2\*, could compete with the state-of-art of BCI speller with language model [13, 14]. Furthermore, considering for example a subject who reaches an accuracy of 100 % in the spelling of a sentence (figure 5 Session 1), we can obtain with the AC2\* method an ITR of around 54 bpm with 3 stimulus cycles and around 52 bpm with 4 stimulus cycles. These theoretical results can effectively compete with, for example, the work proposed by Gembler et al. [15], in which they reported an average ITR value of 57 bpm for the spelling of a sentence in a c-VEP BCI.

## 6.2. Further remarks

The proposed auto-calibration is one component contributing to the performance of c-VEP. At least two more improvements should be made. First, the language model should be integrated into an online application, including specific features such as error correction and word completion [14]. Second, different strategies may improve the repeatability of the VEP responses, which is crucial for the performance of the method. One way is to investigate the influence of the stimulus modulation in order to design an appropriate interface for the auto-calibration c-VEP BCI system. Indeed the stimulus modulation is crucial to obtain a high performance c-VEP BCI. Many



studies [20, 21, 22] investigated several parameters to understand their influence in a c-VEP system but analyzing these studies is clear that it is not possible to define an universal optimal stimulus parameter setting suitable for each BCI user in a c-VEP BCI system.

## 7. Conclusion

The method presented in our work is a proof of concept of auto-calibration for a c-VEP BCI speller. The method exploits only prior language information and not prior information from other training subjects. The intuition of the method is to use the fundamental property that characterizes the VEP response elicited from pseudo-random stimulus sequence. The response of each character is identical, but circularly-shifted with a specific time sample lag. The method is based on the extraction of the time lag of the VEP response of each character with respect to the VEP of the first character during the spelling of the target word. After the spelling of each character, the system finds within the dictionary all the possible words whose letters respect the sequence of relative lags. Offline experiments were simulated to test the proposed auto-calibration method compared to the standard calibration method. We demonstrated that our method can compete with the current state-of-art of BCI spellers [14, 15].

Furthermore, the proposed method shows promising results to develop a calibration-free c-VEP BCI. This opens also a new perspective to the diffusion of BCI systems more user-friendly and adaptable to each user.

## References

- [1] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain-computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–791
- [2] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–523
- [3] Blankertz B, Tangermann M, Vidaurre C, Fazli S, Sannelli C, Haufe S, Maeder C, Ramsey L E, Sturm I, Curio G *et al.* 2010 The berlin brain-computer interface: non-medical uses of bci technology *Front. Neurosci.* **4** 198
- [4] Bin G, Gao X, Wang Y, Hong B and Gao S 2009 Vep-based brain-computer interfaces: time, frequency, and code modulations [research frontier] *IEEE Comput. Intell.* **4** 22–26
- [5] Cecotti H 2010 A self-paced and calibration-less ssvep-based brain-computer interface speller *IEEE Trans. Neural Syst. Rehabil. Eng.* **18** 127–133
- [6] Bin G, Gao X, Wang Y, Li Y, Hong B and Gao S 2011 A high-speed bci based on code modulation vep *J. Neural Eng.* **8** 025015
- [7] Millán J, Rupp R, Müller-Putz G, Murray-Smith R, Giugliemma C, Tangermann M, Vidaurre C, Cincotti F, Kubler A, Leeb R *et al.* 2010 Combining brain-computer interfaces and assistive technologies: state-of-the-art and challenges *Front. Neurosci.* **4** 161
- [8] Krauledat M, Tangermann M, Blankertz B and Müller K R 2008 Towards zero training for brain-computer interfacing *PloS One* **3** e2967
- [9] Jayaram V, Alamgir M, Altun Y, Scholkopf B and Grosse-Wentrup M 2016 Transfer learning in brain-computer interfaces *IEEE Comput. Intell. Mag.* **11** 20–31

- [10] Lu S, Guan C and Zhang H 2009 Unsupervised brain computer interface based on intersubject information and online adaptation *IEEE Trans. Neural Syst. Rehabil. Eng.* **17** 135–145
- [11] Lotte F 2011 Generating artificial eeg signals to reduce bci calibration time *5th International Brain-Computer Interface Workshop* pp 176–179
- [12] Kindermans P J, Verschore H, Verstraeten D and Schrauwen B 2012 A p300 bci for the masses: Prior information enables instant unsupervised spelling *Advances in Neural Information Processing Systems* pp 710–718
- [13] Mora-Cortes A, Manyakov N V, Chumerin N and Van Hulle M M 2014 Language model applications to spelling with brain-computer interfaces *Sensors* **14** 5967–5993
- [14] Speier W, Arnold C and Pouratian N 2016 Integrating language models into classifiers for bci communication: a review *J. Neural Eng.* **13** 031002
- [15] Gemblar F and Volosyak I 2019 A novel dictionary-driven mental spelling application based on code-modulated visual evoked potentials *Computers* **8** 33
- [16] SW G 1982 *Shift Register Sequences* (Aegean Park Press, Laguna Hill)
- [17] Spüler M, Rosenstiel W and Bogdan M 2012 Online adaptation of a c-vep brain-computer interface (bci) based on error-related potentials and unsupervised learning *PloS One* **7** e51077
- [18] Spüler M, Walter A, Rosenstiel W and Bogdan M 2014 Spatial filtering based on canonical correlation analysis for classification of evoked or event-related potentials in eeg data *IEEE Trans. Neural Syst. Rehabil. Eng.* **22** 1097–1103
- [19] Kleih S and Kübler A 2015 Psychological factors influencing brain-computer interface (bci) performance *SMC, 2015 IEEE Intern. Conf.* pp 3192–3196
- [20] Wei Q, Feng S and Lu Z 2016 Stimulus specificity of brain-computer interfaces based on code modulation visual evoked potentials *PloS One* **11** e0156416
- [21] Isaksen J, Mohebbi A and Puthusserypady S 2017 Optimal pseudorandom sequence selection for online c-vep based bci control applications *PloS One* **12** e0184785
- [22] Turi F and Clerc M 2019 Adaptive parameter setting in a code modulated visual evoked potentials bci *8th Graz Brain-Computer Interface Conference*