Closed-loop EEG study on visual recognition during driving

Ruslan Aydarkhanov[1], Marija Ušćumlić[2], Ricardo Chavarriaga[3,4], Lucian Gheorghe[5], José del R Millán[3,6,7]

[1]Medical Image Processing Laboratory, Center for Neuroprosthetics, Interschool Institute of Bioengineering, École Polytechnique Fédérale de Lausanne (EPFL), Campus Biotech H4, 1202 Geneva, Switzerland

[2]Nissan International SA, La Pièce 12, 1180 Rolle, Switzerland

[3]École Polytechnique Fédérale de Lausanne (EPFL), Campus Biotech H4, 1202 Geneva, Switzerland

[4]ZHAW Datalab, Zurich University of Applied Sciences, Winterthur, Switzerland

[5] Advanced Materials and Processing Laboratory, Nissan Research Center, Nissan Motors Co. LTD, 1, Natsushima, Yokosuka-shi, Kanagawa-ken, 237-8523, Japan

[6]Dept. of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712, USA

[7]Dept. of Neurology, The University of Texas at Austin, Austin, TX 78712, USA

E-mail: `ruslan.aydarkhanov@epfl.ch`

Abstract. *Objective.* In contrast to the classical visual BCI paradigms which adhere to a rigid trial structure and restricted user behavior, EEG-based visual recognition decoding during our daily activities remains challenging. The objective of this study is to explore the feasibility to decode the EEG signature of visual recognition in experimental conditions promoting our natural ocular behavior when interacting with our dynamic environment. *Approach.* In our experiment, subjects visually search for a target object among suddenly appearing objects in the environment while driving a car-simulator. Given that subjects exhibit an unconstrained overt visual behavior, we based our study on eye fixation related potentials (EFRP). We report on visual behavior and single-trial EFRP decoding performance (fixations on task-related vs. distracting objects). In addition, we demonstrate the application of our approach in a closed-loop BCI setup. *Main results.* By integrating decoding probabilities of multiple EFRP the online accuracy reached 0.37 on average in identifying one target out of four board types. Using the acquired data, we performed a comparative study of classification algorithms and feature spaces in a simulated online scenario. The EEG approaches yielded similar moderate performances not higher than 0.6 AUC, yet significantly above the chance level. In addition, the gaze duration (dwell time) appears to be an additional informative feature in this context. *Significance.* These results show that visual recognition of sudden events can be decoded during active driving. It lays a foundation for assistive and recommender systems based on driver's brain signals.

Keywords: Brain-computer interfaces, Electroencephalography, Eye tracking, Driving,

1. Introduction

Brain-computer interfaces (BCI), especially those based on non-invasive electroencephalogram (EEG) signals, are not only proving their value as assistive tools for people with disabilities [1, 2, 3, 4, 5, 6] and potential rehabilitation tools for neurological patients [7, 8, 9, 10] but also open opportunities to augment interaction for users without disabilities [11, 12, 13]. In this later respect, a promising possibility is to decode neural correlates of perceptual and cognitive processes while people overtly interact with real-world environments [14, 15, 16]. In this work, we explore the feasibility to decode visual recognition from EEG in experimental conditions mimicking our daily activities which involve overt visual search. In particular, we performed a study on visual BCI in driving – aiming at decoding visual recognition while a driver is primarily engaged in controlling a car simulator, actively exploring the visual environment to complete a visual search task.

Previous works have reported certain EEG signatures of visual recognition in simple, well-controlled experimental setups [17, 18, 19, 15]. The classical restrictive conditions require that the stimuli are static and undergo sharp transitions (e.g. flashing) as well as subjects sit still in front of a screen [18]. In such experiments with an oddball paradigm where subjects had to recognize rare targets stimuli in a sequence, the recognition process is reflected in EEG as the well-known P300 component of Event-Related Potential (ERP). P300 is a positive deflection in the parietal region which occurs approximately between 250 and 500 ms after the stimulus onset [20]. This signal has been successfully used in various BCI applications such as P300-based speller because of high decoding performance. With the typing speed of 10 characters per minute [21], the home use of such spellers may improve the quality of life of people with severe motor disabilities, such as ALS [22, 23, 24]. For people without motor impairments, however, this setup and typing performance does not bring much value.

Successful decoding of visual recognition in free viewing tasks would open space for BCI application for healthy users by providing them assistance and recommendations. Car drivers are exposed to a richer environment and need to constantly interact with it by controlling the car, visually exploring the surrounding and planning upcoming actions. A hypothetical application scenario may include the driver's recognition of road signs of a particular target category, e.g. parking. The driver's interest in parking signs decoded from the EEG will allow to provide timely recommendations for the nearest parking.

In free viewing conditions, people need to fixate or track moving objects to perceive them in all details. Fixations evoke a type of ERP called Eye Fixation Related Potentials (EFRP) as opposed to external events in more classical ERP designs. Early EEG deflections in EFRP reflect the processing of low-level features of visual input projected from the retina to the occipital lobe of the cortex and are manifested mainly in high-amplitude of the P100 component, also referred to as lambda component [25]. Later

EEG deflections in EFRP resemble the P300 component in the oddball paradigm [26].

Visual stimuli used in previous ERP and EFRP studies range from simple geometric shapes in static scenes to static natural images or dynamic scenes with geometric shapes [27, 28]. Only a few attempts on decoding ERP or EFRP in videos or virtual reality (VR) simulations have been reported, e.g. [19, 15]. However, the experimental conditions in those studies did not fully reflect the real-world dynamics. For instance, in human action recognition from a cartoon animation, the video playback was sped up to limit the time of the recognition process, and other experimental settings made unnecessary to utilize eye-tracking as it would have been the case on more natural setups [19]. The classification performance of Target vs Non-Target event reached average AUC > 0.8. In another study involving maze navigation, subjects experienced fast autonomous driving and had to press a button whenever the car in front braked suddenly [15]. The braking task required constant monitoring of the car in front, leading to a brief visual attendance of the stimuli. It resembled the attentional load of real driving, however, unnaturally high speed and lack of full car control limit the transfer of the results to real driving. The classification performance based on single EFRP was around 0.65 AUC on average across subjects.

In this paper, we report a study on visual recognition decoding during driving in a car simulator based on the Eye Fixation Related Potentials and visual behavior. In our study 13 subjects visually attended sign boards along the roads while primarily engaged in active driving through an urban environment with a natural speed of ~ 80 km/h. We introduced, nevertheless, certain limitation concerning the stimuli presentation. The boards were invisible until the driver gets close enough, then they pop out at random locations above the sidewalks. Such popping up of the board ensured that once drivers gaze at a board, they can recognize the content. Otherwise, there would be 2 challenges in decoding visual recognition due to fixations on distant and poorly recognizable boards: 1) to locate the recognition timing within long attendance, 2) to select one of multiple returning fixations when the recognition occurs. The latter challenges collecting clean data while the former challenges the EFRP decoding itself.

The experiment consisted of 2 phases: offline and online. The goal of the offline phase was two-fold: (1) to analyze behavioral and neural correlates of visual recognition and (2) to gather data to train a classifier to be used in the online phase. Then we conducted the online phase to study the performance of the decoder in a closed-loop setting and, as a major objective, to verify the feasibility of the corresponding BCI application.

Our protocol closely resembles the real-life scenario of traffic sign recognition on the roads. In the application, the BCI system that decodes the cognitive response on the target sign recognition can be integrated into the advanced driver-assistance system (ADAS) of the car. In this case, the target is naturally chosen by the driver according to the current driving situation. The accurate decoding of the selection of target traffic signs from the drivers brainwaves can shed light on the drivers intentions and goals, from which the ADAS system could derive useful recommendations and assistance.

Figure 1. Visualization of the protocol. Top left: The experimental setup with the driving simulator (chair and 3 3D screens), Eye-tracker and subject with EEG cap. Middle left: screenshot from the beginning of the road with the cue on the floor and the first board on the left-hand side. Bottom left: screenshot with the online phase feedback for correct decoding, i.e. the 2D board at the bottom of the screen and a green circle. Right: schematic drawing of one road with the target cue, boards and finish line.

2. Experimental setup and protocol

2.1. Data collection

13 volunteers (age of 28 ± 6, 3 female) participated in the study. The study was approved by the ethical committees of the Cantons of Vaud and Geneva, Switzerland (Commission cantonale d'éthique de la recherche sur l'être humain, study no. PB_2017-00295), and all the subjects provided written consent. Experiment lasted 3 hours including the set up (∼ 1 h), the offline phase (∼ 45 min), one pause period (∼ 30 min) and the online phase (∼ 45 min). The extended pause period was necessary to process all the data from the offline phase and to train the classifier for the online phase. The offline phase consisted of 3 runs through the city whereas the online phase could comprise from 3 to 5 runs depending on the available time and subject's fatigue. One run included 20 roads with 12 boards each resulting in 240 boards per run. Before each run, the subjects were

asked to move their eyes up-down and left-right for one minute in order to collect the data for eye movement artifact removal.

The EEG was acquired with BioSemi ActiveTwo system with 64 electrodes at 2 kHz sampling rate. Additionally, we recorded 3 EOG channels to collect the eye movement data: two electrodes next to the outer canthi of the eyes and one above the nasion. The real-time processing of EEG in the online phase was done on a dedicated computer.

The eye gaze was recorded with the SMI RED Eye tracking system with sampling rate of 120 Hz. The chair and eye tracker positions were adjusted for each subject. The eye tracker was calibrated with 13 points once after the EEG setup and before the beginning of the experiment.

The driving simulator logged the car location in the virtual environment, the controllers state (gas and brake pedals, steering wheel, buttons on the wheel) and the 2D position of boards on the screen at a sampling rate of 256 Hz. In order to synchronize the data acquisition on three separate machines (EEG, eye tracking and driving simulator) at different sampling rates, a square pulse of 4 Hz was generated by the driving simulator and sent to the eye tracker through TCP connection and to BioSemi through the parallel port.

2.2. Driving simulator

In our study, we used the driving simulator previously used in [12, 11, 29]. It allows for immersive driving experience through the utilization of real Nissan driving chair with steering wheel and two pedals (gas and brake). The simulated car has an automatic transmission which excludes actions for the manual gear shift. The visual input is provided with three 3D monitors which create multiple renders for different angles. The virtual environment is implemented using the open-source driving simulator project VDrift. The environment resembles a regular grid city with static objects, i.e. building, traffic lights, fields. The task-related objects include direction indications on the road, target cue, boards with symbols and finish lines (Figure 1).

2.3. Tasks

The experimental session had 2 phases: offline and online. The data collected in the offline phase allows to train a classifier that is used in the online phase to provide feedback on the decoding results. In both phases, the subject is instructed to drive through the city while following indications left or right turn at the crossings (Figure 1). At the beginning of the road segment, a symbol (the cue) is depicted on the ground. Subjects must memorize and search for it among the boards appearing through the road segment. Subjects are asked to count the number of boards with the target symbol in each segment. To keep subjects engaged in the task, they have to report the number of targets in the offline phase or observe feedback from the decoder in the online phase of the experiment after they cross a finish line at the end of each road.

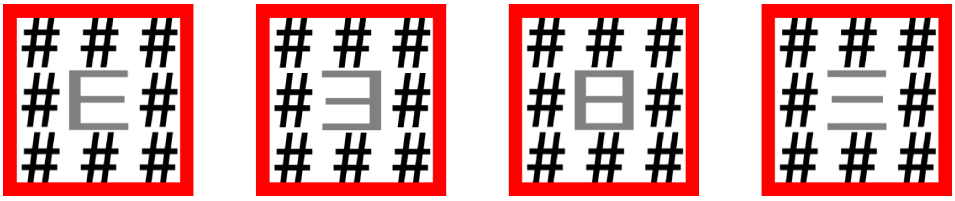

Figure 2. The boards with the used symbols. In the offline phase, we used only boards with symbol E and flipped E whereas in the online phase we used all 4 boards.

We included empty roads where subjects have no additional tasks except driving, allowing them to rest from the visual recognition task. In total, a quarter of the roads were empty.

Offline phase. After crossing the finish line the subject presses a button located at the steering wheel as many times as the number of target boards he/she counted along the current road.

Online phase. In the online phase we want to estimate the system's performance in detecting the target symbol (1 out of 4) among all presented symbols (see Section 2.4). After crossing the finish line the board with the predicted target is rendered at the bottom of the screen as a 2D object (Figure 1). Subjects are instructed to pay attention to this feedback. The online phase resembles a closed-loop interaction because the subjects perceive the decoder result shortly after looking at the stimuli, within 15 seconds after the first board on the road.

2.4. Stimuli presentation

2.4.1. Stimuli visibility. The board presentation is carefully adjusted to guide the visual behavior of subjects. First of all, boards are invisible unless the driver approaches them close enough making them pop up. Their positions are generated using the following rules. The boards appear along the road at an equal distance between them, randomly on either side of the road above the sidewalks with a maximum of 2 boards on the same side in a row (e.g., 2 boards on the left in Figure 1). The number of boards on the left and right sides is balanced. Since the pop-up distance was greater than the distance between the boards along the road, multiple boards from the same side were visible at the same time, so their horizontal and vertical positions were adjusted to avoid the overlap for the driver view.

The maximum speed of the car was limited to 80 km/h to ensure that all the targets can be attended. The subjects were allowed to slow down if it is necessary to attend all the boards and count the targets. Nonetheless, all the subjects practiced until they felt comfortable with completing the recognition task at the maximum speed during the EEG setup. Due to constant speed and regular placement of the boards, the boards popped up at a regular pace with a 900 ms period.

Table 1. The protocol differences between offline and online phases.

	Offline	Online
Types of boards	2	4
Task	Count silently Button press at the end	Count silently Observe the feedback

2.4.2. Stimuli content. In order to link the perception of the symbols on the board with the eye fixations, the recognition by peripheral vision must be avoided. Therefore, the target and distracting symbols were similar and surrounded by the # character to create a crowding effect and force the foveating on the main symbol as done in previous studies [30]. Additionally, we added a bright red border around the board similar to the traffic signs to create a contrast with the environment and facilitate their identification (Figure 2).

Symbols in the offline phase. In the offline phase, one of the two symbols were depicted on each board: E and ∃. One of them was randomly chosen as a target per road segment and was presented as the cue at the beginning of the road. There were 2-4 targets out of 12 boards on each road with the average fraction of the targets of 0.25 in total.

Symbols in the online phase. To allow equal accumulation for each symbol by keeping a low fraction of the targets, we introduced 2 more symbols resulting in 4 symbols in total. There were between 2 and 4 boards of each symbol resulting in 12 boards on the road. Only one of them was a target on each road. Therefore, the fraction of the targets was 0.25 on average as in the offline phase.

To decode the target at the end of the road each EFRP associated with the symbol attendance is assigned a probability to belong to the target class. At the end of the road segment, the probabilities of all EFRPs are aggregated symbol-wise by averaging. The symbol that has the highest probability of a target class is selected as a predicted target. Therefore, to allow equal accumulation on average for each symbol we increased the number of symbols to four. Each symbol appears between 2 and 4 times resulting in 12 boards on the road. The fraction of the targets is 0.25 in total.

3. Methods

Our analysis of EEG and behavioral data, as well as visual recognition decoding, has different objectives in the offline and online phases. The offline phase is necessary for collecting good quality data for training a classifier to be used in closed-loop during the online phase and to estimate its performance. However, it also allows us to study EEG correlates of visual recognition and perform a comparative study of different classifiers.

The online phase allows the estimation of the closed-loop BCI application performance. In the online phase, we apply a single decoding approach. In order to

make a comparative analysis of different decoding approaches under online conditions, we also performed a simulated online analysis.

3.1. Fixation extraction and analysis

There exist numerous methods to extract eye movement events from the eye gaze direction. We used a more robust and accurate method of Identification by 2-Means Clustering (I2MC) [31] for the offline phase data. In the online phase, and simulated online analysis, we used Identification by Dispersion-Threshold (IDT), a simpler and easy to implement for real-time processing [32]

Offline phase. We relied on the I2MC implementation provided by the authors of the method using the default parameters. The main idea behind this method is to find the transition between two consecutive fixations by applying 2-mean clustering in a sliding window manner. During a fixation, the eyes do not move so if we can clearly detect 2 clusters of gaze direction within the sliding window it means that they correspond to two fixations. This method is more precise and robust to noisy outliers which allows to obtain a training dataset of higher quality.

Online phase. Since the available implementation of I2MC method cannot be applied in real time to extract eye fixations, we used the Identification by Dispersion-Threshold (IDT) supplied with our eye tracking system. Fixation in IDT is extracted when the signals lie within the dispersion thresholds for at least a minimum fixation duration. It requires two parameters: we used 100 ms for the minimum fixation duration and 200 pixels for the maximum dispersion.

Simulated online phase. The same procedure for fixation extraction is applied as in the online phase.

Fixation analysis. The P300-like components of cognitive response are stronger when the stimulus is perceived and recognized for the first time [28]. We assume that subjects categorized the symbol at the first attendance so we use only the first fixations on the boards for our analysis. As mentioned in Section 2.4, the popping up effect is introduced and the boards are specially designed to ensure that recognition takes place near fixation on the symbol.

The visual input during the task is dynamic. Due to driving through the virtual environment the objects, including the boards, are also moving on the screen. So we assume that most of the board attendances are done with smooth pursuit rather than fixations. To the best of our knowledge, there is no available algorithm for efficient extraction of smooth pursuit for eye movement data sampled at 120 Hz. The only consequence of extracting fixation from smooth pursuit is that a single smooth pursuit may be oversegmented into multiple fixations. For the sake of our analysis, we do not

need to differentiate between fixations and smooth pursuit movements. The onset of the first fixation on a board will coincide with the onset of smooth pursuit.

Each detected fixation was assigned to a target board, a non-target board or non-board. Because (1) the visual span for text reading extends up to several degrees, (2) the boards were moving and (3) eye tracking data were noisy, we applied the following approach to assign the boards to fixations. For each eye gaze sample, we estimate the probability of fixating eyes on the center of the board according to a normal distribution. After averaging log-probabilities across the fixation time window we apply a hard threshold to assign the fixation to a board or a non-board category. This procedure was implemented in the analysis of the data. The real-time assignment of fixations on boards in the online phase was modified for computational efficiency. It was based on the first 100 ms after fixation onset and used on the average fixation direction within 100 ms.

For the behavioral analysis, we estimate the dwell time – the uninterrupted time which the driver spent looking at a board. The dwells were created by merging all the fixations on the same board with the inter-fixation interval between them below 50 ms. For the sake of valid comparison between offline and online phases in behavioral analysis, we used IDT method to extract fixations.

3.2. EEG processing

EEG processing of offline data. All the EEG and EOG were downsampled from 2 kHz to 256 Hz using FIR antialiasing lowpass filter and further filtered non-causally with a Butterworth band-pass filter of order 4 within the band [1, 10] Hz. Due to the low conductivity of the skull and the skin, the EEG signal is spatially smoothed so a high contrast between nearby channels is a result of noise and movement artifacts. We remove this noise by keeping only low spatial frequency components after decomposition EEG with SPHARA [33]. Horizontal and vertical components of eye movement were estimated which allowed to remove the eye movement artifacts from EEG using multiple regression as in [34] using only horizontal and vertical components as well as the intercept. The coefficients of multiple regression were estimated from the one-minute session of eye movements before the corresponding run. Then the signal is spatially filtered with common-average-reference (CAR).

EEG processing in the simulated online analysis. The EEG processing was identical to the offline analysis except for the eye movement artifacts removal. The multiple regression model was obtained from offline data.

EEG processing in the online analysis. The closed-loop interaction in the online phase required real-time processing. SMI system provides a real-time eye fixation detection. The fixations were buffered by a parallel process within the driving simulator, matched with the boards, and a trigger was sent to the BioSemi system 3 s after the onset of

each the fixation on a board. We chose 3 s delay because we applied a non-causal filter on EEG data and need to diminish the edge effect of the backward filter pass. EEG processing was identical to the offline procedure except for 2 steps:

- the spectral filtering was done on a 5 s buffer of data, approximately around [-2, 3] s around the fixation onset;
- the eye movement artifacts were removed based on the multiple regression coefficients trained with offline phase eye movement data.

3.3. Decoding approaches

To decode visual recognition from EEG signals we extracted the epochs around the fixation onset – EEG segment between 200 ms to 1000ms after the fixation onset. In addition, we extracted dwell time as eye-gaze based features to detect visual recognition (see Section 3.1). We investigate and compare different decoding approaches which consist of various feature sets and classifiers.

Offline and simulated online decoding. For the offline analysis we chose the following combination of features and classifiers:

- **PLR-Waveform**. Penalized logistic regression (PLR) trained on waveform features (i.e. the signal's value at each channel and time point) after reducing the dimensionality with PCA. Only the components which explain 90% of the variance are kept.
- **PLR-Dwell**. PLR trained on dwell time on the boards.
- **PLR-Combined**. PLR trained on the combination of waveform features with dwell time. We concatenate the dwell time and EEG waveform features after applying PCA to keep 95% of variance.
- **RF-Waveform**. Random forest trained on waveform features. We use 100 decision trees and with a maximum depth of 5.
- **PLR-Riemann**. PLR trained on Riemannian features from simple EEG epochs. To build Riemannian features we estimate spatial covariance matrix with shrinkage and project it to the tangent space according to the classical Riemannian geometry on SPD matrices [35]. We subselected 8 channels based on the mean Fisher score across the epoch.
- **PLR-Riemann+**. PLR trained on Riemannian features from augmented epochs. Before computing the covariance matrix we augment the epoch with the averaged ERP for each class (target and non-target). Otherwise, it is identical to the previous approach.

Since PLR is a linear regularized classifier we standardize all the features to z-score when using PLR. For PLR we applied elastic net regularization with a major contribution from ℓ^2-norm (1000 times greater than ℓ^1-norm).

Online decoding. The preliminary tests on two subjects showed that RF-Waveform classifier performed better than the PLR-Waveform classifier. We did not consider PLR-Riemann classifier at this stage because its training was too slow to apply between offline and online phases. Therefore, on the data obtained in the offline phase, we trained individual RF-Waveform classifiers and applied it in real time in the online phase. The probability for the target class was estimated on each instance of every symbol and sent back to the driving simulator. After crossing the finish line the probabilities were averaged per each symbol (1 out of 4). The symbol which had the highest probability of being a target was shown to the subject on the screen.

3.4. Performance evaluation

Offline performance evaluation. We employ nested cross-validation to adjust various hyperparameters in the inner loop: regularization term for PLR and the tree depth in Random Forest. The purpose of the outer loop is to obtain an unbiased performance estimation so it is critical to avoid training and testing on correlated data. We achieve it by performing leave-one-run-out for the outer loop, although we had only 3 offline runs. The inner loop is implemented with 4-fold cross-validation while keeping the temporal order of the trials before the split. Since the classes of target and non-target eye fixations are unbalanced, we utilized AUC to measure the classification performance.

Simulated online performance evaluation. After training the classifiers on all the offline data we applied them to all the online data and obtained a single AUC value.

Online performance evaluation. During the online phase, we predicted the target symbol from the EFRP classification. We assess the overall performance with accuracy and confusion matrices for 4 symbols. Accuracies were estimated per subject and averaged across subjects. The significance of single-subject performance was tested with the exact binomial test against 0.25 (the random level with 4 balanced classes). Due to the high number of tests, we applied the Benjamini-Hochberg correction [36]. The average performance was tested with Student's t-test. The confusion matrices were tested for independence with Pearson's chi-square test [37] separately for each subject with Benjamini-Hochberg correction.

4. Results

4.1. Behavioral analysis

Comparison of fixation extraction methods I2MC vs IDT We used two different methods to extract fixations in the offline and online phase of the experiment. We compared them using the offline phase data via two measures: 1) the average number of all extracted dwells on boards per run and 2) the average dwell time (Table 2). The number of dwells obtained with IDT method is 10% smaller (260 vs 284) which

Table 2. The average number of dwells on boards and the average dwell time per run based on fixation extraction with I2MC and IDT methods in the offline phase. The numbers represent the average and standard deviation across subjects.

	# of dwells	Dwell time [ms]
I2MC	284 ± 59	408 ± 54
IDT	260 ± 64	433 ± 86

Table 3. The board attendance rate for targets and non-targets in the offline and online phases with IDT method of fixation extraction. The numbers represent the average and standard deviation across subjects.

	Offline	Online
Targets	0.92 ± 0.07	0.47 ± 0.06
Non-Targets	0.92 ± 0.07	0.44 ± 0.09

is a statistically significant difference (p-values < 0.0001 with one-tailed Student's t-test). Since we used IDT method in the online phase, we missed 10% of the fixations which led to either missed attendance or labeling some of the second fixations as first fixations. The dwell duration is slightly longer with IDT method (433 vs 408 ms) and it is not significantly different compared to I2MC method (p-values $= 0.06$ with one-tailed Student's t-test).

Board attendance One of the subjects had a poor quality of eye tracker calibration which led to poor fixation extraction. This subject is excluded from the behavioral analysis. For the analysis of attendance rate, we used IDT method to extract fixations for the data from both offline and online phases. Subjects attended most of the boards in the offline phase and approximately half of them in the online phase. The average attendance rate is shown in Table 3. Repeated measures ANOVA shows a significant difference between all 4 groups: targets in offline, targets in online, non-targets in offline and non-targets in online, with p-value < 0.0001. The post hoc analysis shows that it is driven by the difference between offline and online phases with p-value < 0.0001 (paired t-test). The difference between targets and non-targets is not significant with p-value $= 0.048$ so we can assume that subjects could not differentiate the symbols with peripheral vision.

Counting The total number of targets in the offline phase is 173. We analyzed the button presses after each road segment which should be equal to the number of targets on each road. The average number of incorrect counts (both missed and extra counts) was 5, the worst performance was at 15 errors (see the number of mis-counts in Figure 6).

Dwell time We analyzed the distribution of dwell times on targets vs non-targets in the offline and online phases (Figure 3). Most of the dwells are limited to the time between

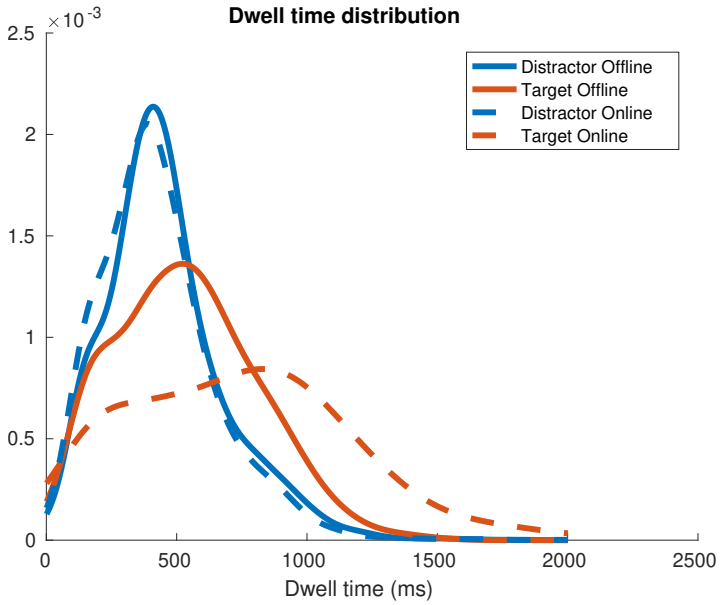

Figure 3. Dwell time distribution for first dwells on targets vs distractors in the offline and online phases. The single-subject distributions were smoothed and averaged across subjects.

the boards pop up equal to 900 ms. The dwell times are identical for non-targets in both phases and significantly shorter than for targets (p-value < 0.0001). The median dwell time for targets is significantly longer in the online phase (p-value < 0.0001).

4.2. EFRP waveform

We present the analysis of the aggregated EFRP waveform for the 4 subjects who demonstrated the highest classification performance in the offline phase with the PLR-Waveform classifier because it allows for direct representation of the features used by a linear classifier (see Section 4.3).

We visualized the representative Cz channel signal for each eye fixation while ordering them by the dwell time (Figure 4). The amplitude of the presented EFRP is limited to the range [-1.5, 1.5] uV. The complex of components right after the fixation onset ranging from 100 to 300 ms reflects the evoked potentials from the fixation itself. It contains negative and positive deflections. We can observe the same complex of components after the dwell offset. The shift of gaze happens right where we expect the P300-like component so it can be masked by this evoked activity. The positive deflection occurs at the end of the dwell for both targets and non-targets, however, it has a greater amplitude for targets as seen on the averaged EFRP.

The univariate discriminant power is shown in Figure 5. The results are similar for the offline and online phases, with online data having twice as higher discriminant power with up to 0.01 of signed R^2. The greatest values are mainly confined within the region between 100 and 700 ms. The higher discriminant power is spread across the whole scalp which can be the consequence of using CAR in the processing. Nonetheless,

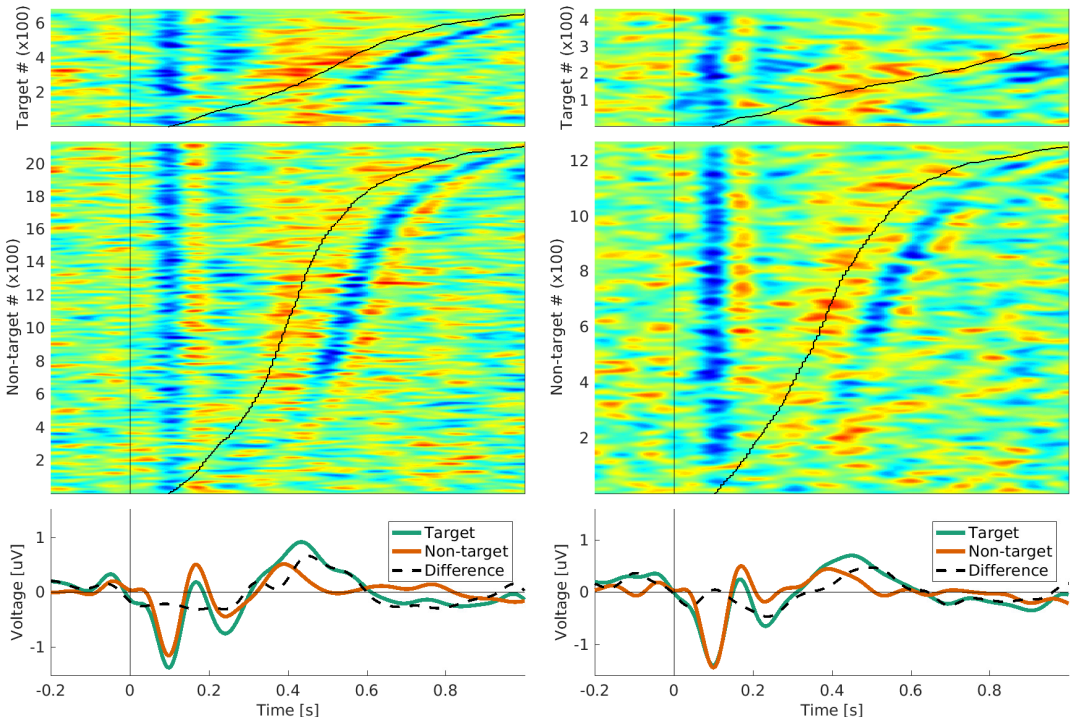

Figure 4. The signal of Cz channel for extracted epochs aggregated for the 4 best performing subjects with the PLR-Waveform approach. Left: offline, Right: online. The epochs are ordered according to the dwell time shown with the black curve.

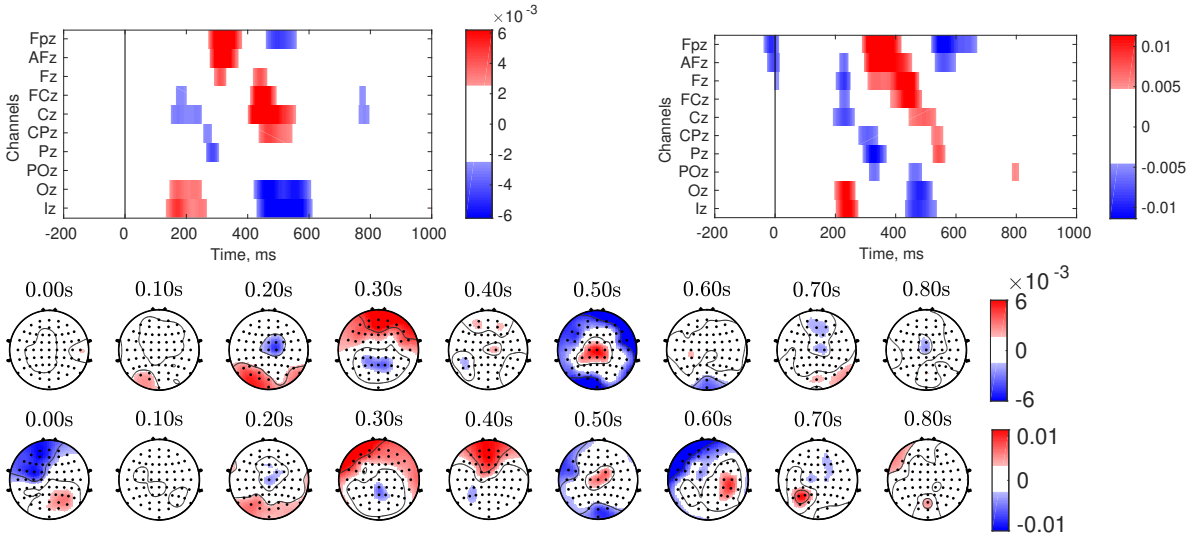

Figure 5. The discriminant power for the aggregated epochs of 4 best subjects with the offline AUC above 0.6 with the PLR-Waveform approach. Signed R^2 is demonstrated for midline channels across around the eye fixation onset (top left: offline, top right: online) and on topographic maps (top map: offline, bottom map: online).

the P300-like component of EFRP can be seen at 500 ms after the fixation onset.

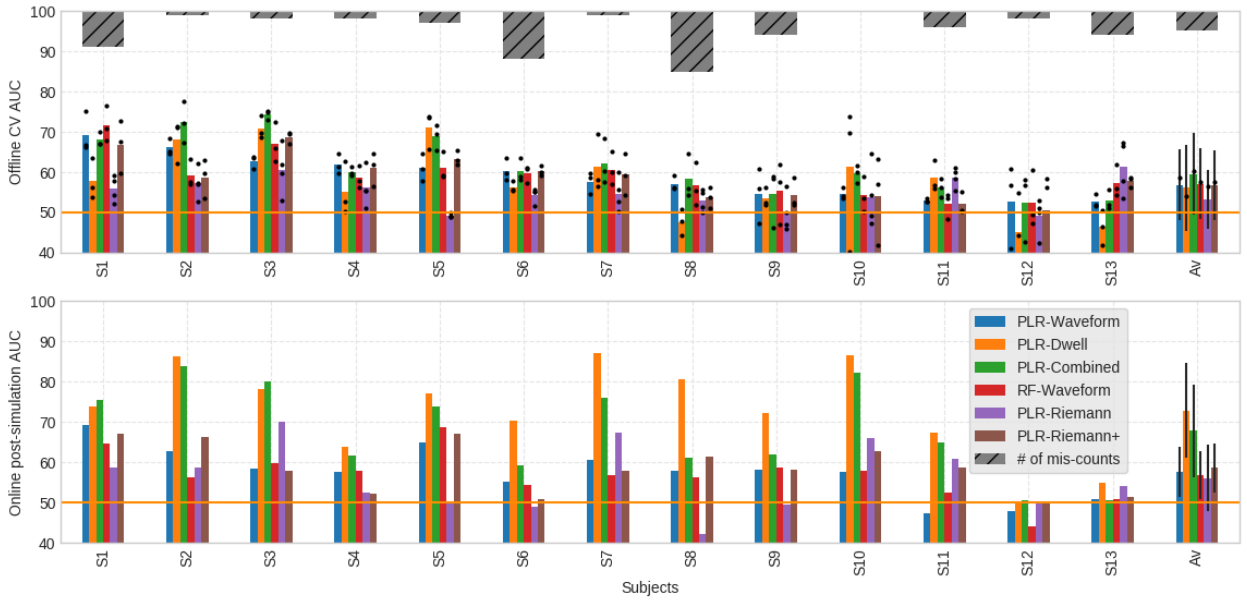

Figure 6. Performance of EFRP classification with various approaches in the offline analysis (top) and the simulated online analysis (bottom) as a percentage of AUC. Each dot shows single fold performance in a leave-one-run-out cross-validation for the corresponding classification approach. The overlaid gray bars at the top show the behavioral performance: the number of mis-counts in the offline phase on the same scale as AUC percentage.

4.3. Comparison of decoding approaches

Offline All the classification methods yielded a single-trial performance between 0.53 and 0.6 AUC on average (Figure 6) which is statistically significant against the random level of 0.5 for all methods (p-values < 0.05 with Student's t-test after Benjamini-Hochberg correction for 6 methods). Eight out of thirteen subjects achieve performance above 0.6 for at least one of the approaches based on EEG features. The performance of different approaches has a different ranking per subject, however, the differences between the approaches are not statistically significant (p-value $= 0.16$ with repeated measures ANOVA). It is worth noticing that the combination of both dwell time and EEG waveform is not always better than just one of these feature sets.

Simulated online It is worth noting that the performance of EEG-based approaches on online data is consistent with the training performance on offline data. The average AUC values lie between 0.56 and 0.59 for each approach which is statistically significant against 0.5 for all approaches. (p-values < 0.05 with Student's t-test after Benjamini-Hochberg correction for 6 approaches).

For approaches relying on the dwell time, however, the performance drastically improved for multiple subjects compared to the offline performance. The average AUC

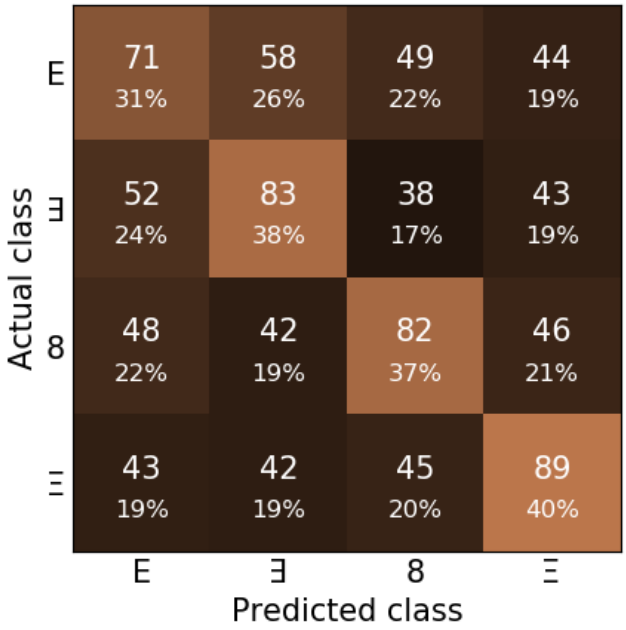

Figure 7. The aggregated confusion matrix of online decoding for each type of symbol being a target class. The absolute number of cases is presented and normalized as a percentage relative to the actual class (row-wise normalization).

Table 4. Online performance. The test results show the p-values after Benjamini-Hochberg correction. Statistically significant results are typed in bold.

Subjects	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Accuracy	0.44	0.38	0.45	0.29	0.44	0.35	0.39	0.25	0.31	0.38	0.38	0.4
Accuracy test	**0.001**	**0.04**	**0.001**	0.48	**0.001**	0.07	**0.02**	1.0	0.24	**0.04**	**0.04**	**0.03**
Independence test	**0.004**	0.29	**0.03**	0.74	0.08	0.17	0.1	0.41	0.48	0.19	0.22	0.26

for *dwell* classifier increased from 56 to 73 and for the *PLR-Combined* classifier from 59 to 67. This improvement is a direct consequence of the increased target dwell time together while non-target dwell time remained unchanged, see Figure 3. However, the underlying question of why the dwell time changed between the offline and the online phases still remains.

4.4. Online performance

We assessed the task performance for each subject as the accuracy of decoding target at the end of the road integrated after multiple presentations of each type of boards (Table 4). The averaged accuracy equals 0.37 and it is significantly different from the random level of 0.25 for a multi-class classification with 4 balanced classes (p-value < 0.0001 with Student's t-test). Additionally, we applied a statistical test to assess the accuracy per subject. After adjusting p-values with Benjamini-Hochberg procedure 8 subjects performed statistically above chance level.

To verify the independence of the 4 classes in the online phase we computed the aggregated confusion matrices across all symbols for all subjects (Figure 7). We applied an independence test for confusion matrices per subject. After adjusting p-values with Benjamini-Hochberg procedure only two subjects show a significant imbalance in the confusion matrix. It is linked to the high accuracy of the symbol Ξ and low accuracy for the symbol E.

5. Discussion

The integration of the BCI systems in the daily life of healthy/able-bodied users requires the system to be built around the experimental paradigms supporting human natural behavior. To this end, EFRP-based decoding of cognitive processes in overt visual search has the potential to augment human-machine-interaction. In this study, we investigated the decoding of visual recognition in a driving scenario. It resembles one of the typical everyday activity and provides the associated challenges in decoding visual recognition: free eye gazes, dynamic visual input, primary tasks. For this purpose, we limit the driving task to following the simple route at a comfortable and natural speed. To avoid overloading subjects' attention with too many distractors, we did not include other participants on the road nor moving objects. Nonetheless, due to the movement of the car, the drivers were subjected to a dynamic visual input and perceived a natural optic flow.

Ocular behavior in driving. We considered the random pop-up appearance of the task-relevant stimuli above the sidewalks. In the previous study on active search in a dynamic scene [27], this type of appearance created sufficient locking of the cognitive EFRP's components to the fixation onset. At the same time, their results showed that the time spent on the stimuli was not informative about the stimuli type (target vs. non-target) in contrast to the more attentionally demanding motion appearance conditions. Interestingly, we observed the time spent on the stimuli to be discriminative in our exp. scenario although the same pop-up stimuli appearance was considered. This illustrates the effect of the user/driver motion relative to the visual content on the visual information processing and the attentional load. Moreover, in the online phase subjects looked longer on targets compared to the offline phase. Although the subjects' dwell time was not decoded directly, they were aware that they could potentially influence the decoding quality. This might lead to deliberate or unconscious changes in their behavior. Some subjects could achieve a high decoding performance based only on dwell time in the offline phase. But with the changes in the behavior during the online phase, most of the subjects drastically improve in their decoding performance based on the dwell time.

The board attendance rate reveals two aspects of the task completion: (1) how well the subjects coped with the pace and attentional load and (2) whether they can differentiate between targets and distractors using their peripheral vision. We observed a statistically significant difference in the board attendance between offline and online

phases despite the equal number of boards. On one hand, we expected the subjects to be more engaged in the task due to the interactive feedback part. However, the observed increase of the dwell time on target boards in the online phase made it more challenging to attend all the boards within the limited time. On the other hand, behavioral response based on the number of recognized targets was not required which might relax the cognitive load.

We carefully designed stimuli to ensure that their recognition requires foveal vision. It is confirmed by the similar attendance rate on targets and non-targets and by just a few errors in the counting task. We were concerned that the introduction of new symbols in the online phase may change the behavior or the cognitive response due to the novelty. The balanced confusion matrices of decoding target recognition in the online phase confirm that subjects perceived all the symbols equally with regards to the task. However, the presence of additional types of stimuli may have contributed to the increased dwell time on the target boards because the target identification among 4 symbols closely resembling each other is more challenging than among only 2 symbols [38, 39].

Decoding visual recognition in driving. The discriminant analysis of EFRP shows similar results for both offline and online phases. Most of the relevant features lie within [200, 700] ms window which coincides with the dwell durations. The spatial localization of relevant features is consistent with the typical spatial distribution of P300 component in the oddball paradigm. The EFRP waveforms are known to contain a strong P1 component at the occipital area that reflects the beginning of the visual processing of a stable visual input after the saccade. In the analysis of Cz channel, it corresponds to the negative deflection at 100 ms. It is clearly present in most of the first fixations on boards. Moreover, we could also see it for the following fixations. During the dwell time, multiple fixations can be detected what leads to the overlap of the P300-like component locked to the 1st fixation and the following early EFRP component. In contrast to the classical P300 paradigms where the timing of the visual input is controlled (regular timing), here the overlap of EFRP is variable across trials. This is an additional challenge as compared to typical event-related potential detection approaches. Removing the activity related to previous and subsequent fixations from the EEG was attempted by modeling it from various characteristics of the previous and subsequent fixations. [28, 40, 41]. We did not choose to apply this approach as there is a risk to distort the signal of interest. The precise EFRP shape depends on preceding and following eye behavior (e.g. amplitude and direction of the saccades) as well as low-level features of the visual input [42].

We compared multiple classification approaches based on EFRP on offline data and simulated their application to online data. All approaches, including waveform-based linear and non-linear and covariance-based methods, resulted in a similar performance on average across subjects which is significantly above the chance level for all the approaches. However, there is no single best approach for all subjects.

The actual online performance in closed-loop measured by the target symbol identification is significantly above the chance level for 8 subjects. Nonetheless, on average across subjects, the accuracy of 0.37 is significantly higher than 0.25.

One can argue that the relatively good performance of EFRP-based classifiers is due to the strong and well-aligned evoked potentials after the fixations offset, which reflects the difference between the target and non-targets dwell times. However, this was not the case as the simulated online performance of the EEG-based decoders did not exhibit the improvement achieved by the classification models based on dwell time. Also, a combination of the two sources of information (EEG and dwell time) did not lead to significant improvement as compared to EEG alone. Future work will require incorporating techniques in the BCI decoders to cope with the natural temporal variability of EEG potentials [43].

Finally, in this study, our primary goal was to explore the neural correlates of the object recognition. Ocular correlates such as pupil dilation along with multi-modal integration of neural and ocular correlates may be further investigated in future studies. Multimodal integration may improve the decoding performance [15].

6. Conclusion

In this study, we have demonstrated the feasibility to decode recognition of visual targets from brain signals in challenging conditions of a naturalistic driving scenario. Our approach and results, although far from optimal, would enhance interaction by enabling intelligent cars provide valuable recommendations to drivers, who could quickly accept or ignore. Still, our approach can benefit from improvements in different components, such as decoders robust to the natural temporal variability of EEG potentials, reliable and computationally efficient eye fixation algorithms, longer training of subjects over multiple sessions, and multimodal integration of neural and ocular data.

Acknowledgments

The authors thank Nissan Motor Co. Ltd for supporting this work.

References

[1] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398(6725):297–298, March 1999.

[2] J. R. Wolpaw and D. J. McFarland. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51):17849–17854, December 2004.

[3] E. M. Holz, J. Höhne, P. Staiger-Sälzer, M. Tangermann, and A. Kübler. Brain–computer interface controlled gaming: Evaluation of usability by severely motor restricted end-users. *Artificial Intelligence in Medicine*, 59(2):111–120, October 2013.

[4] R. Leeb, L. Tonin, M. Rohm, L. Desideri, T. Carlson, and J. d. R. Millán. Towards Independence: A BCI Telepresence Robot for People With Severe Motor Disabilities. *Proceedings of the IEEE*, 103(6):969–982, June 2015.

[5] S. Saeedi, R. Chavarriaga, and J. d. R. Millán. Long-Term Stable Control of Motor-Imagery BCI by a Locked-In User Through Adaptive Assistance. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(4):380–391, April 2017.

[6] S. Perdikis, L. Tonin, S. Saeedi, C. Schneider, and J. d. R. Millán. The Cybathlon BCI race: Successful longitudinal mutual learning with two tetraplegic users. *PLOS Biology*, 16(5):e2003787, May 2018.

[7] A. Ramos-Murguialday, D. Broetz, M. Rea, L. Läer, Ö. Yilmaz, F. L. Brasil, G. Liberati, M. R. Curado, E. Garcia-Cossio, A. Vyziotis, W. Cho, M. Agostini, E. Soares, S. Soekadar, A. Caria, L. G. Cohen, and N. Birbaumer. Brain-Machine-Interface in Chronic Stroke Rehabilitation: A Controlled Study. *Annals of Neurology*, 74(1):100–108, July 2013.

[8] F. Pichiorri, G. Morone, M. Petti, J. Toppi, I. Pisotta, M. Molinari, S. Paolucci, M. Inghilleri, L. Astolfi, F. Cincotti, and D. Mattia. Brain–computer interface boosts motor imagery practice during stroke recovery. *Annals of Neurology*, 77(5):851–865, 2015.

[9] A. Biasiucci, R. Leeb, I. Iturrate, S. Perdikis, A. Al-Khodairy, T. Corbet, A. Schnider, T. Schmidlin, H. Zhang, M. Bassolino, D. Viceic, P. Vuadens, A. G. Guggisberg, and J. d. R. Millán. Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke. *Nature Communications*, 9(1):1–13, June 2018.

[10] M. A. Cervera, S. R. Soekadar, J. Ushiba, J. d. R. Millán, M. Liu, N. Birbaumer, and G. Garipelli. Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis. *Annals of Clinical and Translational Neurology*, 5(5):651–663, March 2018.

[11] H. Zhang, R. Chavarriaga, Z. Khaliliardali, L. Gheorghe, I. Iturrate, and J. d. R. Millán. EEG-based decoding of error-related brain activity in a real-world driving task. *Journal of Neural Engineering*, 12(6):066028, 2015.

[12] Z. Khaliliardali, R. Chavarriaga, L. A. Gheorghe, and J. d. R. Millán. Action prediction based on anticipatory brain potentials during simulated driving. *Journal of Neural Engineering*, 12(6):066006, 2015.

[13] R. Chavarriaga, M. Uščumlić, H. Zhang, Z. Khaliliardali, R. Aydarkhanov, S. Saeedi, L. Gheorghe, and J. d. R. Millán. Decoding Neural Correlates of Cognitive States to Enhance Driving Experience. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2(4):288–297, August 2018.

[14] M. Uščumlić, R. Chavarriaga, and J. d. R. Millán. An Iterative Framework for EEG-based Image Search: Robust Retrieval with Weak Classifiers. *PLOS ONE*, 8(8):e72018, August 2013.

[15] D. C. Jangraw, J. Wang, B. J. Lance, S.-F. Chang, and P. Sajda. Neurally and ocularly informed graph-based models for searching 3D environments. *Journal of Neural Engineering*, 11(4):046003, 2014.

[16] S. Haufe, J.-W. Kim, I.-H. Kim, A. Sonnleitner, M. Schrauf, G. Curio, and B. Blankertz. Electrophysiology-based detection of emergency braking intention in real-world driving. *Journal of Neural Engineering*, 11(5):056011, August 2014.

[17] A. D. Gerson, L. C. Parra, and P. Sajda. Cortically coupled computer vision for rapid image search. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):174–179, June 2006.

[18] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw. Toward enhanced P300 speller performance. *Journal of Neuroscience Methods*, 167(1):15–21, January 2008.

[19] D. Rosenthal, P. DeGuzman, L. C. Parra, and P. Sajda. Evoked Neural Responses to Events in Video. *IEEE Journal of Selected Topics in Signal Processing*, 8(3):358–365, June 2014.

[20] J. Polich. Updating P300: An Integrative Theory of P3a and P3b. *Clinical Neurophysiology*, 118(10):2128–2148, October 2007.

[21] A. Rezeika, M. Benda, P. Stawicki, F. Gembler, A. Saboor, and I. Volosyak. Brain–computer interface spellers: A review. *Brain Sciences*, 8(4), March 2018.

[22] E. W. Sellers, T. M. Vaughan, and J. R. Wolpaw. A brain-computer interface for long-term independent home use. *Amyotrophic Lateral Sclerosis*, 11(5):449–455, October 2010.

[23] E. M. Holz, L. Botrel, T. Kaufmann, and A. Kübler. Long-Term Independent Brain-Computer Interface Home Use Improves Quality of Life of a Patient in the Locked-In State: A Case Study. *Archives of Physical Medicine and Rehabilitation*, 96(3, Supplement):S16–S26, March 2015.

[24] J. R. Wolpaw, R. S. Bedlack, D. J. Reda, R. J. Ringer, P. G. Banks, T. M. Vaughan, S. M. Heckman, L. M. McCane, C. S. Carmack, S. Winden, D. J. McFarland, E. W. Sellers, H. Shi, T. Paine, D. S. Higgins, A. C. Lo, H. S. Patwa, K. J. Hill, G. D. Huang, and R. L. Ruff. Independent home use of a brain-computer interface by people with amyotrophic lateral sclerosis. *Neurology*, 91(3):e258–e267, July 2018.

[25] C. Evans. Some further observations on occipital sharp waves (lambda waves). *Electroencephalography and Clinical Neuro- physiology*, (4):371, 1952.

[26] A.-M. Brouwer, B. Reuderink, J. Vincent, M. A. J. v. Gerven, and J. B. F. v. Erp. Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials. *Journal of Vision*, 13(3):17–17, September 2013.

[27] M. Uščumlić and B. Blankertz. Active visual search in non-stationary scenes: coping with temporal variability and uncertainty. *Journal of Neural Engineering*, 13(1):016015, 2016.

[28] H. Devillez, E. Kristensen, N. Guyader, B. Rivet, and A. Guérin-Dugué. The P300 potential for fixations onto target object when exploring natural scenes during a visual task after denoising overlapped EFRP. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 1024–1027, April 2015.

[29] H. Renold, R. Chavarriaga, L. A. Gheorghe, and J. d. R. Millán. EEG correlates of active visual search during simulated driving: An exploratory study. In *2014 IEEE International Conference on Systems, Man, and Cybernetics*, 2014.

[30] J. E. Kamienkowski, M. J. Ison, R. Q. Quiroga, and M. Sigman. Fixation-related potentials in visual search: A combined EEG and eye tracking study. *Journal of Vision*, 12(7):4–4, July 2012.

[31] R. S. Hessels, D. C. Niehorster, C. Kemner, and I. T. C. Hooge. Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC). *Behavior Research Methods*, 49(5):1802–1823, 2017.

[32] D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, pages 71–78, November 2000.

[33] U. Graichen, R. Eichardt, P. Fiedler, D. Strohmeier, F. Zanow, and J. Haueisen. SPHARA - A Generalized Spatial Fourier Analysis for Multi-Sensor Systems with Non-Uniformly Arranged Sensors: Application to EEG. *PLOS ONE*, 10(4):e0121741, April 2015.

[34] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller. A fully automated correction method of EOG artifacts in EEG recordings. *Clinical Neurophysiology*, 118(1):98–104, January 2007.

[35] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten. Multiclass brain-computer interface classification by Riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, March 2012.

[36] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[37] K. P. F.R.S. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, July 1900.

[38] A. O. Dick. Processing time for naming and categorization of letters and numbers. *Perception & Psychophysics*, 9(3):350–352, May 1971.

[39] J. Alegria and P. Bertelson. Time uncertainty, number of alternatives and particular signal-response pair as determinants of choice reaction time. *Acta Psychologica*, 33:36–44, January 1970.

[40] H. Devillez, N. Guyader, and A. Guérin-Dugué. An eye fixation–related potentials analysis of the P300 potential for fixations onto a target object when exploring natural scenes. *Journal of Vision*, 15(13):20, September 2015.

[41] E. Kristensen, A. Guerin-Dugué, and B. Rivet. Comparison between adjar and xDawn algorithms to estimate Eye-Fixation Related Potentials distorted by overlapping. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 976–979, April 2015. ISSN: 1948-3554.

[42] A. R. Nikolaev, R. N. Meghanathan, and C. van Leeuwen. Combining EEG and eye movement recording in free viewing: Pitfalls and possibilities. *Brain and Cognition*, 107:55–83, August 2016.

[43] R. Aydarkhanov, M. Uscumlic, R. Chavarriaga, L. Gheorghe, and J. d. R. Millan. Spatial covariance improves BCI performance for late ERPs components with high temporal variability. *Journal of Neural Engineering*, 2020.