Novel Sequential BCI Speller based on ERPs and Event-Related Slow Cortical Potentials

Accepted for publication in the Journal of Neural Engineering.

http://doi.org/10.1088/1741-2552/add772

Novel Sequential BCI Speller based on ERPs and Event-Related Slow Cortical Potentials

Riccardo Poli, Ahmet Can Mercimek, Caterina Cinel

Brain-Computer Interfaces and Neural Engineering Laboratory,
School of Computer Science and Electronic Engineering,
University of Essex, UK

E-mail: `rpoli@essex.ac.uk`

Abstract. *Objectives.* One of the most effective Brain-Computer Interfaces (BCI) spellers, Donchin and Farwell's matrix speller, uses visual stimulus presentation and the oddball effect, eliciting P300 event-related potentials to rare and randomly presented stimuli of interest. Although proposed almost 4 decades ago, most BCI spellers still rely on this principle and the original matrix speller design although some of the issues that affect oddball spellers have progressively been addressed over the years with significant, but very incremental, performance improvements.

Farwell and Donchin seminal paper suggested the future possibility of abandoning the oddball paradigm, for a regular/periodic presentation pattern which they predicted might produce a Contingent Negative Variation (CNV) thus improve speller performance. However, this has never been investigated.

Building on our past research on a BCI for cursor control which adopted a periodic stimulation protocol, here we explore whether a periodic presentation pattern could be a viable alternative to the oddball paradigm in a BCI speller.

Approach. We tested the periodic presentation principle in a BCI speller where 36 letters are organised around a circle and are highlighted sequentially, and compared it to the original matrix speller at two stimulus presentation rates.

Main Results. Our periodic speller produces not only clear P300s, but also equally clear CNVs, as postulated by Farwell and Donchin, as well as other Slow Cortical Potentials (SCPs). At the higher stimulation rate, this leads to significantly higher AUC, classification accuracy, ITR and utility w.r.t. Donchin's speller.

Significance. Our findings suggest that periodic stimulation can not only produce clear P300s but also a variety of event-related SCPs, leading to significant performance improvements over Donchin's paradigm. This work opens new avenues for BCI spelling where ERPs are combined with naturally-triggered (rather than trained) SCPs, that will hopefully result in more efficient communication systems for individuals with severe motor impairments.

1. Introduction

Brain-Computer Interface (BCI) spellers — systems that enable users to communicate by directly translating brain signals into text —– represent the most extensively researched area within BCIs. This is evidenced by the publication of hundreds of articles, including numerous recent reviews [1, 2, 3, 4, 5, 6, 7, 8].

This all started with the seminal work by Farwell and Donchin [9] (more on this below) based on the oddball paradigm and Event Related Potentials (ERPs) [10], and many modern spellers are still based on ERPs, including the speller proposed in this work. For this reason, in our literature review in the subsections below, we will focus on ERP-based spellers.

However, other brain waves — such as Slow Cortical Potentials (SCPs) [11] (slow shifts in the EEG signal, associated with cortical excitability) and Steady State Visually Evoked Potentials (SSVEPs) [12] (brain responses to repetitive visual stimulation) — and other perceptual modalities — such as auditory [13, 14] and tactile [15, 16] — and their hybridisations (e.g., systems relying on both ERPs and SSVEPs [17, 18], or multiple modalities, e.g., visual and auditory [19]) have been explored, too.

1.1. ERPs, Oddball Paradigm and P300s

ERPs are relatively well defined EEG variations which are elicited by a stimulus or event (visual, auditory, or somatosensory). They include an exogenous response, caused by early, perceptual processing of the stimulus, as well as a later endogenous response, which is a reflection of higher level cognitive processing of the stimulus, such as attention [10, 20].

The P300 wave is an endogenous ERP with a typical latency of 300 to 600 ms, which is elicited by rare and/or significant stimuli, provided the user attends to such stimuli [21, 22, 23]. P300s are often studied via the *oddball paradigm* [24, 25, 26], which involves presenting a series of stimuli, where a rare deviant stimulus is randomly interspersed within “standard” stimuli.

The amplitude and latency of the P300 depend on several factors [23]. For instance, the P300 amplitude increases as target probability decreases (see [20] for a review), and is also positively correlated with the Inter-Stimulus Interval (ISI: the time interval between the end of a stimulus and the beginning of the following stimuli) and Stimulus Onset Asynchrony (SOA: the time interval between the beginning of two consecutive stimuli) [27, 28, 29]. There is also a positive correlation between P300 amplitude and the number of non-target stimuli preceding a target (e.g., [30, 31]). However, some studies suggest that the Target-to-Target Interval (TTI) is the true factor underlying the P300 amplitude variations attributed to target probability, ISI, SOA and number of preceding non-targets [32, 33].

1.2. Row-Column ERP Spellers

As indicated above, the first speller based on ERPs was proposed by Farwell and Donchin [9] and is alternatively known as the *matrix speller*, the *row-column speller* or *Donchin's speller*. In this speller, the letters of the alphabet are organised in a matrix, the rows and columns of which are flashed/highlighted one at a time but in random order without replacement, i.e., a row or column is not allowed to flash again until all others have flashed. Participants need to focus their attention on the letter they want to spell and mentally count the number of times the row or column containing such a letter are flashed. Because flashes of the target letter are relatively rare within a sequence that also include flashes of non-target rows or columns, they elicit detectable P300s, making it possible to identify the target letter. ‡

It was already recognised in [9] that, in ERP-based spellers, the best P300 recognition accuracy would be obtained by temporally spacing the target stimuli in such a way that their ERPs minimally overlap, i.e., with long ISIs. However, due to the noise inherent in EEG, even with long ISIs, P300 classification algorithms will have a fairly low accuracy, which would make reliable spelling impossible. For this reason, in order to bring accuracy to acceptable levels (80% and 95% were considered in [9]) it is necessary to average ERP classifier outputs (more specifically, the scores produced by the classifiers) over multiple repetitions of stimulus presentation. Acceptable accuracy levels can be obtained by using shorter-than-ideal ISIs (and, so, lower single-trial accuracy) but more repetitions, longer ISIs but fewer repetitions, or anything in between. For a fixed ISI, normally increasing the number of repetitions increases accuracy, but reduces the communication bandwidth, a.k.a. the *Information Transfer Rate* (ITRs) [34, 35].

An argument put forward in [9] in favour of a row/column matrix speller is that it would present an advantage over a *pure oddball speller* (where the letters of the alphabet would be presented or highlighted one at a time in random order): it required participants to wait on average a shorter period of time (a smaller number of non-target stimuli) between flashes of the character they are trying to enter (the target stimulus), while with a pure oddball speller participants would have on average to wait until all other characters in the alphabet (non-target stimuli) have flashed. However, in the light of the need for averaging over multiple trials discussed above, it is not clear whether with suitably short ISIs, even a pure oddball speller might provide acceptable accuracies and ITRs (more on this below).

An issue with all Donchin's spellers is that, because rows and columns are randomly flashed, the TTI varies greatly, and so do, as indicated in Section 1.1, the P300 amplitudes. Therefore, the P300s elicited by target row/column flashes are greatly variable, reducing the performance of classification algorithms (unless explicitly designed to accommodate and exploit this [36]). Another issue of such spellers is *perceptual errors*, such as attentional blink [37] and repetition blindness [38], whereby flashes in rapid succession are less easily perceived and, consequently, trigger no (or a very small) P300s [39, 40].

Over the years a number of variations and improvements of Donchin's speller have been proposed. Some examples of variations include flashing pseudo-random patterns of letters instead of rows and columns [41, 42], replacing the characters with

---

‡ This is the letter at the intersection between the row where a classifier trained to identify P300s produces the highest score out of all the rows flashed, and the column where the classifier produces the highest score out of all the columns flashed.

familiar faces, own faces, or both faces and graphical symbols concurrently instead of highlighting them [43, 44, 45, 46], the use of different changes to letters other than flashing [47], the flashing of small squares at the margins of rows/columns rather than whole rows/columns [48], assigning a colour to each character and modulating it instead of flashing [49], changing the dimensions of the symbols and their distance [50, 51, 52], using three-dimensional character visualisation [53, 54], partitioning the matrix into submatrices [55], and many more. However, below we will focus on reviewing only those spellers that are relevant for the work proposed in this article.

1.3. Rapid Serial Visual Presentation Spellers

A pure oddball speller, based on a Rapid Serial Visual Presentation (RSVP) approach, where random characters are very briefly visually presented, *one at a time, in the middle of a display*, was proposed in [56] and later improved in [57]. Two SOAs were tested: 83ms (in both articles) *vs* 116ms in [57] and 133ms in [56]. Characters were displayed either in black or in colour on a grey background. Results were promising, with ITRs being above 10bits/min. A similar RSVP approach was used in [58], with a SOA of 166ms but without colour, and compared with Donchin's approach. The RSVP speller had ITRs of below 8bits/min, which are comparable with those reported in [57]. So, while these pure-oddball RSVP spellers present the advantage of not requiring gaze control, which is particularly helpful for individuals with severe motor disabilities who may be unable to saccade, they appear to be not as performant as row-column spellers, corroborating the argument put forward in [9] about the superiority of the row-column approach over pure oddball spellers.

Interestingly, much better performance (e.g., an ITR of approximately 20 bits/min) was obtained in [59] by hybridising an RSVP speller with Donchin's approach. In this speller, the characters from an alphabet of 36 are shown in triplets in the middle of the screen for 250ms at each stimulus presentation. This process is repeated for 36 triplets. So, each character is presented three times. However, no character is ever presented with the same two other characters within a sequence of 36 triplets. So, if one extracts from this sequence the triplets where a P300 was detected, the only character they will have in common is the target character.

1.4. Single-Display Spellers

Another pure oddball speller, called *Single Display* (SD) speller, was proposed in [60] where the characters of a Donchin's speller (with the typical 6 × 6 organisation) were flashed *one at the time* (rather than flashing whole rows or columns). The stimuli sequences were random but avoided neighbouring characters flashing immediately after one another. A very short ISI of 60ms was used, corresponding, however, to a rather long mean TTI of 2,160ms, which would be expected to produce large P300s. Indeed, the reported single-trial accuracy was very good (53.4%) suggesting a potentially high ITR (but this was not reported). However, a later comparison between this same paradigm and a region-based speller revealed that the former (i.e., [60]) had a peak ITR of 26 bits/min (obtained with 3 repetitions) [61].

Guan *et al.* [60] SD approach was modified and tested in a recent study [62], where a 4×10 matrix of characters was used. Here, however, characters were flashed/highlighted for 100ms, but flashing a new one every 30ms. So, there were up to four highlighted characters on the screen at any given time. In addition, this protocol was compared with one where stereo visual stimuli were used and, so, characters were perceived as both flashing and moving cubes in 3–D space. In either protocol, all characters were highlighted within $30 \times 40 = 1,200$ms. While impressive accuracies and ITRs were graphically reported in [62], there appear to be inconsistencies in such data.§

Finally, another SD approach called the *Lateral Single Character* (LSC) speller was proposed in [63], which was compared to matrix spellers. In the LSC speller, letters were arranged in two symmetrical groups on the left and right side of the screen respectively and flashed alternately and pseudo-randomly between the left and right sides. By dividing the targets into two separate groups, the LSC speller effectively avoided the issues (commonly associated with matrix spellers) of crowdedness and the influence of flashing non-target letters adjacent to the target. The LSC paradigm demonstrated superior performance compared to the standard matrix speller, proving effective for patients with ALS and other neuromuscular diseases, achieving an ITR of 26.11 bits/min and an accuracy rate of 89.90%.

1.5. Slow Cortical Potentials

Slow Cortical Potentials (SCPs) have been used as an alternative brain signal to control BCIs [64, 65, 11, 66]. SCPs are gradual shifts in the electrical activity of the brain cortex, measured by EEG, that occur over extended periods ranging from 300ms to several seconds, characterised by very low frequencies below 1 Hz[67, 64]. In the aforementioned studies, subjects were trained with neurofeedback to modulate their SCPs, for the purpose of controlling a BCI. The training required weeks.

SCPs are considered to reflect the overall level of cortical excitability, often associated with processes like attention, expectancy, cognitive preparation, decision-making, and motor planning/intention, with *negative SCPs generally indicating a heightened state of cortical readiness/excitability and positive SCPs signifying a decrease in excitability* [67].

Well known SCPs are the *Contingent Negative Variation* [68, 69] and the *Readiness Potential* (RP) [70, 71], which are negative SCPs that occur in anticipation of an event or an action. CNVs have been found in studies with fixed foreperiods [72, 73, 74, 75, 76] (more on this in Section 4.2) preceding the presentation of a stimulus and pure mental tasks [68, 77]. The RP, while normally associated with motor action preparation [78, 79,

§ Based on our best estimates of accuracies and the presentation time of 1200ms, ITRs are approximately 69 bits/min for Guan's protocol and 91 bits/min for the proposed 3–D cubes version. While these are slightly lower than the reported ones, they are still quite impressive.

80], has recently emerged as beeing also associated with the onset of voluntary mental tasks [81].

In terms of amplitudes SCPs can potentially be much bigger than ERPs and can extend over many scalp locations, as indicated in [11] where SCP localisation was studied in 15 healthy subjects finding that participants fell into three groups: nonregulators, local regulators, and global regulators, with global regulators producing extraordinary large SCPs of 30—50 μV in amplitude over many channels.

1.6. Sequential/Periodic Spelling Proposed in this Paper

A common feature of all the ERP-based spelling approaches reviewed above is randomness in the order of presentation of letters (or sets of letters), which, as previously discussed, has typically been considered critical for triggering larger P300 ERPs. Here, however, we question whether such a randomness is required. This is not heresy. In fact, Farwell and Donchin themselves mentioned the possibility of abandoning the oddball paradigm to further improve performance in the conclusions of [9]:

> *"It may also be possible to enhance the speed of the system by incorporating additional components of the ERP. If, for example, we were to present the rows and columns in a regular sequence, one would expect to see a CNV develop as the time for the appearance of the correct column, or row, neared. The relative effectiveness of a random presentation utilizing the P300 solely and a presentation that capitalized on both a CNV and the P300 is a matter for further research."*

However, to the best of our knowledge, *research on spellers using a regular* (hence periodic) *sequence of flashes never took place.* This paper starts filling this knowledge gap to find out whether, indeed, more powerful spellers could be obtained.

We explored this idea in the context of a P300-based BCI mouse in [82, 83]. In that study, a pointer cursor was controlled by participants focusing on eight circular stimuli surrounding the cursor (representing the desire to move the cursor in the directions N, NE, E, SE, S, SW, W and NW), that were highlighted with 100ms SOAs. Different approaches and mental tasks were tested, but the best performing one involved the highlighting of the circles in colour and the participants mentally naming colours of the target. As a result of the offline tests, and maximum ITR of 45 bits/min was achieved.

In this paper, we extend the idea to the domain of spellers to produce a BCI speller with periodic stimulation (like Farwell and Donchin suggested), but that, instead of highlighting multiple characters at a time (like the rows and colours of Donchin's speller), highlights them one at a time, like the SD spellers reviewed in Section 1.4. Because this produces identical and long TTIs, it was hoped that the elicited P300s would be bigger and clearer, and so easier to classify, than for other spellers. In addition, it was hoped that this would elicit CNVs, as postulated by Farwell and Donchin, which could further help in the classification.

This was put to the test in two experiments, where we collected data from participants performing both a spelling task using the new approach proposed here, as well as the classic version of the Donchin's matrix speller, which were then compared.

1.7. Structure of the Article

The structure of the paper is as follows. In Section 2 we discuss the methodology used, including participants, experimental protocols, EEG acquisition as well as feature extraction and classification. In Section 3, we provide an analysis of the ERPs for targets and non-targets, for both our sequential speller and an implementation of Donchin's speller, and we report on the classification results, including AUCs, accuracies, ITRs and Utilities, and the influence that training-set size has on them. We provide some discussion in Section 4 and some conclusions, limitations and indications for future work in Section 5.

2. Methodology

2.1. Participants and Ethics

The study was performed in accordance with the Declaration of Helsinki and received ethical approval from the University of Essex on the 18th of January, 2019. All adult participants provided written informed consent to participate in the study. Experiments were conducted within the Essex Brain-Computer Interfaces and Neural Engineering laboratory.

Two experiments were conducted, Experiment 3s and Experiment 2s, with eight (age 29.9±3.6, 3 females and 5 males) and ten (age 31.1±5.0, 4 females and 6 males) participants, respectively, with normal or corrected-to-normal vision. Participants were recruited from the University of Essex student and staff population and they were required not to have an own or family history of epileptic seizures and have normal colour vision.

2.2. Experimental Protocols

In both Experiment 3s and Experiment 2s, participants were required to use two different BCI protocols in a *counterbalanced* manner:

(i) a version of Farwell and Donchin's classic matrix speller [9] (called "Donchin's speller" hereafter), where rows and columns are highlighted in random order, and

(ii) the new BCI speller we propose here, based on a circular arrangement of characters that are sequentially highlighted either in green or red (randomly) for a short time (called "Circular speller" hereafter).

Both spellers contained 36 characters including the 26 characters of the English alphabet, a space (represented as "_") and the numbers from 1 to 9. In Donchin's speller, these were organised in a 6 by 6 matrix structure. The arrangement of the characters for the two spellers in shown in Figure 1. The mental task associated with targets was *silent flash counting* in the case of Donchin's speller and *silent colour naming* in the case of the Circular speller.

Participants sat on a comfortable chair about 70cm away from a 52.5cm-wide and 30cm-high LCD screen with a resolution of 1920×1080 pixels. In both spellers, characters were rendered using the GNU Free Sans Bold font (see `https://github.com/opensourcedesign/fonts/blob/master/gnu-freefont_freesans/FreeSansBold.ttf`) with a size of 45 pixels (the height of the display divided by 24). The characters subtended a solid angle of approximately 0.0003 steradians. The background was black. In both spellers non-highlighted characters were shown in grey with RGB=(100,100,100). In Donchin's speller characters in highlighted rows and columns were drawn in pure white with RGB=(255,255,255). In the Circular speller highlighted characters were either red with RGB=(255,0,0) or green with RGB=(0,255,0).

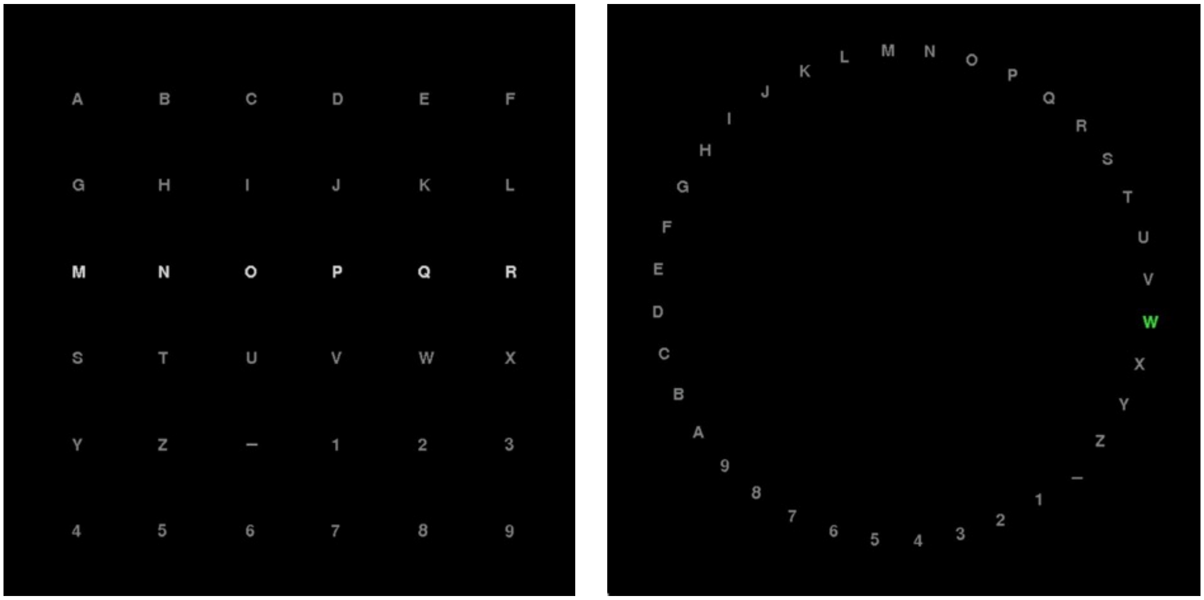
For both spellers, the target character, i.e., the character participants were asked to focus on, was highlighted in blue with RGB=(0,0,255) for two seconds, before the beginning of the flashing.

Both the matrix in Donchin's speller and the letter-circle in the Circular speller were centred in the middle of the display. Their size was determined in such a way to ensure both subtended approximately the same solid angle (0.1 steradians). For the Circular speller this was obtained by setting the radius of the circle to 10 times the font size (i.e., 450 pixels or 12.5cm). For Donchin's speller the matrix edge length was set to 17.5 times the font size (i.e., 787.5 pixels or 21.9 cm). With these dimensions, in the Circular speller characters were approximately 2cm apart, while in Donchin's speller they were regularly spaced at a 4.4cm distance.

After EEG preparation, before starting each protocol, participants were given the protocols' written instructions, followed by a practice session.

In the experimental sessions, trials were divided into blocks (more on this below). Within each block, participants were asked to focus only on the flashing of the target character that was specified at the beginning of the block. At the end of each block, participants could rest until ready for the next block.

The methodology of the two experiments was almost identical, with the only difference being that in Experiment 3s all stimuli (all 12 rows and columns in Donchin's speller and all 36 individual characters in the Circular speller) were highlighted once within a period of $R = 3$ seconds, while in Experiment 2s all stimuli were highlighted within an $R = 2$ seconds (where we use the symbol R as a mnemonic for "*repetition*"). If we denote as S the number of stimuli, ISI = SOA = R/S. So, in Experiment 2s, the ISI is 166.7ms (2,000ms/12) for Donchin's speller but only 55.6ms (2,000ms/36) for the Circular speller, while in Experiment 3s, the ISI is 250ms (3,000ms/12) for Donchin's speller and 83.3ms (3,000ms/36) for the Circular speller. However, in the Circular speller each character remains highlighted for $R/12$ like for Donchin's speller. In other words, similarly to [62] (see Section 1.4), there are 3 neighbouring characters highlighted on the screen at any given time.



(a) (b)

Figure 1. The arrangement of the characters in (a) Donchin's matrix speller and (b) Circular sequential speller.

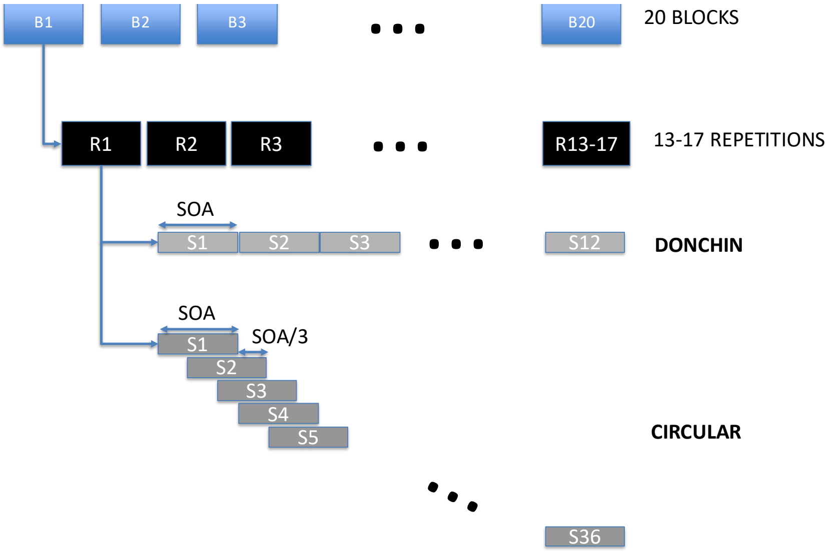


Figure 2. Timeline of the experimental protocols used in this study. Participants spelled 20 characters (one per block), with a variable number of repetitions (between 13 and 17). Each repetition lasted for $12 \times SOA$ irrespective of speller. In the Circular speller stimuli had the same SOA as for the Donchin's speller, but their onset was only 1/3 SOA apart, so that at any given time three neighbouring letters were highlighted. With this choices the character selection time for the two spellers was the same.

As illustrated in Figure 2, in each experiment participants completed 20 blocks of trials with Donchin's speller and 20 blocks with the Circular speller (in counterbalanced order). Once again, in each block participants had to focus on one specific character.

Each block contained a random number (between 13 and 17) of repetitions in which all characters were highlighted once and the participant was asked to perform a mental task when the target character was highlighted. As indicated above, with Donchin's speller, participants were asked to mentally count the number of flashes of the target character in a block; to encourage compliance, at the end of each block participants were asked to report the number of flashes counted. In the case of the Circular speller, participants were asked to mentally name the colour of the target as soon as it flashed (randomly either green or red); to encourage compliance, they were asked to report the colour of the last target flashed in a block.

In Donchin's speller sessions participants were presented with 588 targets (294 rows and 294 columns) and 2,940 non-targets (corresponding to 294 repetitions). In the Circular speller sessions participants were presented with 309 targets and 10,815 non-targets (corresponding to 309 repetitions).

2.3. Data Acquisition

A 64 channel Biosemi ActiveTwo EEG device was used to record brain signals. The data were sampled at 2,048 Hz.

The data were converted from the original Biosemi `bdf` format to MNE's `raw` format [84]. The mean and standard deviation of the standard deviations of all channels were computed, and the channels with standard deviations more than two standard deviations higher than the mean were identified as "bad". Bad channels were removed and reconstructed via the `raw.interpolate_bads()` function of MNE using neighbouring (non-bad) channels.

In preparation for ICA, the data were band-pass filtered between 1Hz and 100Hz and downsampled to 288Hz. This particular sampling frequency was chosen so as to ensure that the accuracy of the timing of trigger events for both spellers and presentation speeds would not suffer from rounding errors (288 it is a multiple of the stimulation frequencies used, namely 4, 6, 12 and 18Hz). ICA was then applied (Infomax method, 50 components) and the IClabel method [85] was used to label the components. Only those labelled as "brain" or "other" were retained. The retained components were then applied to the raw (non-filtered) data.

The resulting data were referenced using the infinite-point REST method [86], band pass filtered between 0.2Hz and 20Hz, and further downsampled to 72Hz (which also respects the timing of events) for further processing.

2.4. ERP and SCP Analysis and Statistical Tests

In order to visualise the ERPs and SCPs recorded with different spellers and presentation rates, we used both grand-averages and individual averages. Individual averages were computed by separately averaging the target and non-target epochs of a participant. We did this both with baseline correction (which makes it simpler to identify ERPs) and without it (which makes it simpler to identify SCPs). Grand averages were computed by averaging the target and non-target averages across participants.

In all cases, both for participant-by-participant results and aggregate (grand-average) results we provide statistical evidence across all channels and times considered provided. For this we used the *threshold-free spatio-temporal cluster permutation test* [87, 88] available in MNE (1,000 permutations, $\alpha = 0.05$, with start and step values for threshold identification both set to 0.4). This was applied using the the individual target and non-target averages as observations to study the statistical significance of grand-averages. It was also applied to the target and non-target epochs, on a pariticipant-by-participant basis, to study the statistical significance of individual averages.

2.5. Feature Selection and Classification

For ERP analysis and classification, we extracted stimulus-locked epochs, starting 200ms before the onset of the flashing of a row or column in Donchin's speller, and of an individual character in the Circular speller, and lasting until 1000ms after stimulus onset. Each epoch thus contained 87 samples for each channel. Baseline correction was applied using the -100ms to 0 period.

Because ERP based spellers are typically based on P300s, we limited our attention the following 19 central, parietal and occipital channels: Cz, CPz, Pz, POz, Oz, P1–P8, PO3, PO4, PO7, PO8, O1 and O2. This brought the number of features from the original 5,568 down to 1,653.

Our classification pipeline starts with the xDAWN algorithm [89] which is a spatial filter that optimises the distinction between the ERPs associated to target and non-target stimuli. We extracted three components, which reduced the number of features to 261. This was followed by a standard scaler, which Z-scores the features.

We should note that there is a very strong imbalance between the target/non-target classes for both Donchin's and the Circular speller. The imbalance is 1 (target-containing row/column) to 5 (non-target rows/columns) for Donchin's speller, but it is much higher for the Circular speller: 1 (target character flash) to 35 (non-target character flashes). Without correction, such strong imbalance would produce poor classification performance. We addressed this imbalance by adding the `RandomOverSampler` class (with `shrinkage=0.5`) from the `imblearn` package [90] to our pipeline, which over-samples the minority class. However, instead of creating exact duplicates, it computes the covariance matrix of the minority class and shrinks it towards the actual samples, generating new samples by adding multivariate normal deviates (with the shrunk covariance matrix) to existing samples. This technique was used to reduce the imbalance for both spellers to 1/4.

Our classification pipeline was completed by an ensemble classifier: `sklearn`'s `VotingClassifier` with soft voting, where final predictions are based on the average of predicted probabilities from all base classifiers, and the class with the highest probability is chosen. The classifier combined a traditional Logistic Regression classifier (with *l1* penalty) and the *CatBoost* classifier [91], which reduces overfitting and improves generalisation compared to other gradient-boosting classifiers.

2.6. Performance Evaluation

As it is customary in machine learning, for each participant and experimental condition, the pipeline described above was used with k-fold cross validation, which splits the epochs into k non-overlapping sections (folds), and trains a classifier k times, each time using a different fold for testing and the remaining $k-1$ folds for training. By combining the classifier outputs obtained in each test folds, we obtained cross-validated predictions across each dataset.

For each fold, performance was evaluated using the usual metric of target-vs-non-target classification performance as well as task (spelling) performance.

Classification performance metrics included receiver operating characteristic (ROC) and the corresponding Area Under the Curve (AUC) as these are particularly important given the strong imbalance in the classes. Task performance evaluation metrics included: task accuracy (i.e., fraction of correctly spelt characters), the Information Transfer Rate (ITRs) (bits per minute), and utility (the actual number of bits per minute transferred, considering that deletions would be needed for any incorrectly spelt character).

The ITR was computed using the standard approximate formula [34, 92, 35]:

$$\mathrm{ITR} = \left(\log_2 N + P \cdot \log_2 P + (1-P) \cdot \log_2\left(\frac{1-P}{N-1}\right)\right) \Big/ \left(\frac{T}{60}\right), \tag{1}$$

where N is the number of classes, P is the accuracy of the classifier and T is the selection time in seconds, while the BCI *utility* measure was computed according to the definition in [93] (with symbols having the same meaning as above):

$$U = (2P-1)\log_2(N-1) \Big/ \left(\frac{T}{60}\right) \tag{2}$$

if $P > 0.5$, and $U = 0$ otherwise.

Because, typically, single-trial task accuracy of BCI spellers is insufficient for communication, as customary, the task performance metrics mentioned above were evaluated also when averaging the classifier scores/proba's over *multiple repetitions* of stimulus presentation, before a character is selected. Specifically, we tested this from 1 to 10 repetitions of the stimulation sequence.

In addition, we studied how performance varies as a function of the training set size, as for spellers to be practical, they must not require participants to spend a long time providing training data for the speller for its online performance to be acceptable. We did this by studying how performance varies with the cross-validation parameter k (varying from 2 to 20). Since this allows only to test from 50% training vs 50% testing to 95% training vs 5% testing, we also used a modified for of cross-validation where: (1) *the roles of training and test folds are swapped*, and (2) *we sample the testing folds* so as to avoid having multiple predictions for the test data. This allowed us to test performance also for much smaller training sets, e.g., as low as 5% training vs 95% testing.‖

3. Results

In this section, we will analyse and compare the ERPs and SCPs observed in grand-averages (and to some extent also individual averages) for targets and non-targets trials obtained for the different protocols, in both Experiments 3s and 2s (Section 3.1). Then, we examine the results of the target *vs* non-target classifications as well as the overall task performance (including ITR and utility) for the spellers under test (Section 3.2).

3.1. ERP and SCP Analysis

In this section we compare the ERPs and SCPs elicited by Donchin's speller and the Circular speller at the two presentation rates. We report results obtained both in the *absence and presence of baseline correction*, as such a correction, while emphasising ERPs, may affect our ability to detect SCPs that are present in some conditions, both before stimulus presentation and after it.

3.1.1. Donchin's Speller

Figure 3 shows the EEG *grand-averages* for targets and non-targets obtained with Donchin's speller.

As one can immediately see, *grand-averages are largely unaffected by the baseline correction*. Both with and without it, we see clearly recognisable ERPs in the presence of targets, including P2, N2, and P300. On the contrary, we only see the exogenous-ERP ripple at the stimulation frequencies for the non-targets. This has a frequency of 4Hz and peak-to-peak amplitude of approximately $1\mu V$ in the slower condition and a frequency of 6Hz and a peak-to-peak amplitude of approximately 0.5μV in the faster condition.

These exogenous SSVEP-like oscillations also affect targets epochs. So, a better way to see the differences between target and non-target ERPs (for both stimulation frequencies) is to look at their difference, which is reported in Figure 4.

From this figure, we can see the morphological differences between the grand-average ERPs associated with targets in the two experiments. In particular, taking as a reference the baseline corrected case (Figure 4 bottom), the P2 has larger amplitude, the N2 — while of similar amplitude — is narrower and the P300 is smaller at the faster presentation rate (Experiment 2s). A smaller P300 could be expected in Experiment 2s

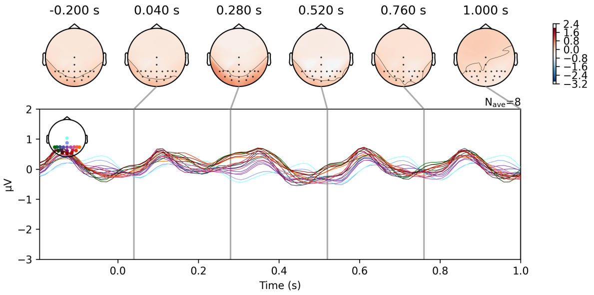
---

‖ For instance, if we swap the role of training and test sets in 3-fold cross validation, we will train three times with one fold (i.e., 33.3% of the data), and we will test three times on two folds (66.7% of the data). As a result, at the end of the process, we will have two cross-validate predictions for each trial, and not one as usual. With our adaptation of the method, we simply sample from the available predictions so that we end up with one cross-validated prediction per trial.

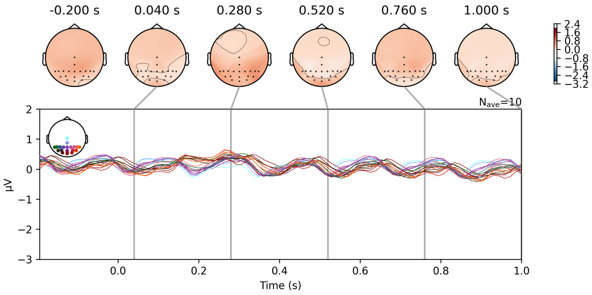
Donchin 3 second experiment | **Donchin 2 second experiment**

Without baseline correction | *Without baseline correction*

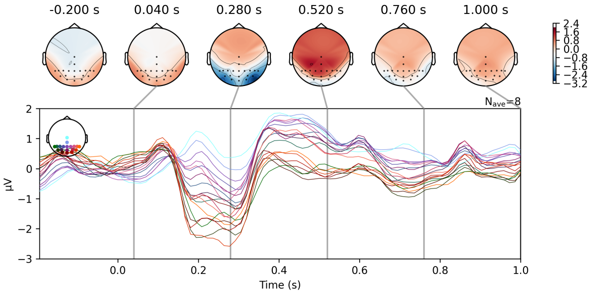
Non-targets



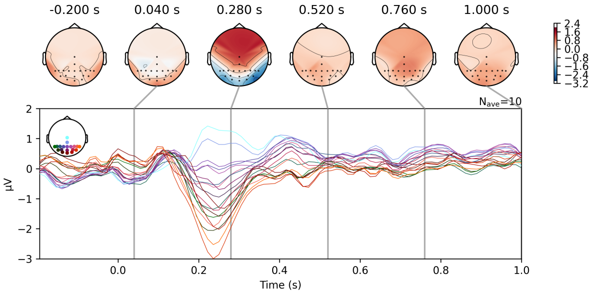
Non-targets



Targets

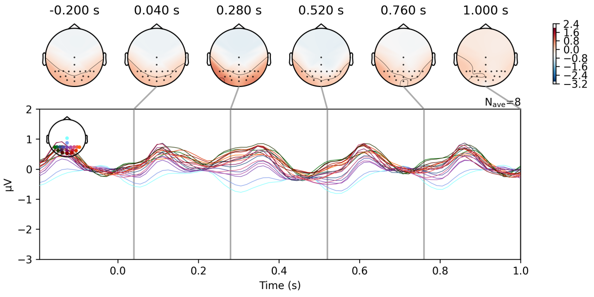


Targets

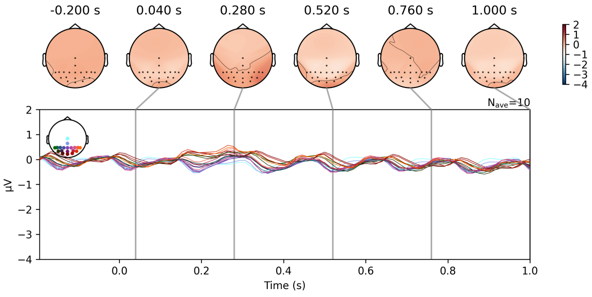


With baseline correction | *With baseline correction*

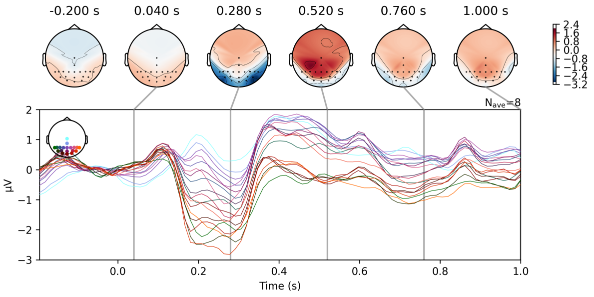
Non-targets



Non-targets



Targets



Targets

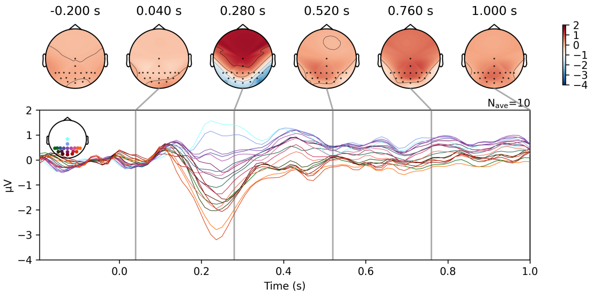


Figure 3. EEG grand averages of targets and non-targets for Donchin's speller in Experiments 3s (left) and 2s (right) without (top) and with (bottom) baseline correction. Different colours correspond to different channels with the mapping specified in the small scalp map in the top left corner of each plot.

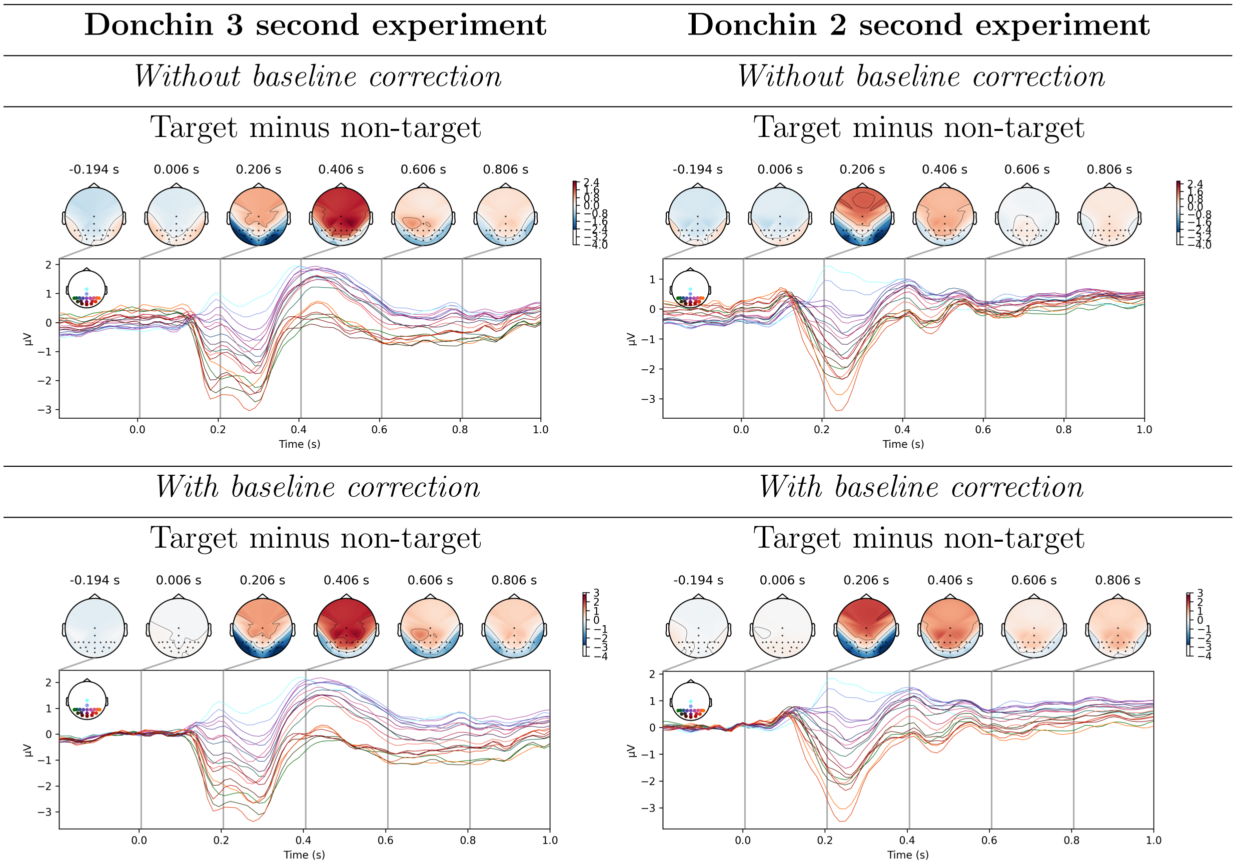


Figure 4. Difference between the grand average for targets and non-targets for Donchin's speller in Experiments 3s (left) and 2s (right) without (top) and with (bottom) baseline correction. Different colours correspond to different channels with the mapping specified in the small scalp map in the top left corner of each plot.

due to its higher target frequency, as discussed in Section 1.1. In any case, the P300s are in both cases fairly small, possibly due to the presence of perceptual errors and near targets [39, 40].

The spatio-temporal clusters where the permutation test (see Section 2.4) has identified *statistically significant differences in the grand-averages for targets and non-targets* for Donchin's speller are shown in Figure 5. There, all areas in blue (red) represent channels and time steps where the differences between the individual averages of targets and non-targets are statistically significantly negative (positive), with the darker the colour, the larger the amplitude difference.

Focusing on the plots on the left, which relate to the slower Donchin's speller, both *N2 and P300 are statistically significant in most channels* but there is also statistical significance for the *P2 in channels Cz and CPz*, irrespective of baseline correction.

The clusters of statistical significance for the faster Donchin's speller (Figure 5 right), however, show a very different picture. While the grand-averages for the 2s condition look very similar to those for the 3s condition and appear to be unaffected by the baseline correction, the reality depicted in Figure 5 is quite different: in the *absence* *of baseline correction* the N2 and P2 are statistically significant (although for very few channels and time steps), but *the P300 does not reach statistical significance.*

On the contrary, *with baseline correction*, not only N2 and P300 are statistically significant, but so is also the P1 and the mostly-positive *low-voltage SCPs following the P300* in the faster protocol. The latter appear to be present also at the slower presentation rate, but not consistently enough across subjects to be statistically significant in the grand average. Only in channel P8 we see a statistically significant SCP in the slower speller.

However, in Donchin's speller nothing noteworthy seems to take place before stimulus presentation nor for the first 90ms following its onset for either presentation rate, irrespective of baseline correction.

3.1.2. Circular Speller

As shown in Figures 6 and 7, the baseline correction process has significant effects on the grand-averages observed for the Circular spellers. Both with and without it, we see some ERPs — specifically, a large occipito-parietal N2 and a large centro-parietal P2 that effectively blends into an early P300 — in the presence of targets. On the contrary, we only see a tiny exogenous ripple at the stimulation frequencies (12Hz and 18Hz) for the non-targets.

In the *absence of baseline correction*, the *P2-P300 complex is statistically significant* in 7 electrodes (Cz, CPz, Pz, P1, P2, P4 and P6) at the slower presentation rate, but only Cz in the faster one (see Figure 8 top). Instead, the observed *N2 is not statistically significant* for either presentation rate.

Before such ERPs, in Figures 6 and 7, we see *very clear pre-stimulus SCPs*, their sign being negative in Cz and CPz (suggesting they might represent a CNV) and positive for several other electrodes. In the faster protocol, these SCPs are statistically significant for more than half of the channels (see Figure 8 top right). In the slower protocol, only the negative SCPs in channels Cz and CPz are statistical significant (see Figure 8 top left).

After the ERPs, still in the non-baselined data, in Figures 6 and 7 we see quite *large late SCPs*, which are mostly *negative* unlike for Donchin's speller, where such potentials were mostly positive and less prominent. Despite being larger at the slower presentation rate, such SCPs are not significant. However, their tail ends (700—900ms after target onset) are *statistically significant for the faster presentation rate* across just under half of the channels.

Turning now to the *baselined grand-averages* shown in Figures 6 (bottom half) and 7(bottom), we see that the SCPs preceding the target presentation are, of course, completely zeroed by the baseline correction process. However, *the N2 and the P2-P300 ERPs are much emphasised* and have amplitudes between 50% and 100% larger w.r.t. the no-baseline case. The P2-P300 complex is now much wider too, starting at approximately 100ms and ending at 500ms after stimulus onset. We also see that the aforementioned *post-ERP SCPs are much more prominent*, becoming approximately

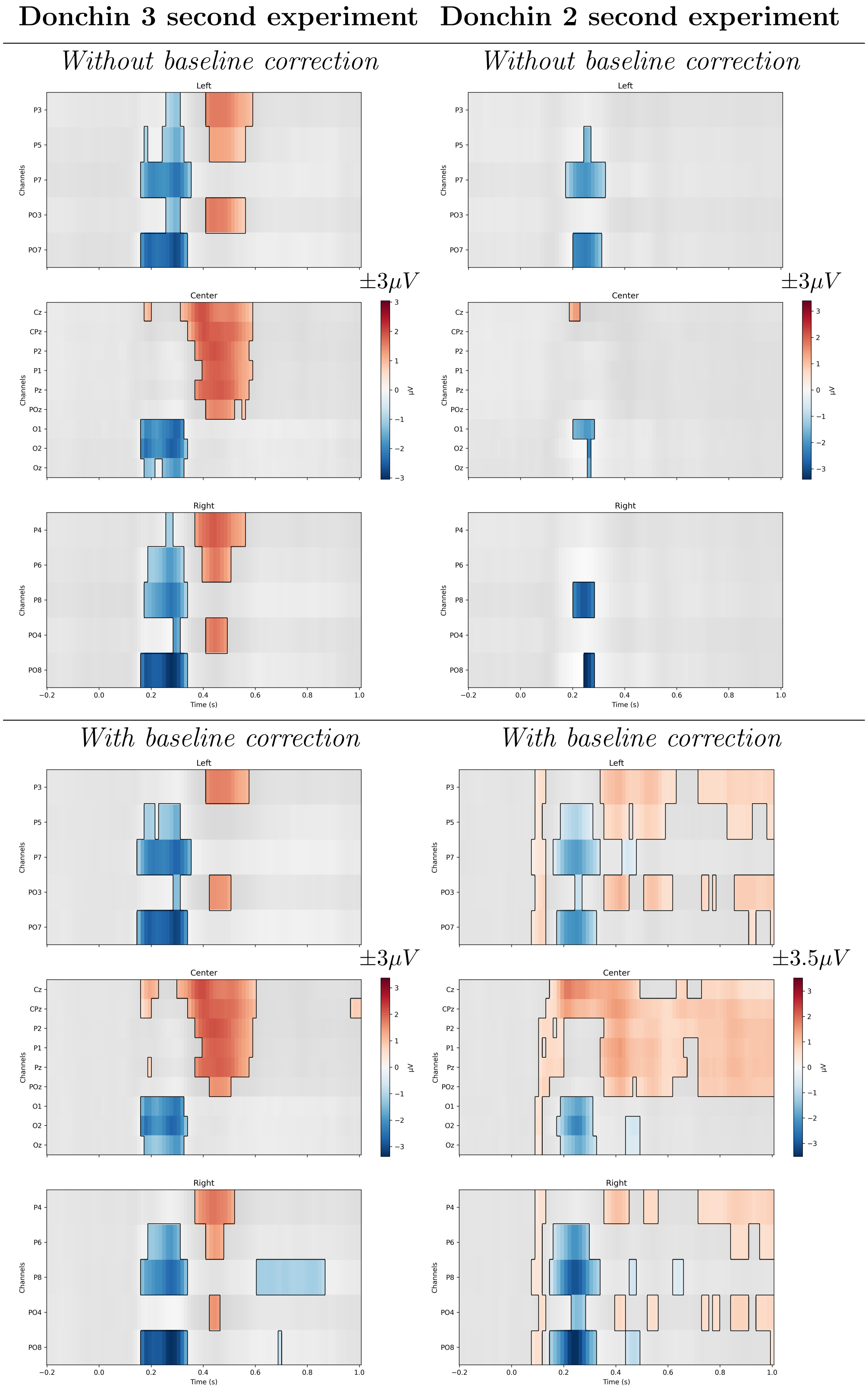


Figure 5. Spatio-temporal cluster permutation test results for Donchin's speller in Experiments 3s (left) and 2s (right) without (top) and with (bottom) baseline correction. The background represents the difference of the grand averages of targets and non-targets (as in Figure 4). The coloured regions represent the statistically significant ($\alpha = 0.05$) clusters of channels and times.

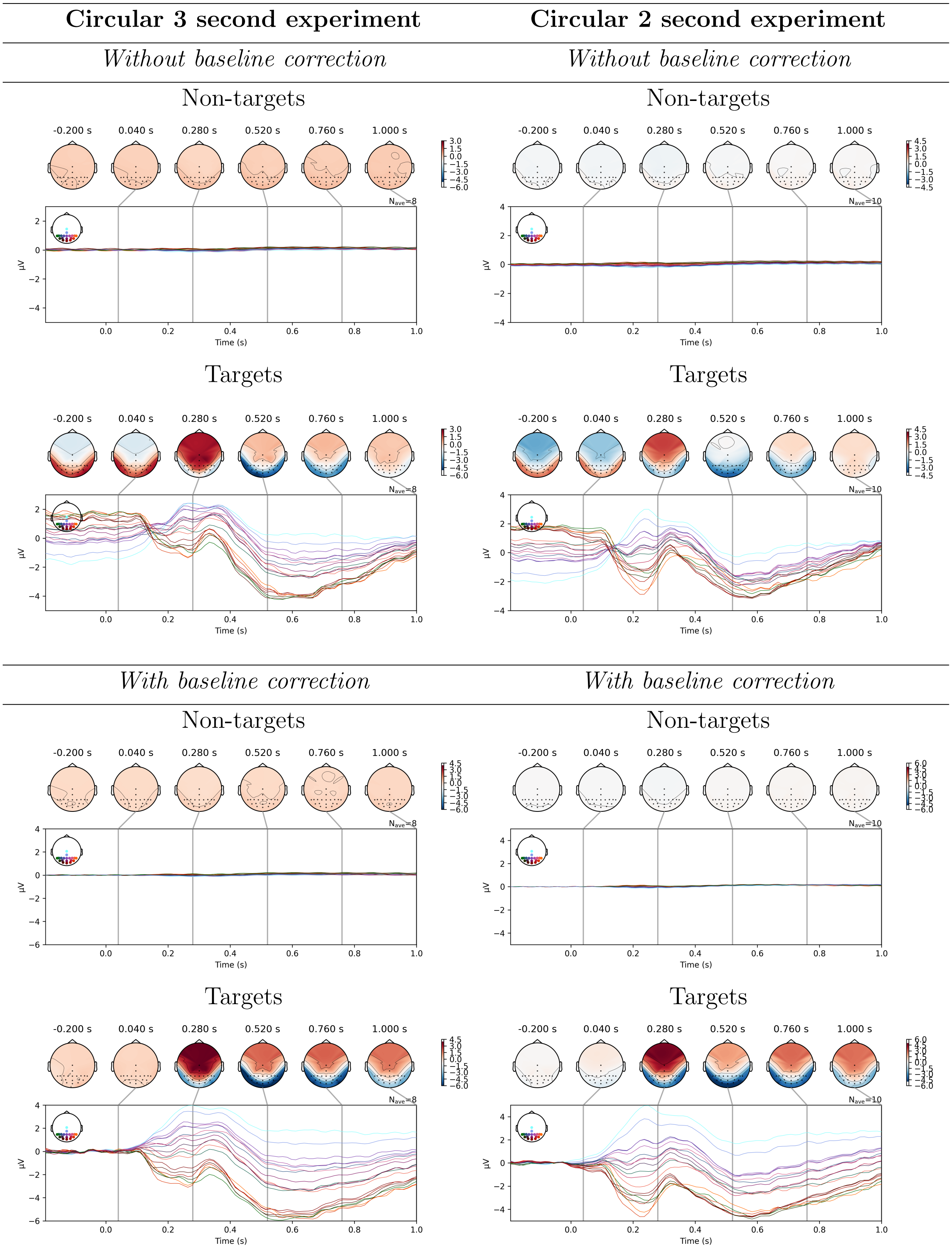


Figure 6. EEG grand averages of targets and non-targets for the Circular speller in Experiments 3s (left) and 2s (right) without (top) and with (bottom) baseline correction. Different colours correspond to different channels with the mapping specified in the small scalp map in the top left corner of each plot.

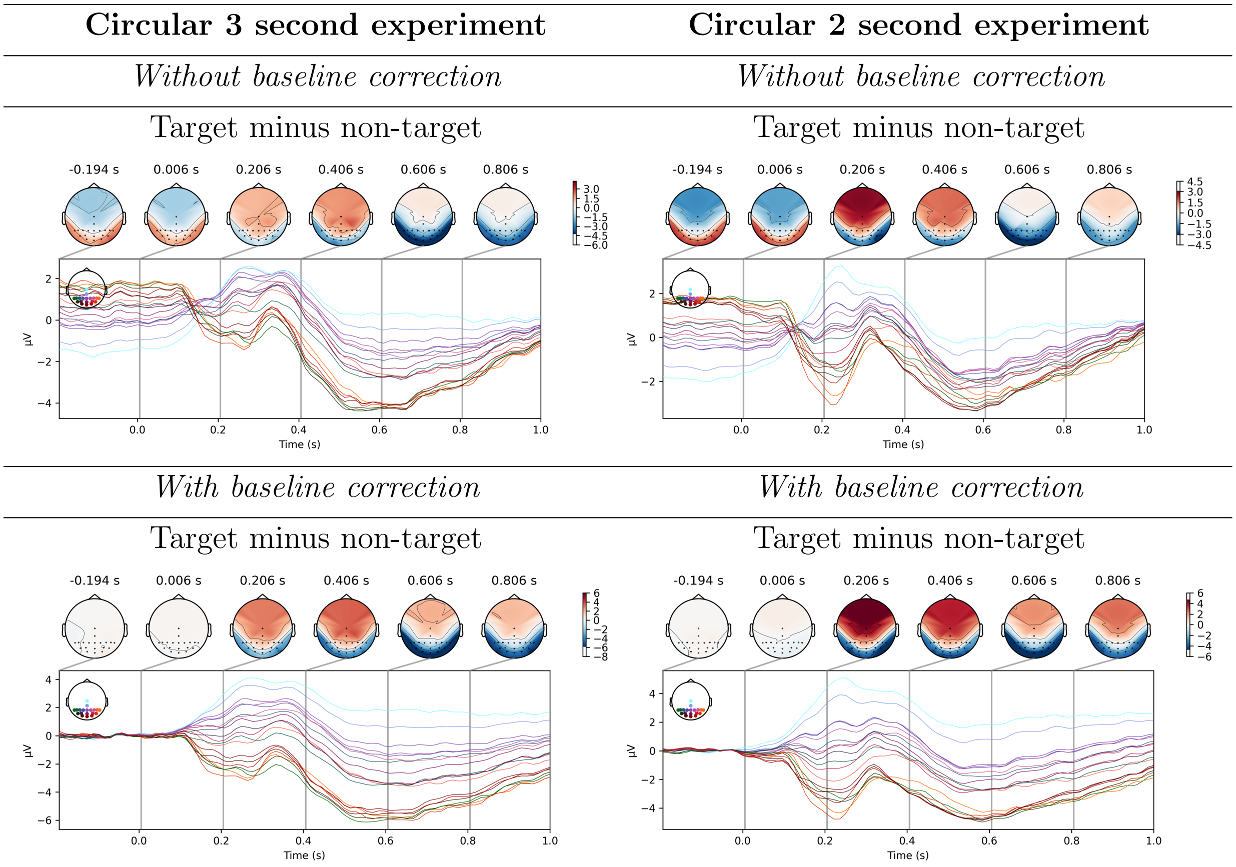


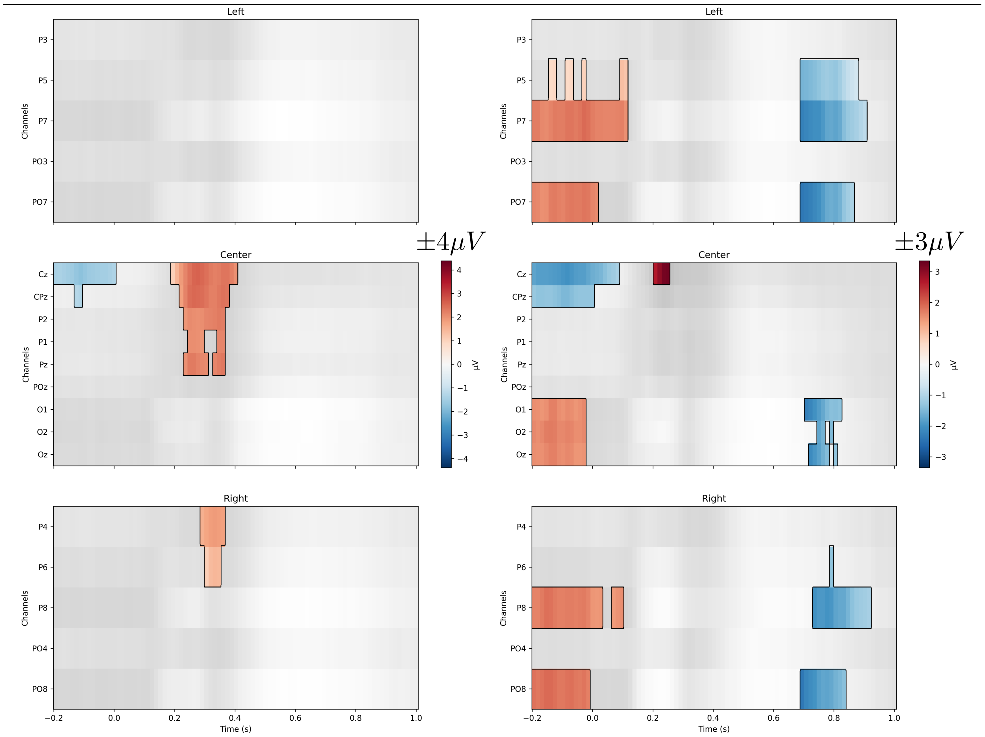
Figure 7. Difference between the grand average for targets and non-targets for the Circular speller in Experiments 3s (left) and 2s (right) without (top) and with (bottom) baseline correction. Different colours correspond to different channels with the mapping specified in the small scalp map in the top left corner of each plot.

50% bigger in voltage than without baseline correction. As a result, with baseline correction the *N2 and P2-P300 complex are statistically significant at both presentation rates*, collectively in almost all channels as shown in Figure 8(bottom). Also, the *post-ERP SCPs are statically significant at the faster presentation rate for more than half of the channels* until the very end of the epoch.

3.1.3. Event Related Slow Cortical Potentials

We should note that the *anticipatory and posticipatory SCPs* elicited by Donchin's speller and the Circular speller at the faster presentation rate, described in the previous two sections, *are not trained responses*. Instead, they are naturally triggered by preparations taking place both before the predictable onset of the target stimulus and relaxation/reduction in excitability after the stimulus presentation has taken place. In other words, they too are *event-related*. For this reason, we will call them *Event-Related Slow Cortical Potentials* (ERSCPs) here after.

Circular 3 second experiment	**Circular 2 second experiment**
Without baseline correction	*Without baseline correction*



With baseline correction	*With baseline correction*

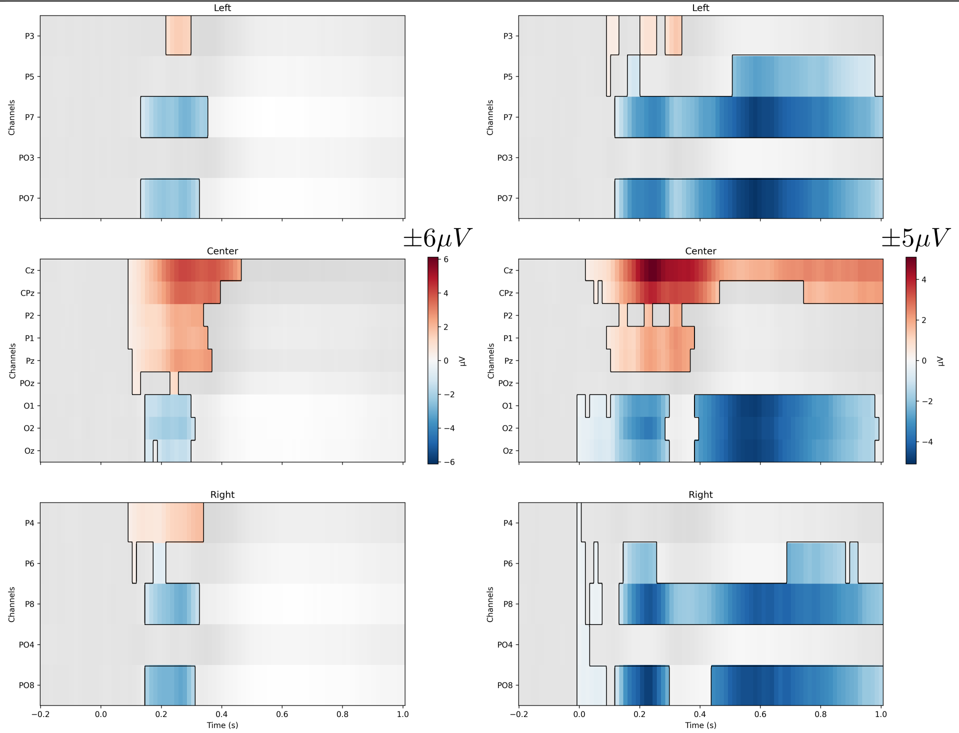


Figure 8. Spatio-temporal cluster permutation test results for the Circular speller in Experiments 3s (left) and 2s (right) without (top) and with (bottom) baseline correction. The background represents the difference of the grand averages of targets and non-targets (as in Figure 7). The coloured regions represent the statistically significant ($\alpha = 0.05$) clusters of channels and times.

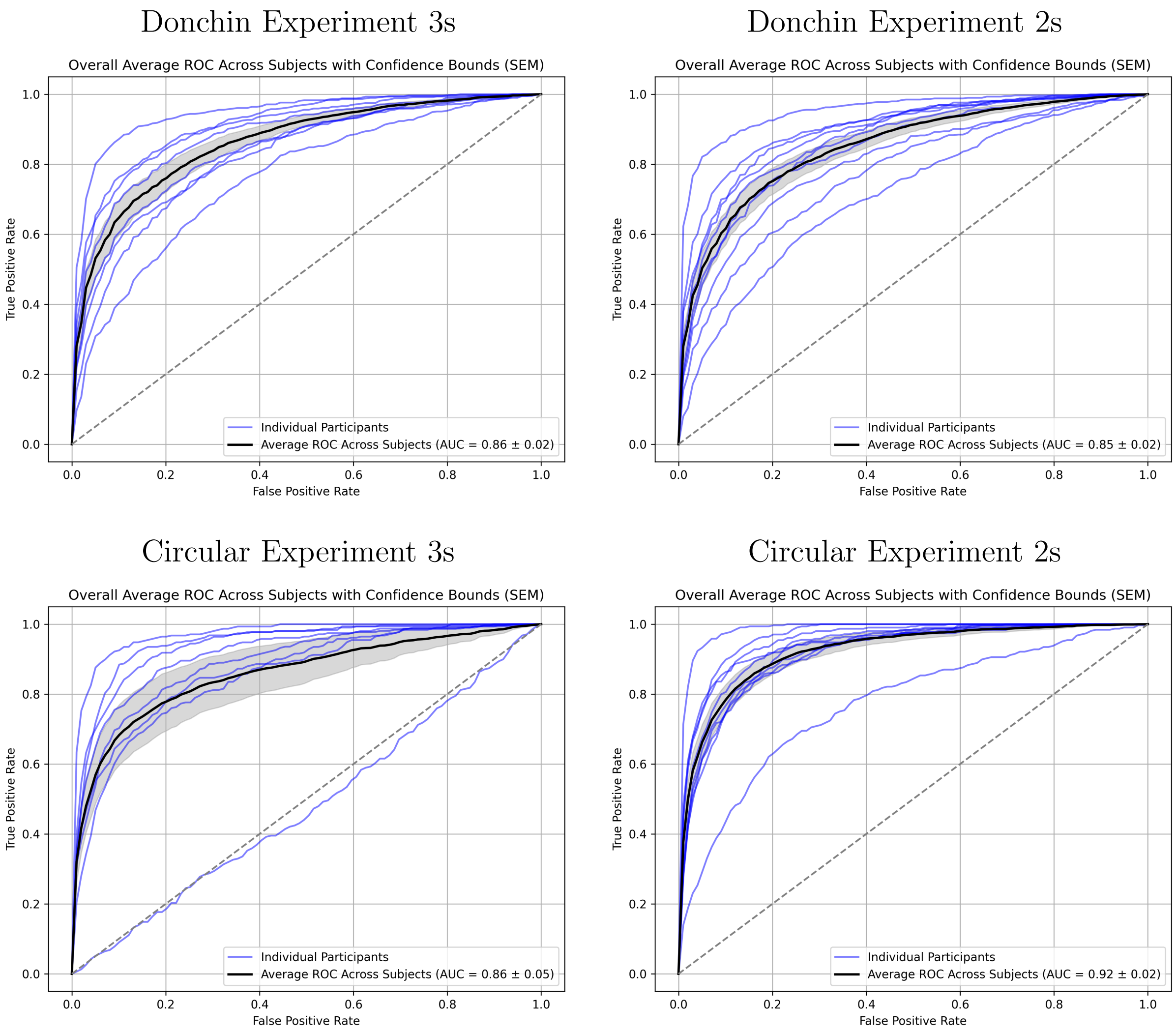


Figure 9. Mean receiver operating characteristic (ROC) curves and AUCs (with their SEM bounds) for the circular sequential speller and Donchin's speller for Experiments 3s and 2s. The blue lines represent the ROC curves for each participant.

3.2. Classification Results

In this section we will first look at the ability of the classifier to distinguish targets from non-targets for the two spellers and the two stimulus presentation rates (Section 3.2.1). We will then focus on task/spelling performance (Section 3.2.2). Finally, we will study how the performance of the classifiers is affected by the size of training sets (Section 3.2.3).

3.2.1. Receiver Operating Characteristics and AUC for Epoch Classification

Figure 9 shows the cross-validated ($k = 20$) Receiver Operating Characteristic (ROC) curve for the classifiers, with the corresponding Area Under the Curve (AUC), where the dashed lines represent the performance of a random classifier. The classifier was trained and tested on baseline-corrected epochs.

Looking first at Experiment 3s (left column), we see that the *average ROCs* (thick black line) *for the Donchin and Circular spellers are almost identical*, and the corresponding *AUCs are identical* (0.86). This seems to be aligned with the similarity in the respective grand-average statistical results reported in Figures 5(bottom left) and 8(bottom left). However, individual spatio-temporal cluster permutation tests (not reported) for the 3s Circular speller indicate that for 6 out of the 8 participants there are ERSCPs that are statistically significant for many channels and time steps, but their polarity is mainly positive for 3 and negative for the other 3. As a result they do no show up nor can be statistically significant at grand-average level. These participant-level differences can be exploited by the classifier though. Indeed, the ROC curve and the AUC for the Circular 3s speller have been depressed by the poor below-the-diagonal ROC (with AUC=0.48) of one participant (Participant 8). Had we excluded this participant, the mean AUC would have been 0.91 (with an SEM=0.02).

If we now turn to the ROCs and AUCs of Experiment 2s (Figure 9 right column), we see that Donchin's AUC (0.85) is almost identical to those of the two 3s spellers discussed above (and with a low SEM of 0.02). This might seem counterintuitive as the P300s observed in Donchin's speller at this presentation rate are the smallest among all conditions.¶ However, as we discussed earlier, in this speller there are small but consistent post-P300 ERSCPs which are statistically significant at grand-average level (see Figure 5 bottom right). We believe that these ERSCPs make up for the reduction in P300 amplitude.

On the contrary, *in the 2s condition, for the Circular speller we see a markedly better ROC curve and AUC* (AUC=0.92). This result makes sense given that this condition produces large and statistically significant N2s and P300s as well as large and distinctive post-P300 ERSCPs which are statistically significant for 15 out of 19 channels (Figure 8 bottom right).

3.2.2. Character Classification Accuracy, Information Transfer Rate and Utility

The selection of characters in Donchin's speller was based on identifying the row (column) with the highest classifier score (technically, `proba`) out of the six rows (columns) that flashed within a 2s for Experiment 2s, and within a 3s for Experiment 3s. So, the problem is a difficult 12 class problem, with a chance level of 8.3%. In the Circular speller, the character with the highest score was selected out of the 36 characters flashed within the same period of time. In the case of the Circular speller, character classification is an even more difficult 36-class problem, where random chance level is 2.8%.

In Table 1, we report the overall classification accuracy obtained after one presentation of the 12 or 36 stimuli associated with each speller protocol for both experiments. Here the two Donchin's spellers and the Circular 3s speller are broadly

---

¶ Possibly linked to the fact that reducing the TTI in the oddball task, and more specifically with a Donchin's speller, reduces the amplitude of the P300s elicited by flashes of target characters [94, 26, 9, 36].

Table 1. Single-repetition character classification accuracy for different protocols and experiments. Bold face indicates the superior speller for each experiment. Underlined is the top performer across all conditions.

Experiment	Protocol	Accuracy
3s	Circular	**40%**
	Donchin's	39%
2s	Circular	**<u>48%</u>**
	Donchin's	38%

equivalent with accuracies between 38% and 40% (albeit, Participant 8 depressed significantly the average for the Circular speller, which would have been 45% had we excluded them). On the contrary, *at the faster stimulation rate, the Circular speller, with 48% accuracy, is clearly dominant over all other spellers and stimulation rates.* We should note that single-trial 48% accuracy is extremely high for a 36-class problem, and compares very well with the 55% accuracy we obtained for pointer control [82], which was an 8-class problem.

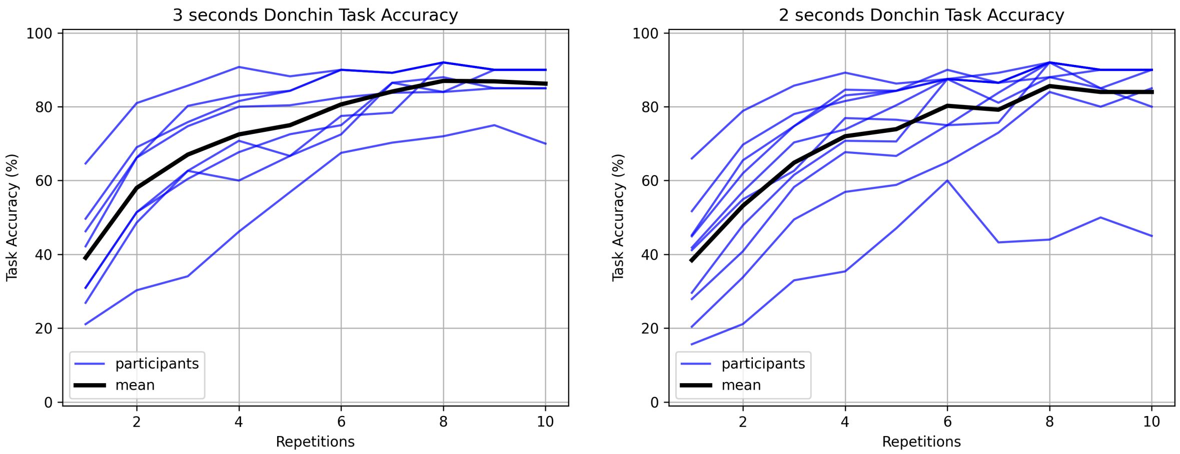
We followed the standard approach of averaging the score of the classifier over multiple repetitions of target presentation, and then selecting the row/column with maximum average score in Donchin's speller, and the character with maximum average score in the Circular speller. We then computed the resulting accuracies (%), ITRs (bits/minute) and Utilities (bits/minute), as a function of the number of trials averaged, for each speller and experiment.

Results for all spellers, presentation speeds and metrics are reported in Figures 10, 11 and 12.

Looking first at the mean *task accuracy* (Figure 10), for all spellers and presentation rates, we see *healthy increases in task accuracies as the number of repetitions increases.* The only key difference between Donchin's and the Circular speller is that with the Donchin's none of the participants achieved accuracies above 90% even with 10 repetitions, while for the Circular spellers 5 out of the 8 participants reached 95% or higher in the 3s protocol and 7 out of 10 reached that level in the 2s protocol.

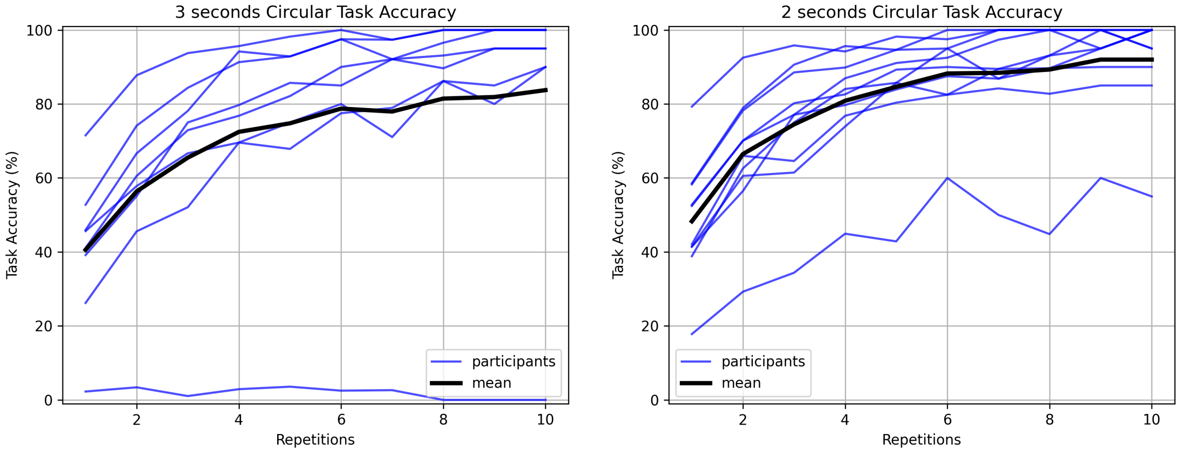
In terms of peak performance, the Circular protocol with 2s revolutions, as expected given its superior single trial accuracy, provides better and more consistent performance across participants with an average task accuracy reaching 92% at 9 and 10 repetitions. While the mean task accuracy for the 3s Circular speller in the worst, this has been significantly affected by Participant 8's task accuracy being between 0 and 2% irrespective of repetitions. Had this participant been excluded, this speller would have been a comfortable second.

Turning to the *ITR* (see Figure 11), we see for all spellers the mean ITR starts at above 20 bits/min and then decays as frequently observed in other studies. Because the ITR is inversely proportional to the selection time T (see Equation (1)), for spellers with similar task accuracies, such as the 3s and 2s Donchin's spellers, we should expect



Donchin Experiment 3s

Donchin Experiment 2s



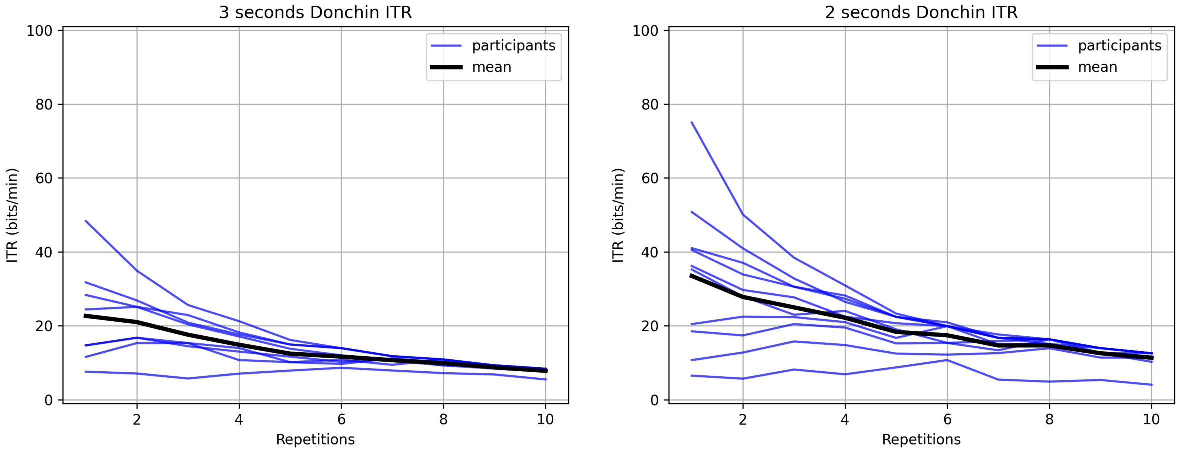
Circular Experiment 3s

Circular Experiment 2s

Figure 10. Task accuracies for Donchin's speller (top) and the Circular speller (bottom) for Experiments 3s (left) and 2s (right) as a function of the number of repetitions. The blue lines represent the accuracy curves for each participant.

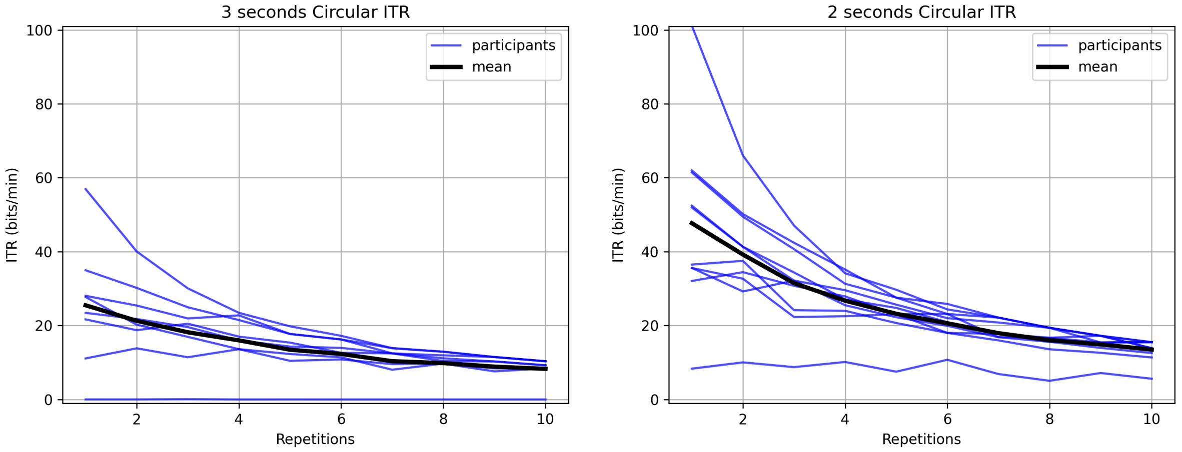
the ITR for a 2s protocol to be 50% higher than the ITR for the corresponding 3s version. This is almost exactly what we see for Donchin's speller where the ITR for the 2s protocol starts at approximately 34 bits/min which is 55% higher than for the 3s protocol. However, in the case of the Circular speller, the ITR for the 2s protocol is *48 bits/min* which is over 90% higher than the corresponding 3s protocol. Once again this is due to the influence of Participant 8. Had we removed the participant, the top mean ITR for the Circular 3s speller would have been approximately to 29 bits/min, suggesting that its task classification performance is only slight inferior to the 2s speller.

An important element to consider in relation to the ITR is that, while it *tends* to monotonically decrease with the number of repetitions, this does not apply to every single participant. Therefore, in addition to looking at the *maximum of the average ITR* (as we have just done), another useful measure of performance to consider might



Donchin Experiment 3s

Donchin Experiment 2s



Circular Experiment 3s

Circular Experiment 2s

Figure 11. ITR for Donchin's speller (top) and the Circular speller (bottom) for Experiments 3s (left) and 2s (right) as a function of the number of repetitions. The blue lines represent the ITR curves for each participant.

be *the average of the maximum ITR* across participants, where for each participant we measure the ITR at their individual best repetition number. The values of the mean of the maximum ITR are reported in Table 2. These are only marginally higher than the peak values in Figure 11 and, so, they confirm the dominance of the 2s Circular speller.

Finally, let us now turn to the all-important *utility* measure, which is the acid test for any BCI speller, in that it directly relates to the number of characters that users can reliably enter (after accounting for mistakes and corresponding corrections). As shown in Figure 12(left), the two 3s protocols are almost equivalent, with mean utility peaking at between 13 and 15 bits/min at 3 repetitions. Naturally, the two 2s protocols are better, with Donchin peaking at 18 bits/min at 4 repetitions and *Circular peaking at 28 bits/min at 2 repetitions.*

Because the utility measure *does not* tend to monotonically decrease with the

Table 2. Mean of the maximum ITR recorded for each participant, in both experiments and with both speller protocols. Bold face indicates the superior speller for each experiment. Underlined is the top performer across all conditions.

Experiment	Protocol	ITR (bits/min)
3s	Circular	**26**
	Donchin's	24
2s	Circular	<u>**48**</u>
	Donchin's	35

Table 3. Average of the maximum utility recorded for each participant, in both experiments and with both speller protocols. Bold face indicates the superior speller for each experiment. Underlined is the top performer across all conditions.

Experiment	Protocol	Utility (bits/min)
3s	Circular	**18**
	Donchin's	16
2s	Circular	<u>**35**</u>
	Donchin's	23

number of repetitions, rather than looking at the maximum average utility, a more precise measure of performance is *the average of the maximum utility* for each participant. We report the values of the latter in Table 3.

Once again this shows the superiority of the 2s Circular speller. At 35 bits/min, users of the Circular speller would be able to spell approximately 7 characters per minute, which seems a significant result for a newly developed protocol.$^+$

3.2.3. Training Set Size and Training Time Effects

To study how performance of the spellers varies as a function of the training set size at both presentation rates, we varied the cross-validation parameter k from 2 to 20 and also used the *swapped cross-validation* approach described in Section 2.6, with the same values of k. Together they made it possible to obtain cross-validated performance results for training sets from 5% to 95% of the data available. The results are reported in Figure 13 in terms of AUCs, maximum ITR and maximum utility.

In the figure, each row corresponds to one metric: average AUC (top), average maximum ITR (middle), and average maximum utility (bottom). The two columns represent the same data but the one on the left adopts the *percentage of training set used for training* as the abscissa scale while the one on the right uses the number of *minutes required for a participant to acquire a training set or training time.* *

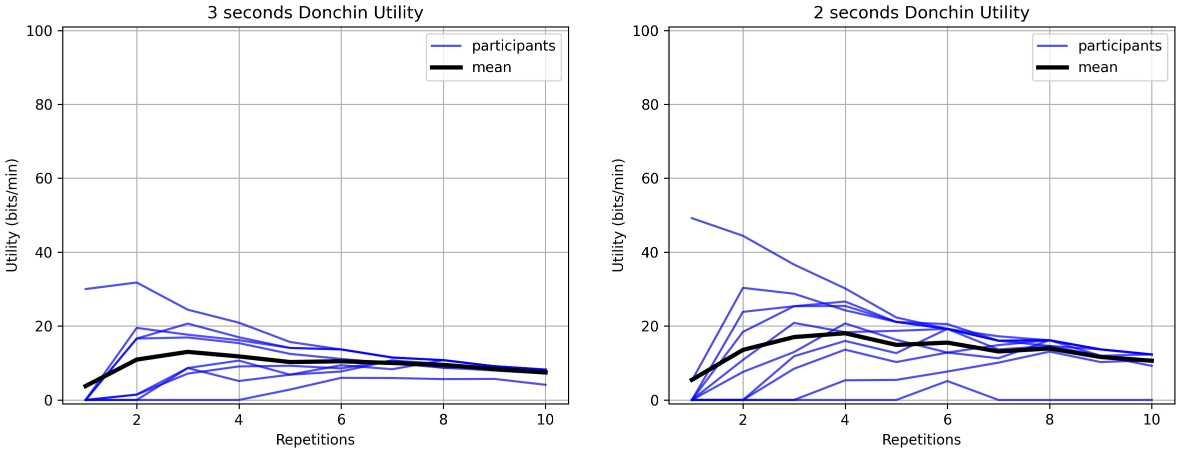
Looking at how the AUCs vary with the amount of training data (Figure 13 top),

---

[+] The number of bits required to encode one of 36 possible characters is $\log_2(36) = 5.17$ bits.

[*] To reiterate: the term "training time" here refers to the time the users need to spend gathering data to train the BCI. This must not be confused with the time required train the BCI's classifier, which is of the order of seconds.

Donchin Experiment 3s   Donchin Experiment 2s



Circular Experiment 3s   Circular Experiment 2s

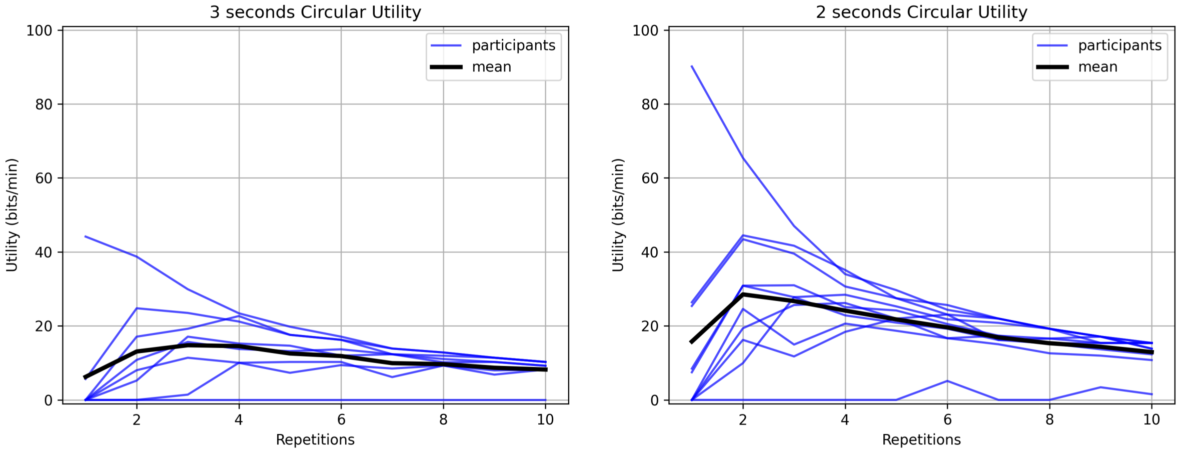
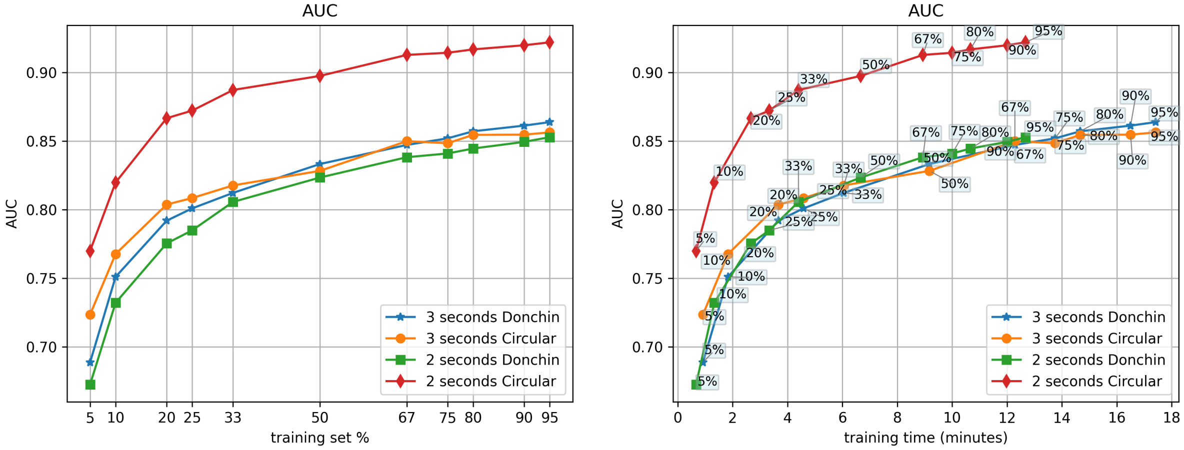


Figure 12. Utility for Donchin's speller (top) and the Circular speller (bottom) for Experiments 3s (left) and 2s (right) as a function of the number of repetitions. The blue lines represent the utility curves for each participant.

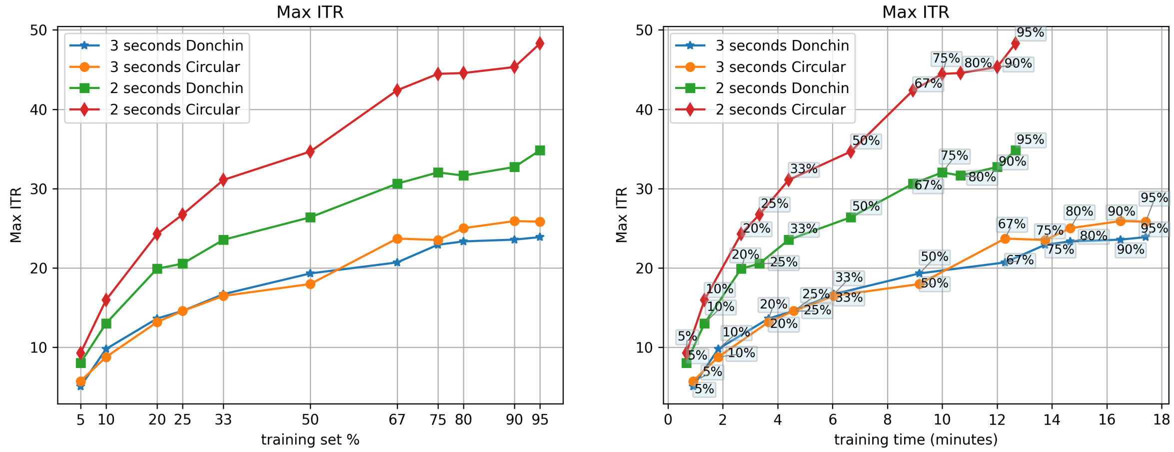
we see that *both 3s spellers and Donchin's 2s speller* generalise in effectively the same manner. They all struggle with very small training sets the AUC plots presenting a knee at approximately 20% of training data, but then show an essentially linear improvement with training set size thereafter. As a result, with 50% of the data for training, their AUC is only approximately 0.03-0.04 away from the maximum value achieved at 95% training. This is a satisfying result as participants would only need to spend between 7 and 9 minutes collecting training data for the BCI (see Figure 13 top right).

Focusing now on the *2s Circular speller* AUCs, we see a similar knee at approximately 20% of training data, and a similar linear growth thereafter. However, all this happens at a much higher AUC, so much so that *this speller would only need 10% of the data, corresponding to approximately 1 and 1/2 minutes of training, to achieve the performance the other spellers achieve with 50% of the data* (i.e., in 7 to 9 minutes

AUC vs training set size (%) | AUC vs training time (minutes)



Max ITR vs training set size (%) | Max ITR vs training time (minutes)



Max utility vs training set size (%) | Max utility vs training time (minutes)

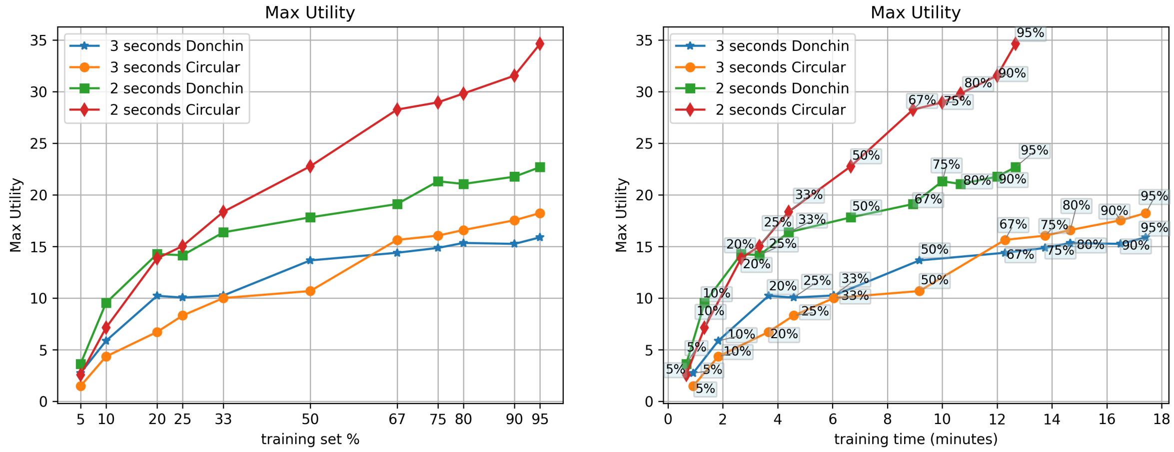


Figure 13. AUC, maximum ITR and maximum utility for the Circular and Donchin's spellers for Experiments 3s and 2s as a function of the proportion of training set (left) used and the training set acquisition time (right). The marker labels in the plots on the right provide the percentage of training set corresponding to each specific training time.

of training).

Looking at the *maximum ITR* (Figure 13 middle), we see a slightly different picture in that the 2s Donchin speller is now superior to the two 3s spellers due to its faster selection time. Now this speller would only require somewhere in between 33% and 50% of the training set (requiring 4 to 7 minutes of training time on the part of the user) to achieve the ITR the 3s spellers achieve with 95% of the data (requiring approximately 18 minutes of training). Here, too, however, the 2s Circular speller excels requiring only between 20 and 25% of the training data (approximately 3 minutes of training time) to achieve the same benchmark. Also, directly comparing the 2s protocols, we see that the Circular one requires only 50% of the data (7 minutes of training) to match the maximum ITR achievable by the 2s Donchin's speller (with 95% of the data or 18 minutes of training time).

Looking at the *maximum utility* (Figure 13 bottom), we see a very similar picture to the ITR, i.e., the 2s Circular speller is superior to all other spellers if trained with the same fraction of training data, or it is requires much less data for the same level of performance. The only exception to this is that the 2s Circular speller has slightly worse utility than the 2s Donchin speller with very small training sets.

Extrapolating the observable trends in Figure 13, it is clear that all the spellers tested have not reached a plateau in their performance, and so might benefit from larger training sets or from meta-learning approaches that make better use of training data [95].

4. Discussion

In this section, firstly (Section 4.1) we discuss Farwell and Donchin's periodic speller hypothesis presented in Section 1.6 in the light of the evidence gathered in our study. Secondly, in Section 4.2 we discuss the relationship between TTI and ERSCPs in spellers with studies on the preparedness and the foreperiod. Finally, we formulate the hypothesis that preparatory ERSCPs are present also in Donchin's speller, but only in a small fraction of the trials (Section 4.3).

4.1. Farwell and Donchin's Periodic Speller Hypothesis

In this paper, we put to the test the hypothesis, originally formulated by Farwell and Donchin [9], that a periodic highlighting of stimuli in a BCI speller might elicit slow preparatory potentials (e.g., CNVs), in addition to P300 ERPs, and that this might result in improved classification performance.

In doing so, we did not exactly follow Farewell and Donchin's suggestion to sequentially flash the rows and columns of a matrix speller, because of the issues associated with this paradigm discussed in Section 1, which later research on RSVP and SD spelling approaches sought to address. Instead, we adopted a circular organisation of the letters, inspired by our earlier work on a sequential BCI mouse [82] and somehow reminiscent of the LSC speller [63] (where each letter had 4 neighbours and not 2, like in our speller) and we flashed letters sequentially one at a time (unlike [63] where letters were flashed randomly, the only periodicity being the alternation of flashes between the left and right side of the screen).

Another departure from [9] is that our system also made use of colour, specifically in the intensification of the letters. However, unlike previous speller protocols where colour was only used to make stimuli more salient (e.g., [49]) and the mental task was still counting the number of highlights of the target, in our new approach we used a mental task (silently naming the colour) that required paying attention to the colour of the target characters highlighting, which was randomly green or red. This might not seem a big difference, but in [82], when stimulation was periodic, we found that the mental task of counting target flashes *did not elicit any P300 ERPs while mentally naming the colour of targets produced the largest P300s.*♯

With the colour naming task, we found that our protocol not only elicited robust CNVs and P300s in the presence of targets, as originally predicted in [9], but also a *wider spectrum of statistically significant ERSCPs* (some positive and some negative) both preceding the target presentation but also following the P300. The fact that such ERSCPs can be exploited to increase the statistical separation between targets and non-targets via a simple baseline correction, and the classification results showing the superiority of the 2s Circular speller, further *confirm that Farwell and Donchin's were right in hoping that the combination of P300 and SCPs might lead to improved classification performance.*

4.2. ERSCPs, Foreperiod and Mental Task Jitter

As indicated in Section 1.5, ERSCPs are often found in studies on *temporal preparedness* and the *foreperiod effect*. The latter is studied in experiments where a warning stimulus precedes a target/imperative stimulus requiring a response [72, 73, 74, 75, 76]. The warning stimulus enhances the participant's preparation, leading to shorter reaction times (with a narrower distribution) and higher accuracy. Typically, when the time between the warning and target stimuli — *the foreperiod* — is constant and above a certain minimum threshold, reaction times tend to increase as the foreperiod lengthens, likely due to difficulties in precisely estimating elapsed time. Participants are better able to anticipate and allocate attentional resources for shorter constant foreperiods, leading to less variability in reaction times. However, if the foreperiod is variable (randomly distributed), longer foreperiods, while still associated with increased reaction times and reduced accuracy compared to shorter ones, do not necessarily produce the same decline in performance seen in long constant foreperiods. This is because

---

♯ It is not inconceivable that in the past 3+ decades other researchers attempted to explore Farwell and Donchin's suggestion of periodically flashing rows and columns of the speller to improve classification accuracy, only to find that accuracy was worse due to the "vanishing of the P300s" associated with the reported negative interaction between periodicity and the traditional target-flash counting task.

participants understand that the probability of the target appearing increases the longer they have waited (consistent with the hazard function), leading to a gradual increase in preparedness over time.

In both Donchin's speller and the Circular speller, it seems reasonable to *consider the previous-target flash both as an imperative stimulus* (requiring the mental task to be carried out) *and as a warning stimulus for the following target.* As a consequence *the temporal distance between two consecutive targets (TTI) can be seen as a foreperiod.* If we accept this view, based on what we discussed above, we should expect more variability in the timing of the mental tasks (akin to the response time in preparedness studies), and also in the corresponding ERPs, with longer foreperiods, i.e., in Experiment 3s than in Experiment 2s.

Naturally, *for the Circular spellers the foreperiod is constant* (2,000 and 3,000ms, respectively), and so we should expect participants to be able to do reasonable preparations but to struggle more with timing the focusing of their attentional resources at the slower presentation rate than with the faster. This may lead to more mental task jitter and thus misalignment of ERPs and ERSCPs across epochs in the slower Circular than in the faster one. This might be a contributing reason as to why we see both higher statistical significance and better classification performance with the Circular speller at the faster presentation rate.

4.3. Potential Evidence of ERSCPs in Donchin's speller

In Donchin's case, the foreperiod/TTI is a random variable following a very skewed distribution with first decile $2 \times SOA$, first quartile $3 \times SOA$, median/mean $6 \times SOA$, third quartile $8 \times SOA$ and tenth decile $11 \times SOA$. Only for 8% of the random values $TTI \geq 12 \times SOA$, i.e., comparable with the Circular TTIs.

So, while some preparatory mental activity might also take place with Donchin's speller, it can only really happen in a fraction of the trials. We hypothesise that in those trials pre-stimulus ERSCPs will be elicited, including the CNV, and that they are the cause of the small shift in baseline observable in most channels of grand-averages before stimulus onset, in the absence of baseline correction (see Figure 4 top). The shift is most prominent in Cz, where it is approximately -0.5μV.

The shifts are small enough to *not* change significantly the morphology of the grand-averages upon application of baseline correction and are *not* statistically significant at grand-average level.[††] However, they do produce quite dramatic improvements in statistical significance in the baselined data (see Figure 8 bottom), including at the level of post-P300 ERSCPs. We believe that this is indirect evidence that also Donchin's speller elicits preparatory ERSCPs. We also hypothesise that pre-stimulus ERSCPs paired up with baseline correction may be a contributing factor to the well-known P300

[††] At individual-average level, only in one participant in Experiment 2s and one in Experiment 3s pre-stimulus negativity in centro-parietal channels was statistically significant (spatio-temporal cluster permutation test, data not reported).

amplitude increases associated with longer TTI in Donchin's speller [36].

5. Conclusions, Limitations and Future Work

In this section we summarise the main contributions of the work (Section 5.1), we then discuss its limitations and how we propose to address them in future research (Section 5.2), and, finally, we reflect on the outlook for future BCI systems jointly exploiting ERPs and event-related slow cortical potentials (Section 5.3).

5.1. Main Contributions

In this paper, we have introduced a new form of EEG-based speller which, thanks to a periodic stimulation paradigm, *elicits both P300 ERPs and slow cortical potentials*. That a speller of this type might be viable was originally postulated by Farwell and Donchin as possible future work in their 1988 seminal paper [9]. However, no published work appears to have implemented and tested this idea, with much research, instead, focusing on incremental variations of the original row-column oddball speller.

For our speller we chose a circular arrangement of letters, adopting a single-display approach for their highlighting, and we highlighted them in colour, following our successful experience with the development of a BCI mouse [82]. Stimuli were flashed sequentially with very short ISI (83ms and 55ms). However, and we think this was a major ingredient in the success of the approach, we *modified the mental task* from the traditional target-flash counting to the *silent naming of the colour* of target flashes.

Results showed that our sequential speller produces not only clear P300s, but also equally clear SCPs, including CNVs as postulated by Farwell and Donchin. Surprisingly, the SCPs were both before stimulus presentation and after the P300 ERP. Because these SCPs are not voluntary modulated but elicited by an event (the presentation of a target, either just occurred or expected and about to occur), we termed them *Event-Related Slow Cortical Potentials* (ERSCPs).

We found that by simply baseline correcting epochs, we could exploit the pre-stimulus ERSCPs to markedly boost the statistical significance of both the ERPs and the post-P300 ERSCPs.

At the slower stimulation rate, the joint availability of P300s and ERSCPs in the Circular speller did not provide significant improvements of accuracy and ITR w.r.t. Donchin's speller. However, at the faster presentation rate, the Circular speller had much better AUC (0.92), single-trial classification accuracy (48%), ITR (48 bits/min) and utility (35 bits/min) compared to Donchin's speller at both presentation rates and the slower Circular speller.

Surprisingly, we also found that even with very small training sets (requiring 3 to 4 minutes of user time to be acquired), all spellers were able to deliver acceptable results, with the Circular speller dominating with an AUC of 0.87, and ITR of 26 bits/min and a utility of 15 bits/min. Of course, performance continued to grow with the training set size, with the AUC of the 2s Circular speller being at least 7% superior to that of the others, but the gap in terms of ITR and utility progressively widening.

5.2. Limitations and Future Work

In this section we list the main limitations of the work looking at how we suggest to address them in future research.

5.2.1. Patient Population and Online Testing

Major limitations of the work are that we have not tested the Circular speller in online experiments and that we have not performed tests with the target user population (patients), making addressing these the most urgent target for future work.

Associated with the latter are the likely challenges faced by patients with severe motor impairments when using the Circular speller design due to their likely difficulty with shifting and fixating their gaze on specific peripheral locations. One solution would be the use of a smaller circle at the centre of the screen (as we did with our Circular BCI mouse [82]) with 6 circles each representing 6 letters in a T9 type of approach similar to [96, 97]. Having obtained a 48% task accuracy on a 36 class problem, our expectation is that with the cascade of two 6-class problems the accuracy would still be acceptable. However, this should be confirmed by future research.

Another possibility to create a speller that could potentially be more suitable for patients who are unable to gaze would be to *present one character at a time*, in the middle of the screen, like in a RSVP speller, but *sequentially instead of randomly.* The characters would need to have some random distinguishing feature (which could be the colour, like in our speller, or any other perceptual feature, e.g.,see [83]). Participants would simply need to mentally name the distinguishing feature of target characters. This would likely trigger both ERPs and ERSCPs, and, so, should be explored in future research.

Another concern associated with the patient population is that the Circular spellers use a high stimulation frequency (12Hz and 18Hz). This may pose usability issues for patients with reduced cognitive focus or delayed processing due to prolonged bedridden states. For these patients, maintaining engagement with rapid flashes may be challenging. While this is clearly an issue that will need direct experimental confirmation, we would like to point out that, for patients who are able to saccade, in the Circular speller the majority of the non-target stimuli will be in the peripheral vision (unlike in Donchin's speller), perhaps ameliorating the issue. Also, as discussed in Section 4.3, while in Donchin's speller target rows and columns flash on average with a TTI of $6 \times SOA$, with many being much shorter, in the Circular spellers the TTI is not only constant but twice as long ($12 \times SOA$), making it possible for patients to time their attentional resources. Future research will need to explore whether this is less or more tiring than maintaining focus with Donchin's speller.

5.2.2. Stimuli Arrangement and Speller Scalability

Another limitation is the potentially ineffective space utilisation with a speller that occupies a sizeable portion of the screen and its practical efficiency especially if more symbols needed to be accommodated. In preliminary research at the beginning of the project, we considered also matrix spellers with the sequential highlighting of characters following scanning or zigzagging patterns. We discarded them in favour of the circular pattern to avoid inhomogeneities of ERPs associated with row/column starts/ends or turning points. However, the potential benefits in terms of space utilisation and scalability might outweigh any loss in accuracy associated to such inhomogeneities. So, sequential matrix-based alternatives should be explored in future research.

In our prior research with a circular arrangement of 8 stimuli for a BCI mouse system [82], we were able to probe the relative importance of the silent naming of target colours vs the traditional flash counting. Specifically, in [82] we compared 4 stimulation conditions all making use of stimuli with randomly determined colour (red or green): (1) periodic flashing with silent naming of the colour, (2) periodic flashing with counting target flashes, (3) random flashing with silent naming, and (4) random flashing with counting. Of these, the two with the counting task (2 and 4) were the worst in terms of classification performance, with the periodic stimulation with silent naming (1) being the best, and random stimulation with silent naming (3) the second best. So, it is also entirely possible that a form of Donchin's speller where rows and columns are highlighted in random order but in colour, with users mentally naming the colour of the targets (rather than counting the number of times they flash), might provide many of the benefits of the Circular speller to a matrix speller. We explored this idea in preliminary research at the beginning of the project, considering both the option where the characters in a row/column flashed had random colours, and the option where all characters in a row/column flashed had the same (but random) colour. These two alternatives were not pursued, but should be tested in future research.

An alternative way of making better use of space, taking inspiration from the approach in [63], might be to spatially organise the stimuli (e.g., in two concentric circles) in such a way to maximise the distance between consecutive flashes (e.g. highlighting the characters in the inner circle out of sync with the characters in the outer circle), so as to minimise near target responses. This too should be explored.

5.2.3. Technical Enhancements

Another limitation of the study is that we used epochs of the same duration for all spellers and presentation rates, which may be quite suboptimal. For instance, it is possible that Donchin's spellers might have benefited from shorter epochs given that after baseline correction there is no ERP or ERSCP that is statistically significant before 100ms from target onset. On the contrary, using longer epochs might have benefited Circular speller BCIs. Also, it is clear that optimising epoch duration on a participant by participant basis might provide better results. All these options should be explored in future research.

A further limitation of the study is the utilisation of only two presentation rates. Our purpose with this first study was to compare benefits and drawbacks with the best established paradigm: Donchin's. The links we have identified with the literature on preparedness and the foreperiod effect (Section 4.2) make it possible to predict that slower presentation rates than for Experiment 3s are unlikely to further improve accuracy (and very unlikely to improve ITR). However, faster presentation rates worked well for our periodic protocol for pointer control [82] where TTIs of 800ms yielded a mean AUC of 0.91. So, it is not unlikely that faster presentation rates than in Experiment 2s may work well also for a speller, and should be explored in future research.

In the Circular spellers, we use temporally overlapping stimuli (three characters are highlighted at any given time), which is uncommon in BCI. This choice was motivated by the desire to make it easier to compare the Circular speller with Donchin's. So, we wanted stimuli to be highlighted on the screen for the same duration as for Donchin's speller while also having identical character selection times. Temporal overlapping stimuli, such as ours, have previously been used in [62] (see Section 1.4), where new characters were highlighted every 30ms, but once highlighted they remained so for 100ms. So, there were up to four highlighted characters on the screen at any given time. The good performance obtained both in [62] and in this article suggests that there is no major issue with temporal overlaps. However, it is conceivable that the nature and duration of the overlaps may affect ERPs and ERSCPs. This is an issue that will need to be investigated in future research.

Finally, we should emphasise that we limited our analysis to 19 channels. It is likely that statistically significant difference between target and non-target epochs, at the level of both ERPs and SCPs, are also present in other channels. Including a wider set of channels might also further improve the performance of all the spellers we tested. We will explore this in future research.

5.3. Future ERSCP-based Spellers

The use of slow cortical potentials in BCI has typically required long training periods for participants, with a proportion of them still failing to control their SCPs. With the Circular speller, instead, we were able to trigger and make use of natural ERSCPs instead of trained ones. However, we think that with our circular speller *we have barely scratched the surface*.

In addition to exploring the avenues for future work highlighted above, *future research will likely identify new and possibly better ways of making use of ERSCPs*, both alone and in combination with ERPs, for both spelling and other BCI applications. As indicated by the examples aprovided in the previous sections, producing new paradigms that do this may be relatively easy.

Acknowledgments

The authors would like to warmly thank the anonymous reviewers and the editor for their extremely helpful and detailed comments. They helped enormously in improving the quality of the manuscript.

References

[1] H. Cecotti, "Spelling with non-invasive brain–computer interfaces–current and future trends," *Journal of Physiology-Paris*, vol. 105, no. 1-3, pp. 106–114, 2011.

[2] S. Karan, "A literature survey on the contemporary methodologies used in brain computer interface for spelling application," in *2013 International Conference on Human Computer Interactions (ICHCI)*, pp. 1–5, IEEE, 2013.

[3] A. Rezeika, M. Benda, P. Stawicki, F. Gembler, A. Saboor, and I. Volosyak, "Brain–computer interface spellers: A review," *Brain sciences*, vol. 8, no. 4, p. 57, 2018.

[4] M.-h. Shi, C.-l. Zhou, J. Xie, S.-z. Li, Q.-y. Hong, M. Jiang, F. Chao, W.-f. Ren, X.-q. Liu, D.-j. Zhou, *et al.*, "Electroencephalogram-based brain-computer interface for the chinese spelling system: a survey," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 3, pp. 423–436, 2018.

[5] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of EEG-based brain–computer interface paradigms," *Journal of neural engineering*, vol. 16, no. 1, p. 011001, 2019.

[6] M. Li, D. He, C. Li, and S. Qi, "Brain–computer interface speller based on steady-state visual evoked potential: a review focusing on the stimulus paradigm and performance," *Brain Sciences*, vol. 11, no. 4, p. 450, 2021.

[7] T. Fang, Z. Song, L. Niu, S. Le, Y. Zhang, X. Zhang, G. Zhan, S. Wang, H. Li, Y. Lin, *et al.*, "Recent advances of P300 speller paradigms and algorithms," in *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, pp. 1–6, IEEE, 2021.

[8] J. Pan, X. Chen, N. Ban, J. He, J. Chen, and H. Huang, "Advances in P300 brain–computer interface spellers: toward paradigm design and performance evaluation," *Frontiers in Human Neuroscience*, vol. 16, 2022.

[9] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[10] S. J. Luck, *An introduction to the event-related potential technique.* MIT press, 2014.

[11] T. Hinterberger, S. Schmidt, N. Neumann, J. Mellinger, B. Blankertz, G. Curio, and N. Birbaumer, "Brain-computer communication and slow cortical potentials," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1011–1018, 2004.

[12] H. Cecotti, "A self-paced and calibration-less SSVEP-based brain–computer interface speller," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 18, no. 2, pp. 127–133, 2010.

[13] A. Furdea, S. Halder, D. Krusienski, D. Bross, F. Nijboer, N. Birbaumer, and A. Kübler, "An auditory oddball (P300) spelling system for brain-computer interfaces," *Psychophysiology*, vol. 46, no. 3, pp. 617–625, 2009.

[14] J. Höhne, M. Schreuder, B. Blankertz, and M. Tangermann, "A novel 9-class auditory ERP paradigm driving a predictive text entry system," *Frontiers in neuroscience*, vol. 5, p. 99, 2011.

[15] A.-M. Brouwer and J. B. Van Erp, "A tactile P300 brain-computer interface," *Frontiers in neuroscience*, p. 19, 2010.

[16] M. van der Waal, M. Severens, J. Geuze, and P. Desain, "Introducing the tactile speller: an ERP-

based brain–computer interface for communication," *Journal of Neural Engineering*, vol. 9, no. 4, p. 045002, 2012.

[17] E. A. Katyal and R. Singla, "EEG-based hybrid qwerty mental speller with high information transfer rate," *Medical & Biological Engineering & Computing*, vol. 59, pp. 633–661, 2021.

[18] J. Han, M. Xu, X. Xiao, W. Yi, T.-P. Jung, and D. Ming, "A high-speed hybrid brain-computer interface with more than 200 targets," *Journal of Neural Engineering*, vol. 20, no. 1, p. 016025, 2023.

[19] A. Belitski, J. Farquhar, and P. Desain, "P300 audio-visual speller," *Journal of neural engineering*, vol. 8, no. 2, p. 025022, 2011.

[20] E. Donchin and M. G. H. Coles, "Is the P300 a manifestation of context updating?," *Behavioral and brain sciences*, vol. 11, pp. 355–372, 1988.

[21] J. Polich and M. D. Comerchero, "P3a from visual stimuli: typicality, task, and topography," *Brain topography*, vol. 15, no. 3, pp. 141–152, 2003.

[22] J. Polich, "Neuropsychology of P3a and P3b: A theoretical overview," in *Brainwaves and mind: recent developments* (N. C. Moore and K. Arikan, eds.), pp. 15–29, Kjellberg Inc., 2004.

[23] J. Polich, "Updating P300: an integrative theory of P3a and P3b," *Clinical neurophysiology*, vol. 118, no. 10, pp. 2128–2148, 2007.

[24] S. Sutton, M. Braren, J. Zubin, and E. John, "Evoked-potential correlates of stimulus uncertainty," *Science*, vol. 150, no. 3700, pp. 1187–1188, 1965.

[25] K. C. Squires, N. K. Squires, and S. A. Hillyard, "Decision-related cortical potentials during an auditory signal detection task with cued observation intervals.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 1, no. 3, p. 268, 1975.

[26] K. C. Squires, E. Donchin, R. I. Herning, and G. McCarthy, "On the influence of task relevance and stimulus probability on event-related-potential components," *Electroencephalography and clinical neurophysiology*, vol. 42, no. 1, pp. 1–14, 1977.

[27] J. Polich, "Probability and inter-stimulus interval effects on the P300 from auditory stimuli.," *International journal of psychophysiology*, vol. 10, pp. 163–170, Dec 1990.

[28] P. G. Fitzgerald and T. W. Picton, "Temporal and sequential probability in evoked potential studies.," *Canadian journal of psychology*, vol. 35, pp. 188–200, Jun 1981.

[29] B. Z. Allison and J. A. Pineda, "Effects of SOA and flash pattern manipulations on ERPs, performance, and preference: implications for a BCI system.," *International journal of psychophysiology*, vol. 59, pp. 127–140, Feb 2006.

[30] K. C. Squires, C. Wickens, N. K. Squires, and E. Donchin, "The effect of stimulus sequence on the waveform of the cortical event-related potential.," *Science (New York, N.Y.)*, vol. 193, pp. 1142–1146, Sep 1976.

[31] C. J. Gonsalvez, E. Gordon, J. Anderson, G. Pettigrew, R. J. Barry, C. Rennie, and R. Meares, "Numbers of preceding nontargets differentially affect responses to targets in normal volunteers and patients with schizophrenia: a study of event-related potentials.," *Psychiatry research*, vol. 58, pp. 69–75, Sep 1995.

[32] C. L. Gonsalvez and J. Polich, "P300 amplitude is determined by target-to-target interval.," *Psychophysiology*, vol. 39, pp. 388–396, May 2002.

[33] R. J. Croft, C. J. Gonsalvez, C. Gabriel, and R. J. Barry, "Target-to-target interval versus probability effects on P300 in one- and two-tone tasks.," *Psychophysiology*, vol. 40, pp. 322–328, May 2003.

[34] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-computer interface technology: a review of the first international meeting.," *IEEE transactions on rehabilitation engineering*, vol. 8, pp. 164–173, Jun 2000.

[35] A. Schlögl, C. Keinrath, R. Scherer, and G. Pfurtscheller, "Information transfer of an EEG-based brain computer interface," in *Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering*, pp. 641–644, 2003.

[36] L. Citi, R. Poli, and C. Cinel, "Documenting, modelling and exploiting P300 amplitude changes due to variable target delays in Donchin's speller," *Journal of Neural Engineering*, vol. 7, no. 5, p. 056006, 2010.

[37] J. E. Raymond, K. L. Shapiro, and K. M. Arnell, "Temporary suppression of visual processing in an RSVP task: An attentional blink?," *Journal of experimental psychology: Human perception and performance*, vol. 18, no. 3, p. 849, 1992.

[38] N. G. Kanwisher, "Repetition blindness: Type recognition without token individuation," *Cognition*, vol. 27, no. 2, pp. 117–143, 1987.

[39] C. Cinel, R. Poli, and L. Citi, "Possible sources of perceptual errors in P300-based speller paradigm," *Biomedizinische Technik*, vol. 49, pp. 39–40, 2004. Proceedings of the 2nd International BCI workshop and training course.

[40] M. Salvaris and F. Sepulveda, "Perceptual errors in the Farwell and Donchin matrix speller," in *2009 4th International IEEE/EMBS Conference on Neural Engineering*, pp. 275–278, 2009.

[41] G. Townsend, B. K. LaPallo, C. B. Boulay, D. J. Krusienski, G. Frye, C. Hauser, N. E. Schwartz, T. M. Vaughan, J. R. Wolpaw, and E. W. Sellers, "A novel P300-based brain–computer interface stimulus presentation paradigm: moving beyond rows and columns," *Clinical neurophysiology*, vol. 121, no. 7, pp. 1109–1120, 2010.

[42] S.-K. Yeom, S. Fazli, and S.-W. Lee, "P300 visual speller based on random set presentation," in *2014 International Winter Workshop on Brain-Computer Interface (BCI)*, pp. 1–2, IEEE, 2014.

[43] T. Kaufmann, S. M. Schulz, C. Grünzinger, and A. Kübler, "Flashing characters with famous faces improves ERP-based brain–computer interface performance," *Journal of neural engineering*, vol. 8, no. 5, p. 056016, 2011.

[44] S.-K. Yeom, S. Fazli, K.-R. Müller, and S.-W. Lee, "An efficient ERP-based brain-computer interface using random set presentation and face familiarity," *PloS one*, vol. 9, no. 11, p. e111157, 2014.

[45] T. Kaufmann and A. Kübler, "Beyond maximum speed—a novel two-stimulus paradigm for brain–computer interfaces based on event-related potentials (P300-BCI)," *Journal of neural engineering*, vol. 11, no. 5, p. 056004, 2014.

[46] Z. Lu, Q. Li, N. Gao, and J. Yang, "The self-face paradigm improves the performance of the P300-speller system," *Frontiers in computational neuroscience*, vol. 13, p. 93, 2020.

[47] Y. Liu, Z. Zhou, and D. Hu, "Comparison of stimulus types in visual P300 speller of brain-computer interfaces," in *9th IEEE International Conference on Cognitive Informatics (ICCI'10)*, pp. 273–279, IEEE, 2010.

[48] Q. T. Obeidat, T. A. Campbell, and J. Kong, "Introducing the edges paradigm: a P300 brain–computer interface for spelling written words," *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 6, pp. 727–738, 2015.

[49] D. Ryan, G. Townsend, N. Gates, K. Colwell, and E. Sellers, "Evaluating brain-computer interface performance using color in the P300 checkerboard speller," *Clinical Neurophysiology*, vol. 128, no. 10, pp. 2050–2057, 2017.

[50] B. Z. Allison and J. A. Pineda, "ERPs evoked by different matrix sizes: implications for a brain computer interface (BCI) system.," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 11, pp. 110–113, Jun 2003.

[51] E. W. Sellers, D. J. Krusienski, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A P300 event-related potential brain-computer interface (BCI): the effects of matrix size and inter stimulus interval on performance.," *Biological psychology*, vol. 73, pp. 242–252, Oct 2006.

[52] M. Salvaris and F. Sepulveda, "Visual modifications on the P300 speller BCI paradigm," *Journal of neural engineering*, vol. 6, no. 4, p. 046011, 2009.

[53] J. Du, Y. Ke, L. Kong, T. Wang, F. He, and D. Ming, "3D stimulus presentation of ERP-speller in virtual reality," in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 167–170, IEEE, 2019.

[54] O. E. Korkmaz, O. Aydemir, E. A. Oral, and I. Y. Ozbek, "An efficient 3D column-only P300 speller

paradigm utilizing few numbers of electrodes and flashings for practical BCI implementation," *PloS one*, vol. 17, no. 4, p. e0265904, 2022.

[55] J.-h. Shi, J.-z. Shen, Y. Ji, and F.-l. Du, "A submatrix-based P300 brain-computer interface stimulus presentation paradigm," *Journal of Zhejiang University SCIENCE C*, vol. 13, pp. 452–459, 2012.

[56] L. Acqualagna, M. S. Treder, M. Schreuder, and B. Blankertz, "A novel brain-computer interface based on the rapid serial visual presentation paradigm," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 2686–2689, IEEE, 2010.

[57] L. Acqualagna and B. Blankertz, "Gaze-independent BCI-spelling using rapid serial visual presentation (RSVP)," *Clinical Neurophysiology*, vol. 124, no. 5, pp. 901–908, 2013.

[58] S. Chennu, A. Alsufyani, M. Filetti, A. M. Owen, and H. Bowman, "The cost of space independence in P300-BCI spellers," *Journal of neuroengineering and rehabilitation*, vol. 10, pp. 1–13, 2013.

[59] Z. Lin, C. Zhang, Y. Zeng, L. Tong, and B. Yan, "A novel P300 BCI speller based on the triple RSVP paradigm," *Scientific reports*, vol. 8, no. 1, p. 3350, 2018.

[60] C. Guan, M. Thulasidas, and J. Wu, "High performance P300 speller for brain-computer interface," in *IEEE International Workshop on Biomedical Circuits and Systems, 2004.*, pp. S3–5, IEEE, 2004.

[61] J. Pan, Y. Li, T. Yu, *et al.*, "A comparison of P300-speller stimuli presentation paradigms for brain–computer interface," in *3rd Annual Summit and Conf. of Asia Pacific Signal and Information Processing Association*, 2011.

[62] J. Qu, F. Wang, Z. Xia, T. Yu, J. Xiao, Z. Yu, Z. Gu, and Y. Li, "A novel three-dimensional P300 speller based on stereo visual stimuli," *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 4, pp. 392–399, 2018.

[63] G. Pires, U. Nunes, and M. Castelo-Branco, "Comparison of a row-column speller vs. a novel lateral single-character speller: Assessment of BCI for severe motor disabled patients," *Clinical Neurophysiology*, vol. 123, no. 6, pp. 1168–1181, 2012.

[64] N. Birbaumer, "Slow cortical potentials: plasticity, operant control, and behavioral effects," *The Neuroscientist*, vol. 5, no. 2, pp. 74–78, 1999.

[65] A. Kübler, B. Kotchoubey, T. Hinterberger, N. Ghanayim, J. Perelmouter, M. Schauer, C. Fritsch, E. Taub, and N. Birbaumer, "The thought translation device: a neurophysiological approach to communication in total motor paralysis," *Experimental brain research*, vol. 124, pp. 223–232, 1999.

[66] M. Pham, T. Hinterberger, N. Neumann, A. Kübler, N. Hofmayer, A. Grether, B. Wilhelm, J.-J. Vatine, and N. Birbaumer, "An auditory brain-computer interface based on the self-regulation of slow cortical potentials.," *Neurorehabilitation and neural repair*, vol. 19, pp. 206–218, Sep 2005.

[67] T. Elbert, *Slow cortical potentials reflect the regulation of cortical excitability.* Springer, 1993.

[68] W. G. Walter, R. Cooper, V. Aldridge, W. McCallum, and A. Winter, "Contingent negative variation: an electric sign of sensori-motor association and expectancy in the human brain," *Nature*, vol. 203, no. 4943, pp. 380–384, 1964.

[69] J. J. Tecce, "Contingent negative variation (CNV) and psychological processes in man.," *Psychological bulletin*, vol. 77, no. 2, p. 73, 1972.

[70] H. H. Kornhuber and L. Deecke, "Hirnpotentialänderungen bei willkürbewegungen und passiven bewegungen des menschen: Bereitschaftspotential und reafferente potentiale," *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, vol. 284, pp. 1–17, 1965.

[71] H. H. Kornhuber and L. Deecke, "Brain potential changes in voluntary and passive movements in humans: readiness potential and reafferent potentials," *Pflügers Archiv-European Journal of Physiology*, vol. 468, pp. 1115–1124, 2016.

[72] A. Gaillard and R. Näätänen, "Slow potential changes and choice reaction time as a function of interstimulus interval," *Acta Psychologica*, vol. 37, no. 3, pp. 173–186, 1973.

[73] P. Niemi and R. Näätänen, "Foreperiod and simple reaction time.," *Psychological bulletin*, vol. 89,

no. 1, p. 133, 1981.

[74] M. B. Steinborn, B. Rolke, D. Bratzke, and R. Ulrich, "Sequential effects within a short foreperiod context: Evidence for the conditioning account of temporal preparation," *Acta Psychologica*, vol. 129, no. 2, pp. 297–307, 2008.

[75] R. Langner, M. B. Steinborn, S. B. Eickhoff, and L. Huestegge, "When specific action biases meet nonspecific preparation: Event repetition modulates the variable-foreperiod effect.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 44, no. 9, p. 1313, 2018.

[76] B. Rolke and P. Hofmann, "Temporal uncertainty degrades perceptual processing," *Psychonomic bulletin & review*, vol. 14, no. 3, pp. 522–526, 2007.

[77] M. Donald, *Electrocortical correlates of fixed-foreperiod decision tasks.* PhD thesis, McGill University, 1968.

[78] B. Libet, B. Libet, C. A. Gleason, E. W. Wright, and D. K. Pearl, "Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential) the unconscious initiation of a freely voluntary act," *Neurophysiology of consciousness*, pp. 249–268, 1993.

[79] B. Libet, "Unconscious cerebral initiative and the role of conscious will in voluntary action," *Behavioral and brain sciences*, vol. 8, no. 4, pp. 529–539, 1985.

[80] B. Libet, E. W. Wright Jr, and C. A. Gleason, "Preparation-or intention-to-act, in relation to pre-event potentials recorded at the vertex," *Electroencephalography and clinical Neurophysiology*, vol. 56, no. 4, pp. 367–372, 1983.

[81] M. Raś, A. Nowik, A. Klawiter, and G. Króliczak, "When is the brain ready for mental actions? Readiness potential for mental calculations," *Acta Neurobiologiae Experimentalis*, vol. 79, no. 4, pp. 386–398, 2019.

[82] M. Salvaris, C. Cinel, L. Citi, and R. Poli, "Novel protocols for P300-based brain–computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 20, no. 1, pp. 8–17, 2012.

[83] M. Salvaris, C. Cinel, R. Poli, L. Citi, and F. Sepulveda, "Exploring multiple protocols for a brain-computer interface mouse," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pp. 4189–4192, IEEE, 2010.

[84] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, *et al.*, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroinformatics*, vol. 7, p. 267, 2013.

[85] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "ICLabel: An automated electroencephalographic independent component classifier, dataset, and website," *NeuroImage*, vol. 198, p. 181–197, Sept. 2019.

[86] D. Yao, "A method to standardize a reference of scalp EEG recordings to a point at infinity," *Physiological measurement*, vol. 22, no. 4, p. 693, 2001.

[87] E. Maris and R. Oostenveld, "Nonparametric statistical testing of EEG-and MEG-data," *Journal of neuroscience methods*, vol. 164, no. 1, pp. 177–190, 2007.

[88] S. M. Smith and T. E. Nichols, "Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference," *Neuroimage*, vol. 44, no. 1, pp. 83–98, 2009.

[89] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, "xDAWN algorithm to enhance evoked potentials: application to brain–computer interface," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 8, pp. 2035–2043, 2009.

[90] G. LemaÃŽtre, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of machine learning research*, vol. 18, no. 17, pp. 1–5, 2017.

[91] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," *Advances in neural information processing systems*, vol. 31, 2018.

[92] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–

computer interfaces for communication and control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[93] B. Dal Seno, M. Matteucci, and L. T. Mainardi, "The utility metric: a novel method to assess the overall performance of discrete brain–computer interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 1, pp. 20–28, 2009.

[94] C. C. Duncan-Johnson and E. Donchin, "On quantifying surprise: The variation of event-related potentials with subjective probability," *Psychophysiology*, vol. 14, no. 5, pp. 456–467, 1977.

[95] C. Tremmel, J. Fernandez-Vargas, D. Stamos, C. Cinel, M. Pontil, L. Citi, and R. Poli, "A meta-learning BCI for estimating decision confidence," *Journal of Neural Engineering*, vol. 19, no. 4, p. 046009, 2022.

[96] R. Ron-Angevin, S. Varona-Moya, and L. da Silva-Sauer, "Initial test of a T9-like P300-based speller by an ALS patient," *Journal of neural engineering*, vol. 12, no. 4, p. 046023, 2015.

[97] F. Akram, S. M. Han, and T.-S. Kim, "An efficient word typing P300-BCI system using a modified T9 interface and random forest classifier," *Computers in biology and medicine*, vol. 56, pp. 30–36, 2015.