PAPER • OPEN ACCESS

Integrating neural and ocular attention reorienting signals in virtual reality

View the article online for updates and enhancements.

Journal of Neural Engineering

PAPER

Integrating neural and ocular attention reorienting signals in virtual reality

Pawan Lapborisuth[1,*], Sharath Koorathota[1] ⓘ, Qi Wang[1] ⓘ and Paul Sajda[1,2,3,4,*] ⓘ

1 Department of Biomedical Engineering, Columbia University, New York, NY, United States of America
2 Department of Radiology , Columbia University Irving Medical Center, New York, NY 10032, United States of America
3 Department of Electrical Engineering , Columbia University, New York, NY 10027, United States of America
4 Data Science Institute, Columbia University , New York, NY 10027, United States of America
* Authors to whom any correspondence should be addressed.

E-mail: pl2622@columbia.edu and psajda@columbia.edu

Keywords: attention, eyetracking, EEG, BCI

Supplementary material for this article is available online

Abstract

Objective. Reorienting is central to how humans direct attention to different stimuli in their environment. Previous studies typically employ well-controlled paradigms with limited eye and head movements to study the neural and physiological processes underlying attention reorienting. Here, we aim to better understand the relationship between gaze and attention reorienting using a naturalistic virtual reality (VR)-based target detection paradigm. *Approach.* Subjects were navigated through a city and instructed to count the number of targets that appeared on the street. Subjects performed the task in a fixed condition with no head movement and in a free condition where head movements were allowed. Electroencephalography (EEG), gaze and pupil data were collected. To investigate how neural and physiological reorienting signals are distributed across different gaze events, we used hierarchical discriminant component analysis (HDCA) to identify EEG and pupil-based discriminating components. Mixed-effects general linear models (GLM) were used to determine the correlation between these discriminating components and the different gaze events time. HDCA was also used to combine EEG, pupil and dwell time signals to classify reorienting events. *Main results.* In both EEG and pupil, dwell time contributes most significantly to the reorienting signals. However, when dwell times were orthogonalized against other gaze events, the distributions of the reorienting signals were different across the two modalities, with EEG reorienting signals leading that of the pupil reorienting signals. We also found that the hybrid classifier that integrates EEG, pupil and dwell time features detects the reorienting signals in both the fixed (AUC = 0.79) and the free (AUC = 0.77) condition. *Significance.* We show that the neural and ocular reorienting signals are distributed differently across gaze events when a subject is immersed in VR, but nevertheless can be captured and integrated to classify target vs. distractor objects to which the human subject orients.

1. Introduction

As humans, we constantly redirect our attention to different objects and stimuli in the environment. The complex set of neural and physiological adjustments we make is known as the reorienting response. The process underlying attention reorienting (e.g. the reorienting response) has been widely studied both in the fields of neuroscience and psychology [1–3]. Previous studies have identified neural and physiological signatures of attention reorienting, including pupil dilation and the P300 wave recorded via electroencephalography (EEG) [2, 4, 5]. These neural and physiological signatures are parts of the larger attention networks in the brain, namely the dorsal and ventral attention networks, which have also been functionally linked to the locus coeruleus-norepinephrine (LC-NE) system [1, 6, 7]. While

the relationship between the P300 signal and pupil dilation remains unclear, both of them have been shown to potentially reflect the phasic activity of the LC nucleus, with the P300 reflecting the cortical signatures of attention reorienting and pupil dilation serving as an index of the subcortical LC-NE system activity [1, 8, 9]. Utilizing these neural and physiological signatures, recent neural engineering studies have developed brain computer interfaces (BCIs) that can perform simple tasks based on the user's attention reorienting response, such as a P300-based speller and computer cursor control [10, 11].

One of the major limitations of prior attention reorienting studies is the unnaturalistic environment in which the subjects performed tasks. These studies typically employed different variations of a cueing or an oddball task presented on a 2D screen to generate the reorienting response [12–14]. While these tasks are simple, well-documented and well-controlled, they do not represent how humans actually reorient their attention in the real world. Take a simple example of a person driving a vehicle down the street. The driver must constantly reorient their attention to different objects and events in the environment as the vehicle moves forward. These objects may be task relevant such as a pedestrian crossing the street or task irrelevant such as an on-ramp sign. At the same time, the real-world field of view is much wider than that of a screen, requiring the person to not only move their eyes but also their head to constantly monitor the surrounding environment. To better understand the neural and physiological basis of attention reorienting in real-world scenarios, a more naturalistic experimental paradigm is needed. This understanding would potentially translate to more robust and reliable attention-based BCI systems that are not confined to a 2D screen and instead enable more natural eye and head movements.

In this study, we employ an immersive 3D-based target detection paradigm presented in a head-mounted virtual reality (VR) display to study attention reorienting signals in a naturalistic and dynamic setting. Subjects travel through a simulated city environment in a moving vehicle with blank white billboards located in between buildings on the left- and right-hand side of the street. They are instructed to count the number of target images that appear on the billboards during each experimental run. Subjects perform the target detection task under two conditions, one without head movement as a control condition and one with head movement as a more naturalistic condition. We simultaneously collect the subjects' EEG, pupil diameter, gaze position and head rotation data. Our aims are twofold. First, we aim to better understand the relationship between eye movements and the reorienting response. In previous reorienting studies, the traditional experimental paradigm typically only allows for minimal or well-controlled eye movements. However, in more naturalistic conditions such as the one in the current study, eye and head movements of the subjects are now coupled to the reorienting process. This effectively decomposes the reorienting process across these movements. Therefore, we aim to investigate how the neural and ocular reorienting signals are reflected in this decomposition. To achieve this goal, we first employ temporal-based EEG-only and pupil diameter-only classifiers to identify the neural and ocular reorienting signatures that differentiate between target and distractor stimuli responses. We then perform general linear model (GLM) analysis to determine the correlation between the length of different gaze events and the reorienting signatures derived from the classifiers. We show that while the dwell time contributes the most to the reorienting response, the distributions are different between the two modalities, with the EEG reorienting response leading that of the pupil reorienting response. Second, we aim to capture and integrate the neural and physiological response underlying attention reorienting in a naturalistic environment. We employ a hierarchical hybrid classifier combining EEG, pupil diameter and dwell time to classify the object in which the subject observes during each trial. We show that the hybrid classifier successfully captures neural and ocular reorienting signals and can classify the target object with relatively high accuracy even when the subject moves their head in a naturalistic environment.

2. Methods

2.1. Subjects
Twenty healthy volunteer subjects (15 male, 5 female, aged 18–40 years old) were recruited for this study. Subjects did not report any neurological illness or medication and all had normal or corrected to normal vision. Informed consent was obtained in writing from all subjects prior to the experiment in accordance with the guidelines and approval of Columbia University Institutional Review Board. Data from two subjects (1 male, 1 female) were excluded from the final analysis due to substantial artifacts in the EEG signals. Data from the eighteen remaining subjects (14 male, 4 female, aged 18–40 years old) were included in the final analysis.

2.2. Virtual environment
The 3D virtual target detection paradigm was developed using the open-source suite Naturalistic Experimental Design Environment [15] which is built on the Unity3D game development software (Unity Technologies, CA). The virtual environment consists of a street in the middle of a simulated city environment. Buildings were placed on the left- and right-hand side of the street, with blank white billboards placed in between the buildings. Images chosen from the CalTech101 database [16] appeared

on the billboards as the subject approached them in the virtual environment. Four categories of images were selected—cameras, laptops, grand pianos and schooners. Each category of images consisted of a total of 50 images. The image which appeared on each billboard was chosen at random and with random placement to the left or right of the street for each trial.

2.3. Experimental paradigm

During each experimental run, the subjects were navigated down the street at a constant speed in an autonomous vehicle. As the subjects approached to each pair of billboards, one of them would display an image chosen at random from the four categories described in the earlier section. Prior to the start of the experiment, the subjects were informed which one of the categories of images was a 'target' image and that the rest were 'distractor' images. The subjects were instructed to internally count the number of target images displayed and to report the final number to the experimenter at the end of each session.

Each subject performed the task under two conditions—fixed and free (figure 1(a)). In the fixed condition, the subjects were instructed to keep their head still throughout the whole experimental session while only using their eyes to saccade to the images displayed on the billboards before returning to center marked by a grey square in the middle of the street (figure 1(b)). In the free condition, the subjects were instructed to turn both their head and their eyes to observe and categorize the images on the billboards before returning to center, similarly marked by a grey square (figure 1(c)). The two conditions were designed to simulate a control condition (the fixed condition) where only eye movements were allowed and a more naturalistic condition (the free condition) where both eye and head movements were allowed.

A total of 40 images were displayed during each experimental block and each block lasted approximately 200 s. Each subject performed four experimental blocks at a time of a single condition and a total of 16 experimental blocks, eight being the fixed condition and eight being the free condition. The order in which the subjects performed each four experimental blocks were chosen at random. A total of 640 images were displayed for each subject and approximately 25% were targets. The target category was randomly selected for each subject.

2.4. Data acquisition

EEG data was collected using a Biosemi ActiveTwo amplifier (Biosemi, Amsterdam, The Netherlands) with 64 Ag/AgCl electrodes at a sampling rate of 2048 Hz. The electrodes were placed according to the international 10–20 system. All electrode impedances were less than 50 kΩ and common average reference was used. Eyetracking data was collected using a built-in Tobii eyetracker (Tobii, Stockholm, Sweden)

within the Tobii Pro headset. The eyetracker was used to collect eye position and pupil diameter data at a sampling rate of 120 Hz. A five-point calibration was performed every time the subject put on the headset prior to the start of the experiment. Re-calibration was performed if the calibration did not display an 'OK' sign at the end of the calibration session. An open-source software library known as lab streaming layer (LSL) was used to synchronize all the data streams together across a local network [17]. All data acquisition was performed in an electromagnetically shielded room.

2.5. Data pre-processing

Eye position and pupillometry data were analyzed using MATLAB (The Mathworks Inc., MA). Eye position data was first epoched from 0 to 3000 ms locking to image onset (IO). In order to study the relationship between gaze events and the reorienting signals, we first divided the continuous gaze data of each trial into distinct gaze events related to visual attention reorienting. For this purpose, we chose to apply piece-wise linear modeling to divide the continuous eye position data into four distinct phases: (1) Peripheral: the time of fixation on the center of the display before any gaze movement was made, (2) First saccade (FS): the duration of FS towards the target or distractor image, (3) Dwell time: the time of fixation on the target or distractor image and (4) Return saccade (RS): the duration of RS towards the center of the display (figures 1(b) and (c)). The trials that did not fit the model were discarded (about 15 percent of total number of trials on average per subject), along with the corresponding pupillometry and EEG trials.

Traditionally, EEG and pupillometry data are epoched by time-locking to the time of stimuli onset. However, as a result of our piece-wise linear modeling of the gaze data, we also identified the time in which the saccades and fixations began and ended for each trial. This allows us to epoch our EEG and pupillometry data based not only on when the stimuli onset occurs but also when the saccade towards the image and when the fixation on the image occur for each trial. We therefore denote these times as our three different 'locking conditions': (1) time of IO, (2) time of FS and (3) time of first fixation (FF).

Pupillometry data was first processed by removing any data during intervals in which the pupil was not detected. Blinks were then removed based on the speed of change of the pupil diameter. Any missing data was interpolated using cubic spline interpolation. Each subject's pupillometry data was then downsampled to 20 Hz and standardized for each experimental run. Pupillometry data was then epoched from 0 to 3000 ms based on locking condition and baseline-corrected using the mean value from −200 to −0 ms.

EEG data was pre-processed using EEGLAB toolbox [18]. The 64 channel EEG data were band-pass
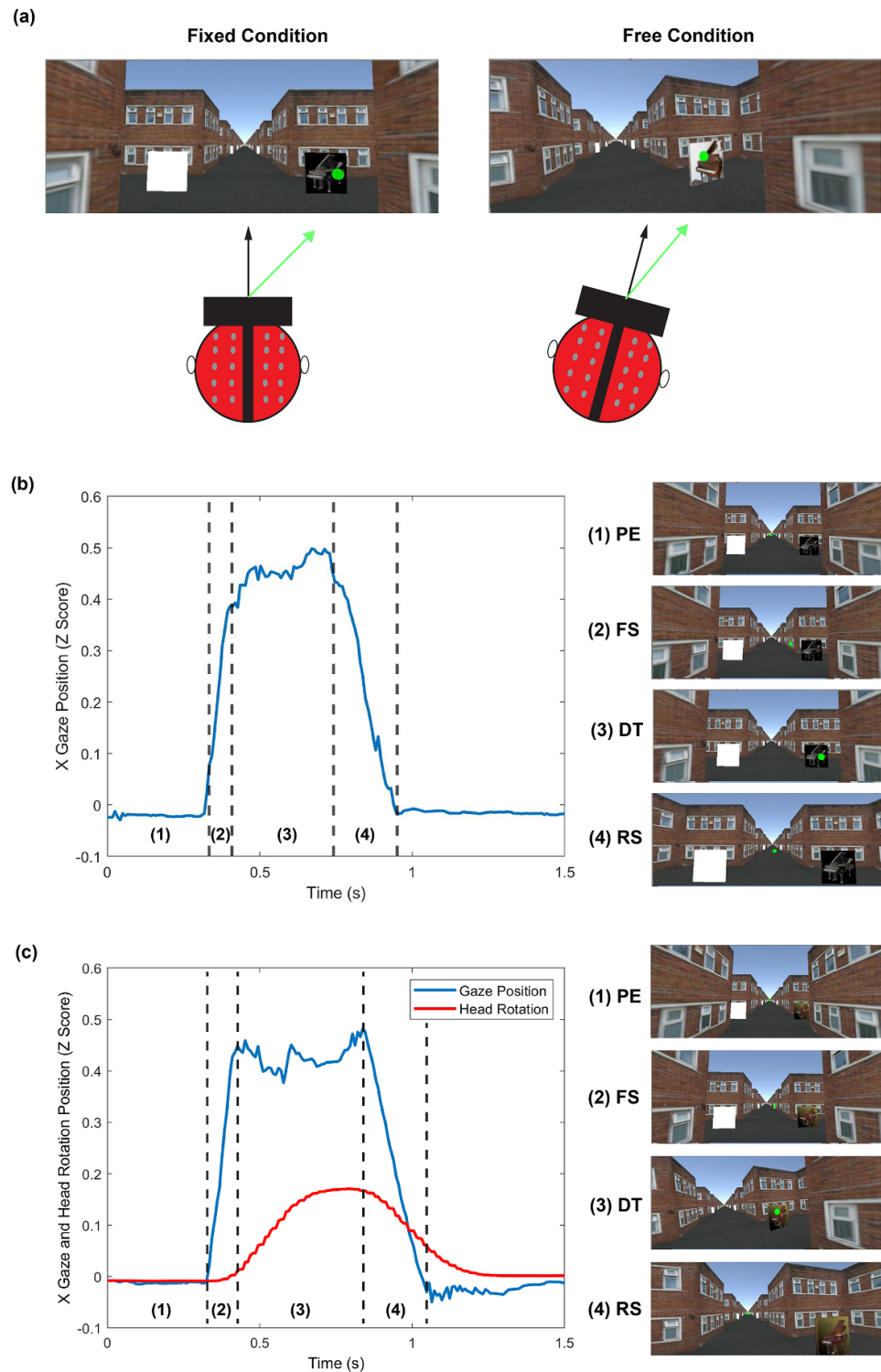
Figure 1. Experimental design and setup. (a) Screenshots of the experimental paradigm displayed inside the HMD along with drawings depicting the head and eye movements of the subject in the fixed condition (left) and free condition (right) when the stimuli appeared in the environment. Black arrow indicates the face-forward direction of the subject's head whereas the green arrow indicates the gaze direction. The green dot in the screenshots indicates where the subject's gaze falls in the virtual environment and is only visible in the experimenter's view. The subject only moves their eyes in the fixed condition to see the stimuli while the subject moves both their eyes and their head in the free condition. (b) The four phase of subject's gaze movement during each experimental trial in the fixed condition. An example of a subject's horizontal gaze movement during a trial in the fixed condition is plotted and divided into four different phases. Peripheral (PE) refers to the time of fixation on the center of the display before any gaze movement was made. First saccade (FS) refers to the duration of FS towards the target or distractor image. Dwell time (DT) refers to the time of fixation on the target or distractor image. Return saccade (RS) refers to the duration of return saccade towards the center of the display. The screenshots to the right correspond to each of the four phases depicted. (c) The four phases of subject's gaze movement during each experimental trial in the free condition. An example of a subject's horizontal gaze movement and head rotation during a trial in the free condition is plotted and divided into four phases similar to those in the fixed condition. The screenshots to the right correspond to each of the four phases depicted. The change in the background view corresponds to the head rotation of the subject.

filtered from 0.5 to 50 Hz and downsampled to 256 Hz. Noisy channels were removed using visual inspection (4 channels removed on average per subject). Independent component analysis (ICA) was performed to remove blinks and horizontal eye movement artifacts. EEG data was then epoched, relative to locking condition, from 0 to 1000 ms and baseline-corrected using the mean value from -200 to 0 ms.

Principal component analysis was then performed on the remaining EEG data and only the top 20 PCs were retained in order to reduce the number of feature space and avoid rank deficiency issues when performing classification. Temporal ICA was then performed on the data to ensure that the temlporal patterns of the activity were statistically independent from each other. The resulting ICs were used as input for the classifier described in the following section and results prior to ICA removal are presented in Supplementary figure 7 (available online at stacks.iop.org/JNE/18/066052/mmedia).

2.6. Data analysis

2.6.1. Hierarchical discriminant component analysis (HDCA)

In order to capture and integrate the neural and ocular reorienting response recorded by the EEG and eye-tracking signals, we adapted the hierarchical discriminant component analysis seen in [19] to build our hybrid classifier. First the epoched EEG ICs data were divided into 10, 100 ms, bins from 0 to 1000 ms relative to locking condition. Fisher linear discriminant analysis (FLDA) was performed on each bin to determine the within-bin weights across ICs:

$$\mathbf{w}_j = (\Sigma_+ + \Sigma_-)^{-1}(\mu_+ + \mu_-), \qquad (1)$$

where $\mathbf{w}_j$ is the vector of within-bin weights for bin j, μ and Σ are the mean and covariance of the EEG data in the current bin, and $+$ and $-$ subscripts refer to target and distractor trials, respectively. The weights $\mathbf{w}_j$ were then applied to the IC activations $\mathbf{x}_{ji}$ to determine the within-bin interest score z_{ji} for each bin i and each trial j:

$$z_{ji} = \mathbf{w}_j^T \mathbf{x}_{ji}. \qquad (2)$$

Similarly, FLDA was performed on the pupil diameter and dwell time data. The epoched pupil diameter data was divided into six 500 ms bin and averaged within each bin from 0 to 3000 ms based on locking condition. The average was passed through FLDA to determine within-bin interest score. The dwell time data was also passed through FLDA. The within-bin interest scores for each feature were then normalized by dividing by their standard deviation across trials. To construct the second-level feature vector, the EEG, pupil diameter and dwell time normalized interest scores were appended into a single column vector.

To visualize the contributions of each EEG data channel to the discriminating components, we calculated and plotted the scalp topography of the forward models for each 100 ms bin of the EEG data. For each bin j, the z_{ij} values were appended across trials into a column vector $\mathbf{z}_j$ and the $\mathbf{x}_{ji}$ vector into matrix $\mathbf{X}_j$. The forward model $\mathbf{a}_j$ can then be calculated as follows:

$$\mathbf{a}_j = \frac{\mathbf{X}_j \mathbf{z}_j}{\mathbf{z}_j^T \mathbf{z}_j}. \qquad (3)$$

For cross-bin classification, logistic regression was applied to the second-level feature vector z_i for each trial to determine the cross-bin weights v (across time bins and modalities):

$$\mathbf{v} = \arg\min_{\mathbf{v}} \left(\sum_i \log\{1 + \exp[-c_i \mathbf{v}^T \mathbf{z}_i]\} \right) \qquad (4)$$

where c_i denotes the class ($+1$ for targets and -1 for distractors) for trial i. The cross-bin weights were then used to calculate the final single cross-bin interest score y_i for each trial:

$$y_i = \mathbf{v}^T \mathbf{z}_i \qquad (5)$$

Ten-fold cross validation was used to create the training and testing sets. The area under the receiver operating characteristic (ROC) curve (AUC) was used to quantify the performance of the classifier. For comparison, we also constructed single-modality classifiers using the same procedures as described above but only using single-modality within-bin interest scores (EEG only, pupil diameter only or dwell time only).

2.6.2. Gaze events-based epoch time-locking

In order to explore the temporal variations in the reorienting signals, the EEG and pupil diameter data were epoched based on the timing of the gaze events during each specific trial—IO, FS and FF. As the name suggested, IO refers to the time point in which the image first appeared on the billboard for that trial. The EEG and pupil diameter data were then epoched with zero starting at the time of IO for that trial. FS refers to the time point in which the subject's eye began moving from center towards the image on the billboard while FF refers to the time point in which the subject's eye began fixating on the image on the billboard. Similarly, the EEG and pupil diameter data were then epoched with zero starting at the time of FS and FF for that trial, respectively.

2.6.3. General linear model (GLM) analysis

We further investigated the relationship between the orienting signals and gaze events by performing a general linear model (GLM) analysis. We fitted the discriminating components (e.g. the cross-bin interest score), y_i derived from the EEG-only and pupil diameter-only classifiers for each trial with the following four measurements derived from the piecewise modeling of gaze data, namely the initial fixation

(peripheral) time, the time of FS, the dwell time and the time of RS. All measurements were normalized within each subject before the GLM was performed. We utilized mixed-effects GLM in order to take into account the variability in the distributions of beta weights across subjects. The setup for our mixed-effects GLM is as followed:

$$Y_i = X_i\beta + Z_i b + \epsilon, \tag{6}$$

where Y_i refers to the vector of the discriminating components y_i, X_i refers to the gaze events time matrix, β refers to the gaze events time-effects vector, Z_i refers to the inter-subject variability design matrix, b refers to the inter-subject variability-effects vector and ϵ to the random error term. We also performed a second set of GLM analysis by first orthogonalizing the four different regressors with the dwell time of each trial before fitting it against the discriminating components derived from the EEG-only and pupil diameter-only classifiers. This is done in order to investigate the contributions of the three remaining time measurements (peripheral, FS and RS) without the effects of the dwell time.

3. Results

3.1. Behavioral gaze results

Gaze behavior of the subjects differs significantly between the fixed and the free conditions as indicated by the amount of time the subjects spent on each gaze event (figure 2). While the subjects spent roughly the same amount of time, on average, in the peripheral gaze event (Fx = 272 ms, Fr = 257 ms), they spent significantly longer amount of time in the free condition for the rest of the gaze events we identified (Student's paired *t*-tests, *p* < .001), including FS (Fx = 73 ms, Fr = 96 ms), dwell time (Fx = 320 ms, Fr = 533 ms) and RS (Fx = 161 ms, Fr = 207 ms). These results are in line with the example shown in figures 1(b) and (c). In the fixed condition, subjects' gaze travels to the image and tracks it during the dwell time section before returning to the middle fixation with no head movement. However, in the free condition, subjects' gaze first travels to the billboard before their head rotation follows, resulting in longer FS and dwell time on the billboard. Furthermore, their gaze then return to the middle fixation prior to their head rotation returning to the starting position, also leading to longer RS time. These results are in line with results found in previous eyetracking studies in which head movements were involved [20, 21].

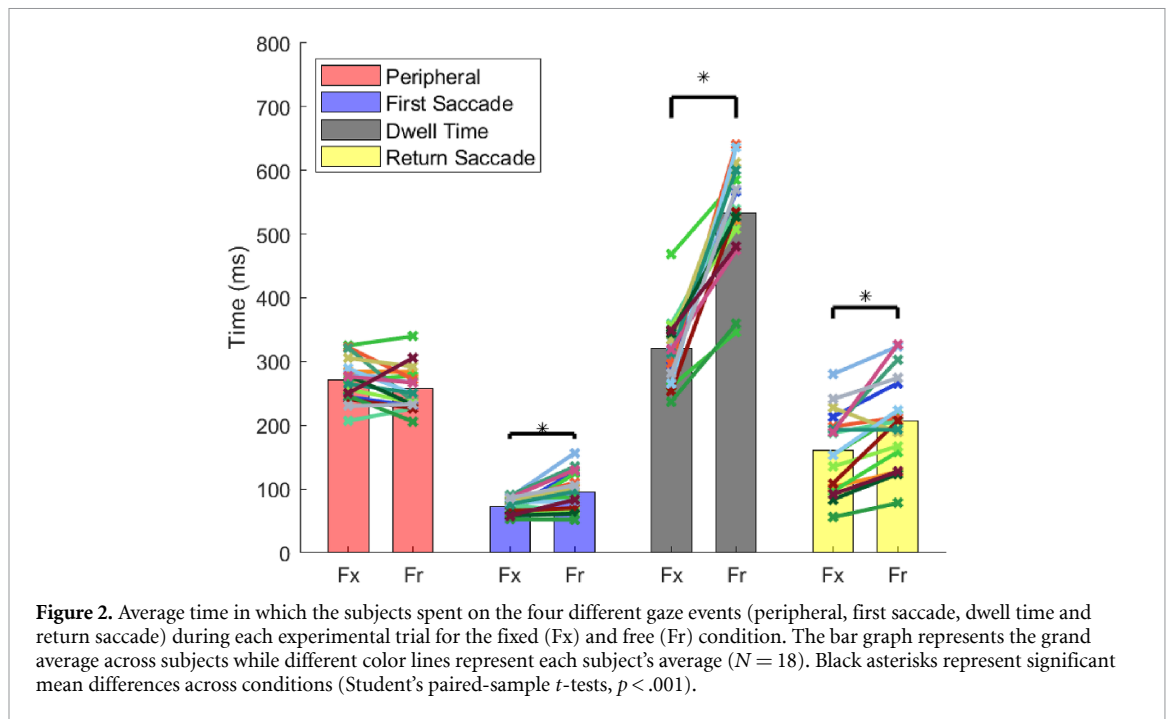## 3.2. Grand average pupil dilation and EEG ERPs results

Grand average EEG event related potentials (ERPs) for the three midline electrodes (Fz, Cz and Pz) are plotted in figure 3(a). The overall pattern and time course for the ERPs are in line with other target

detection studies [22–24]. The separation between the ERPs for the target and distractor trials are more pronounced in the Cz and Pz channels than in the Fz channel. Qualitatively, the P300 peak appears sharper in the fixed condition than in the free condition where it is more distributed over time. This result is expected due to the nature of the paradigm in which the subjects move their head in the free condition and spends more time across different gaze events (figure 2). Grand average pupil dilation across subjects for target and distractor trials are plotted in figure 3(b). The overall time course for pupil dilation (around 1–2 s following stimuli onset) is in line with the results from other target detection studies [19, 25]. Overall the pupil dilates more for target trials than for distractor trials in both the fixed and the free conditions. The sharper pupil dilation more pronounced in the fixed condition around 500 ms following stimuli onset may be explained by the ocular muscle-related dilation from the wide-angle saccade the subjects made to see the images on the billboards [26, 27].

3.3. Relationship between the orienting signals and gaze events

To determine the relationships between the EEG and pupil orienting signals and different gaze events time, we first developed EEG-only and pupil-only classifiers using the HDCA algorithm described in the Methods section. The cross-bin weights of the EEG-only classifier are shown in figure 4(a). The cross-bin weights for both the fixed and free condition peak roughly around 500–600 ms which correspond to the peak time of the P300 signal. Similarly, the forward models calculated from the EEG-only classifiers (figure 4(b)) also show the pattern of the P300 signal peaking roughly between 500 and 600 ms after stimuli onset. Figure 4(c) shows the cross-bin weights of the pupil diameter-only classifier. The cross-bin weights for the pupil diameter-only peak around 1700 ms for both the fixed and free condition, which also correspond to the time of grand average pupil dilation shown in figure 3(a).

Based on the results of the EEG-only and pupil diameter-only classifiers, we used the cross-bin interest scores (e.g. discriminating components) of each trial to be the representative of the strength of the orienting signals of that respective trial. We then performed a mixed-effects GLM fit between the EEG-only and pupil diameter-only discriminating components and the four different gaze events time. We also performed the same analysis after orthogonalizing the four different gaze events time against the dwell time for each trial. The GLM fit estimates for the EEG-only analysis are plotted in figure 5(a). The beta weight estimates (β) for both the fixed and free conditions are greatest for the dwell time. However, the beta weight for FS is only significant in the free and not the fixed condition. After the four regressors were orthogonalized against the dwell time

Figure 2. Average time in which the subjects spent on the four different gaze events (peripheral, first saccade, dwell time and return saccade) during each experimental trial for the fixed (Fx) and free (Fr) condition. The bar graph represents the grand average across subjects while different color lines represent each subject's average ($N = 18$). Black asterisks represent significant mean differences across conditions (Student's paired-sample *t*-tests, $p < .001$).
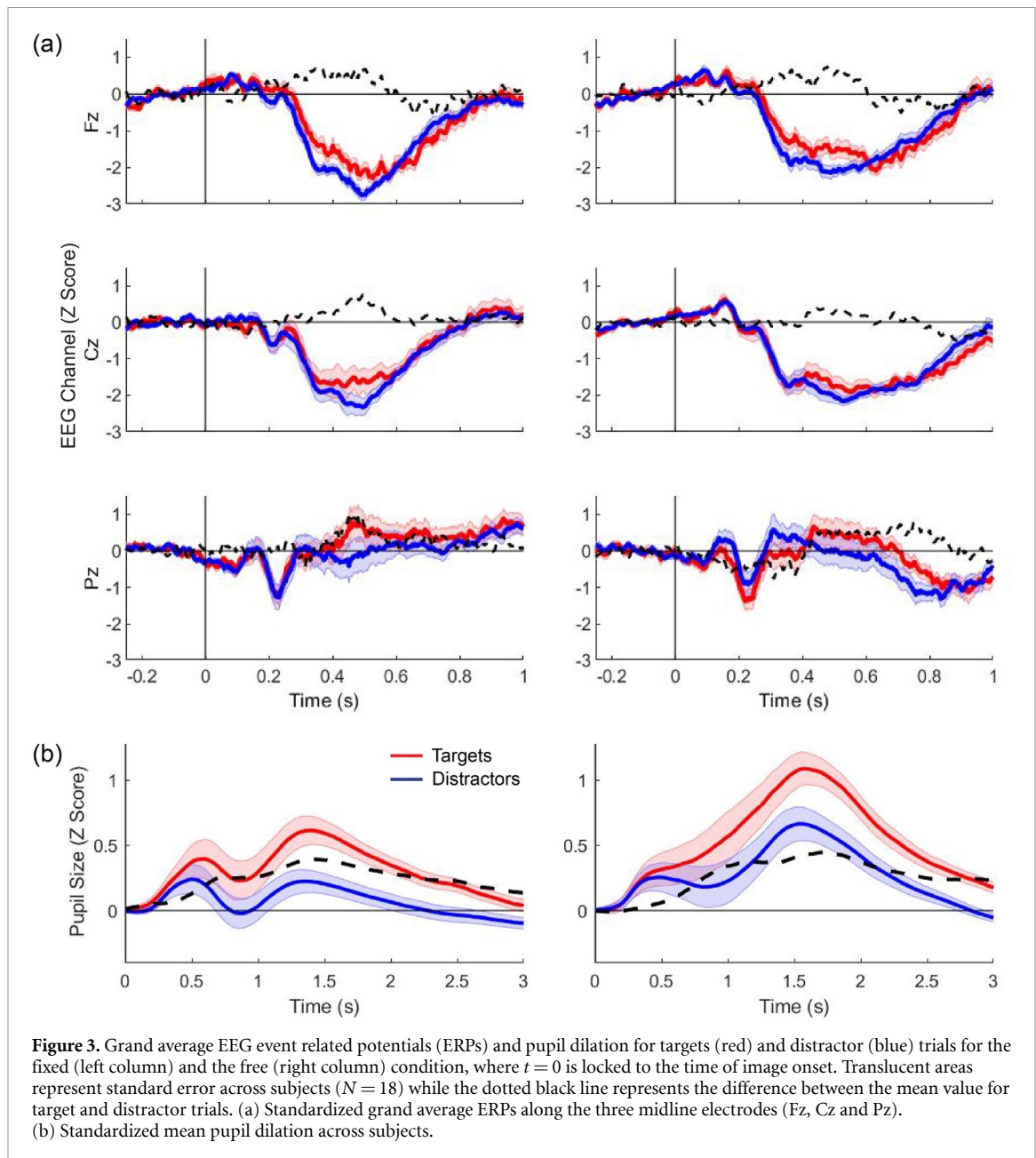
of each trial, the beta weight estimates become negative for the peripheral and FS time in the fixed condition and only for the peripheral time in the free condition. These results suggest that subjects tend to move their eyes away from center (e.g. lower peripheral time) during target trials both in the fixed and free condition and saccade towards targets faster in the fixed condition. Similarly, for the GLM estimates for the pupil diameter-only discriminating components (figure 5(b)), the beta weights are highest for the dwell time in both the fixed and free conditions with the beta weight for the second saccade being significant only in the free and not the fixed condition. The orthogonalized beta weight results for the pupil diameter-only discriminating components show significant negative values for the FS and RS in the fixed condition and peripheral and FS in the free condition. These results both demonstrate a shift forward in time compared to the orthogonalized EEG-only beta weight estimates.

3.4. Hybrid classifier performance

Following the development of the single-modality classifiers, we developed a hybrid classifier using the combination of EEG, pupil diameter and dwell time signals, in which the performance is shown in figure 6. Figure 6(a) shows each subject's AUC for the hybrid classifier compare to the single-modality classifiers. The subjects are sorted in descending order of the EEG-only AUC to highlight the importance of the hybrid classifier. Overall, the AUC of the hybrid classifier tracks and exceeds the AUC of the single-modality classifier which yields the highest AUC for that subject in both the fixed and in the free condition. We show that the hybrid classifier performed

significantly better than each of the single-modality classifier in figure 6(d) (Student's paired-sample *t*-tests, $p < .05$). The cross-bin weights and the EEG forward models of the hybrid classifier are shown in figures 6(b) and (c), respectively. The patterns for the cross-bin weights for both the EEG and the pupil diameter are similar to that of the cross-bin weights derived from the single-modality classifiers shown earlier in figures 4(a) and (c), with the EEG weights peaking around 500–600 ms and the pupil diameter weights peaking around 1700 ms. Similarly, the forward models derived from the hybrid classifier also show the pattern of the P300 signal peaking at approximately 500–600 ms following IO. In addition, we also compared the performance of the hybrid and single-modality classifiers across the fixed and the free conditions as shown in figure 6(e). We did not find any significant difference in the AUC for the hybrid or any of the single-modality classifiers across the two conditions (Student's paired-sample *t*-tests). This result demonstrates that the classifiers are able to capture the reorienting signals both in the control scenario and in the more naturalistic scenario of our experiment. Lastly, we compared the AUC results for the hybrid and single modality classifiers across different types of epoch time locking (as described in the Methods section). We found no significant differences across the three locking types (e.g. IO locked, FS locked and FF locked) for all classifiers in both the fixed and free conditions. This result demonstrates that the reorienting signals are not locked to one particular gaze-based event but are decomposed across multiple different gaze events, which is consistent with other results presented earlier in this study.
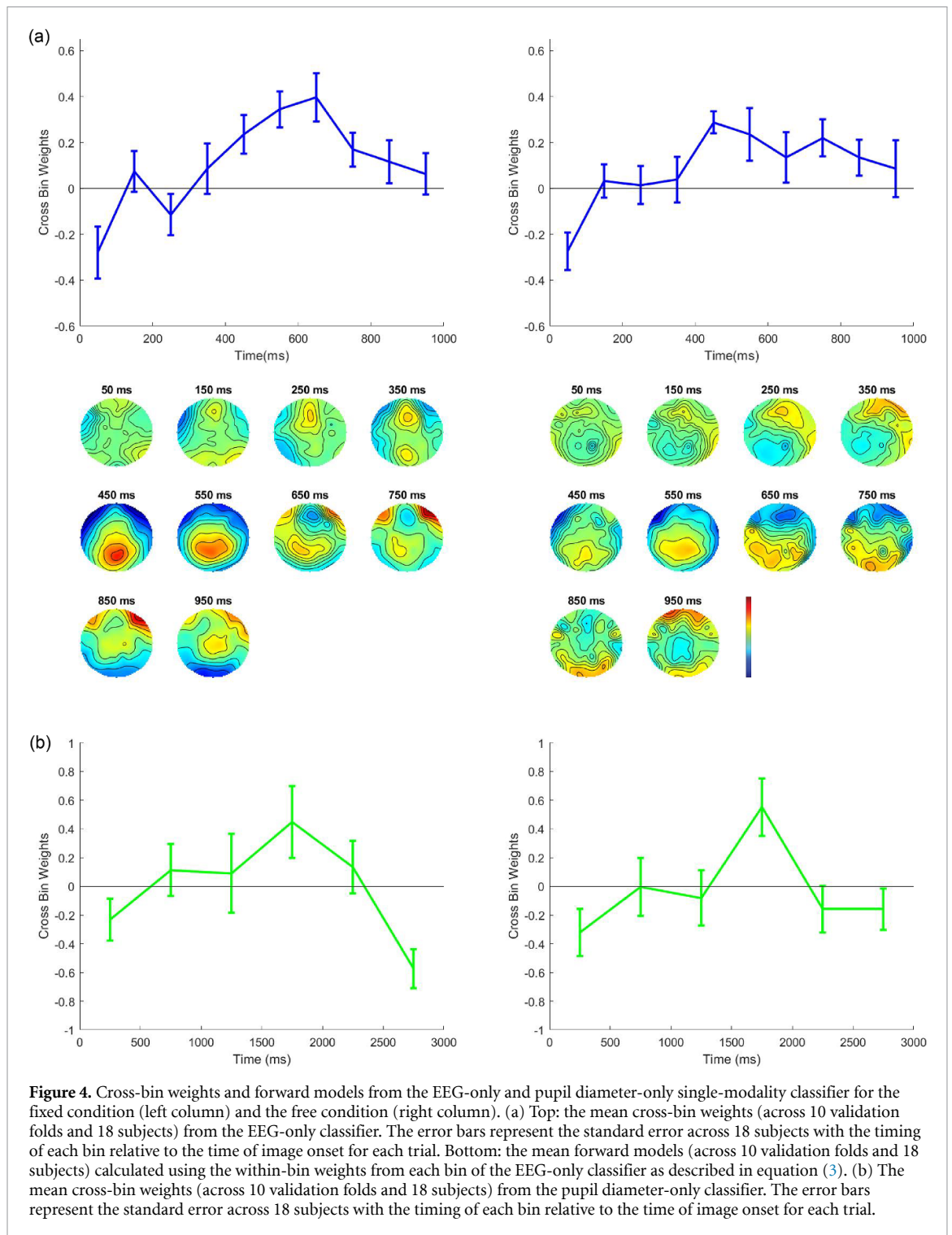
Figure 3. Grand average EEG event related potentials (ERPs) and pupil dilation for targets (red) and distractor (blue) trials for the fixed (left column) and the free (right column) condition, where $t = 0$ is locked to the time of image onset. Translucent areas represent standard error across subjects ($N = 18$) while the dotted black line represents the difference between the mean value for target and distractor trials. (a) Standardized grand average ERPs along the three midline electrodes (Fz, Cz and Pz). (b) Standardized mean pupil dilation across subjects.

4. Discussion

4.1. Moving towards more naturalistic experimental environments

Attention reorienting is without a doubt a complex set of processes. It involves multiple neural and physiological systems working together to redirect our attention to new and novel stimuli in the environment. Using standardized paradigms, typically with no head movement and minimal eye movement, previous studies have identified neural and physiological signatures associated with attention reorienting, namely the EEG P300 and pupil dilation [2, 12, 14, 22]. The fixed condition of our study mimics these standardized paradigms, by limiting the head movement of the subject and only allowing

eye saccades to be made. Unsurprisingly, the grand average ERP and pupil diameter results of the fixed condition show a clear and pronounced P300 and pupil dilation peaks. However, in the free condition where both head and eye movements were allowed, the P300 and pupil dilation become much more spatially and temporally distributed (figure 3). This result coincides with the behavioral results shown in figure 2 where the subjects take significantly longer time to saccade and fixate on the stimuli when head movements were made. Considering that many BCIs utilize these neural and physiological signals as measures of subject's attention, the greater spatial and temporal distributions of these signals pose direct challenge to the performance of these BCIs in more naturalistic environments. To address this issue, we

Figure 4. Cross-bin weights and forward models from the EEG-only and pupil diameter-only single-modality classifier for the fixed condition (left column) and the free condition (right column). (a) Top: the mean cross-bin weights (across 10 validation folds and 18 subjects) from the EEG-only classifier. The error bars represent the standard error across 18 subjects with the timing of each bin relative to the time of image onset for each trial. Bottom: the mean forward models (across 10 validation folds and 18 subjects) calculated using the within-bin weights from each bin of the EEG-only classifier as described in equation (3). (b) The mean cross-bin weights (across 10 validation folds and 18 subjects) from the pupil diameter-only classifier. The error bars represent the standard error across 18 subjects with the timing of each bin relative to the time of image onset for each trial.
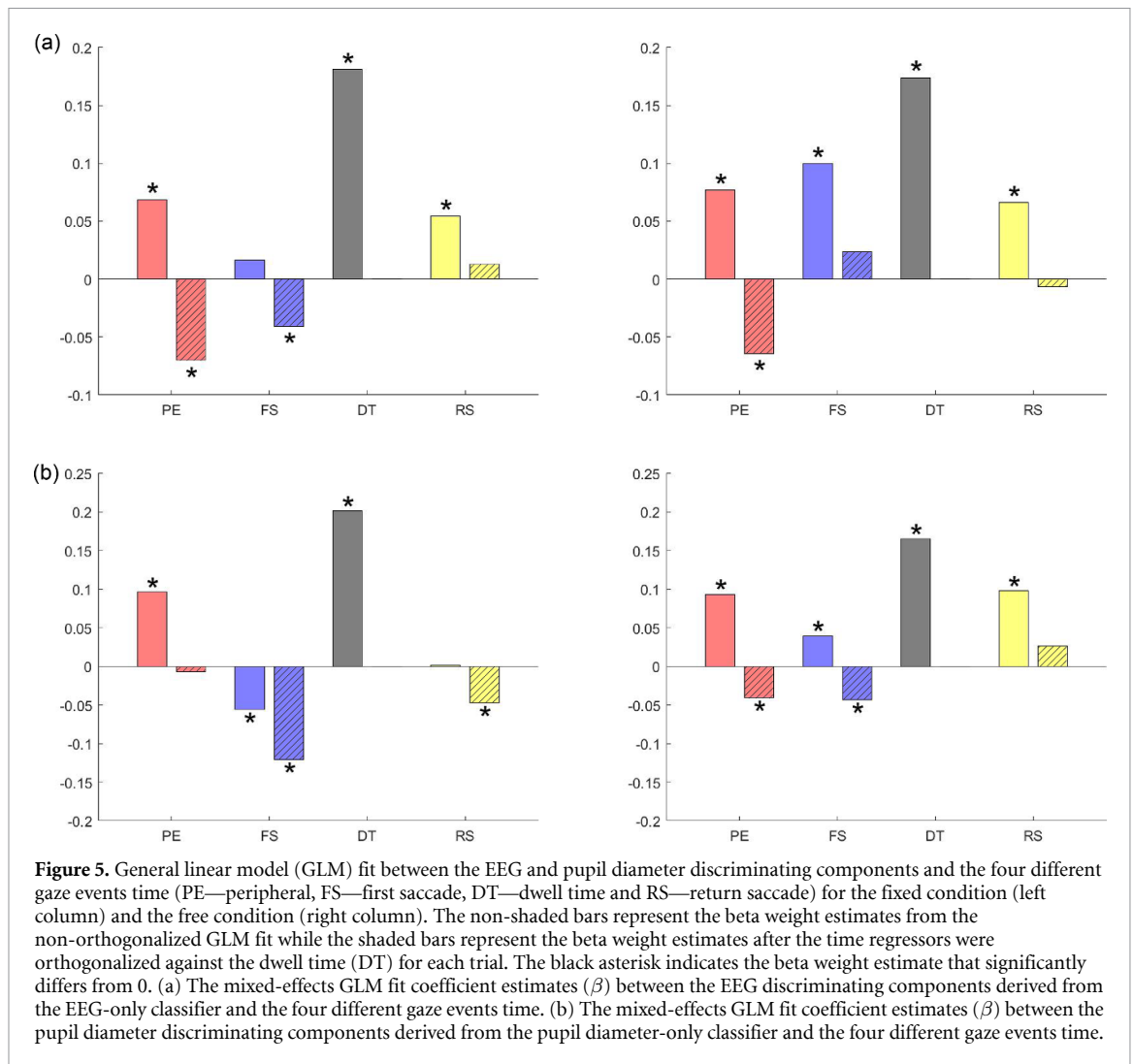
first explore the relationships between the neural and physiological signals associated with attention reorienting and the different gaze events taken place when subjects reorient their visual attention to the stimuli in the environment.

4.2. Relationship between gaze events and attention reorienting

In order to study the relationship between the orienting signals and different gaze events, we must first divide the continuous gaze information collected

from each trial into concrete events. We chose to divide the continuous gaze data into four distinct gaze events, peripheral, FS, dwell time and RS, as they are generally applicable to how a person might observe an object in real world environments and are understood to effect the reorienting response [28, 29]. In realistic scenarios such as the task employed in the current study, the subjects must not only reorient their attention to the stimuli but also reorient their attention back to the center fixation prior to the arrival of the subsequent stimuli. Therefore, we

Figure 5. General linear model (GLM) fit between the EEG and pupil diameter discriminating components and the four different gaze events time (PE—peripheral, FS—first saccade, DT—dwell time and RS—return saccade) for the fixed condition (left column) and the free condition (right column). The non-shaded bars represent the beta weight estimates from the non-orthogonalized GLM fit while the shaded bars represent the beta weight estimates after the time regressors were orthogonalized against the dwell time (DT) for each trial. The black asterisk indicates the beta weight estimate that significantly differs from 0. (a) The mixed-effects GLM fit coefficient estimates (β) between the EEG discriminating components derived from the EEG-only classifier and the four different gaze events time. (b) The mixed-effects GLM fit coefficient estimates (β) between the pupil diameter discriminating components derived from the pupil diameter-only classifier and the four different gaze events time.

consider the RS to be part of the reorientation loop. We performed the GLM analysis using the time of the four different gaze events as the regressors to the discriminating components derived from the EEG-only and the pupil diameter-only classifiers. The beta weight estimates in both the EEG-only and pupil diameter-only analyses and across both the fixed and free condition suggest that the dwell time of each trial contributes most significantly to the reorienting signals. Considering that the dwell time by itself can be used to distinguish between target and distractor stimuli in most subjects (figure 6(a)), and similarly in previous studies [19, 30, 31], this result confirms the importance of dwell time in attention reorienting. Here we also performed the second set of GLM analysis by orthogonalizing the dwell time, the most important contribution to the reorienting signals, against of the other three time regressors. With the dwell time removed, the negative beta weight estimates suggest that while the other gaze events are still important to the reorienting signals, they are negatively correlated. The EEG-only results suggest that the subjects spend less time fixating in the middle (i.e. lower peripheral gaze event time) when

target image appears in both the fixed and free condition and also saccade to the target image faster (e.g. lower FS time) in the fixed condition. The slight positive FS beta weights in the free condition may be explained by the longer FS time overall for that condition. Meanwhile the pupil diameter-only results show negative beta weights estimate for FSs and RSs for the fixed condition and for peripheral and FS for the free condition. These results demonstrate a forward shift in time in comparison to the EEG-only results, suggesting that the neural and ocular reorienting response might be processed by different but connecting brain regions. This theory is in line with recent works connecting the cortical signatures of reorienting mediated by the ventral attention system (e.g. the P300 signal) to that of the subcortical signatures (e.g. pupil dilation) mediated by the LC-NE system [8, 32]. It has been proposed that the activity of the LC is 'informed' by the connecting cortical structures such as the posterior cingulate cortex (PCC) and the anterior cingulate cortex (ACC) [7, 8, 32]. The results of the current study, specifically the forward shift in time in the pupil reorienting signals compared to the EEG reorienting signals as
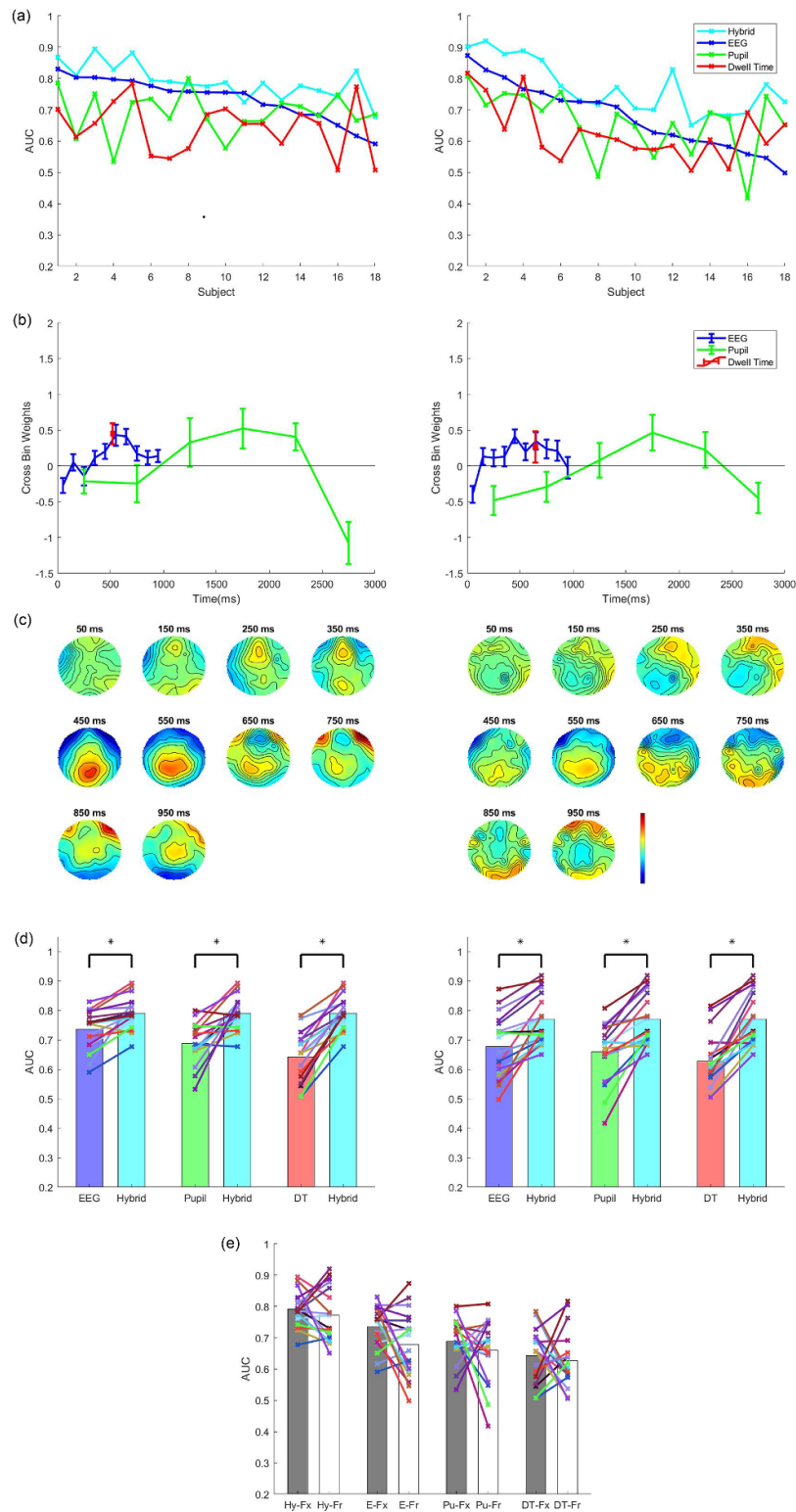
Figure 6. Results and performance of the hybrid classifier in comparison to single-modality classifiers for the fixed condition (left column) and the free condition (right column). (a) The area under the ROC curve (AUC) of the hybrid classifier and single-modality classifiers for all subjects ($N = 18$). The subjects are sorted by the performance of their EEG-only classifier. (b) The mean cross-bin weights (across 10 validation folds and 18 subjects) from the hybrid classifier. The error bars represent the standard error across 18 subjects with the timing of each bin relative to the time of image onset for each trial. The dwell time weight's horizontal position and error bars represent the mean and standard error of the subject's mean dwell time. (c) The mean forward models (across 10 validation folds and 18 subjects) calculated using the within-bin weights from each bin of the hybrid classifier. (d) The bar graphs represent the mean AUC across subjects for the hybrid classifier in comparison to the EEG-only, pupil diameter-only (pupil) and dwell time-only (DT) classifiers. Each line graph represents the subject's AUC for each type of classifier. Hybrid classifiers perform significantly better than all single-modality classifiers in both conditions. (e) The bar graphs represent the mean AUC across subjects for the hybrid (Hy), EEG-only (E), pupil diameter-only (Pu) and dwell time-only (DT) in comparison between the fixed (Fx) and the free (Fr) conditions. Each line graph represents the subject's AUC for each type of classifier and condition. Overall, the mean AUC does not differ significantly across conditions for all types of classifiers.

indexed by the gaze events, provide support for this theory.

4.3. Capturing and integrating attention reorienting signals in naturalistic environments

One of the main aims of the current study is to capture and integrate the neural and physiological signals underlying attention reorienting in naturalistic environments. While the hybrid HDCA classifier has previously been shown to successfully classify target and distractor stimuli in a 2D screen-based environment [19], our study is the first application of the hybrid HDCA classifier in a VR-based 3D environment. The results of the current study show that not only were the hybrid classifier able to classify the type of stimuli the subjects observed in a more immersive and naturalistic environment, it was able to perform equally well even when the subjects moved their heads in the free condition. The implication of this result is that despite the greater temporal distribution of the reorienting signals across trials in the more naturalistic condition, the hybrid classifier is still able to capture and integrate the information within these signals. We also demonstrate the benefits of utilizing multiple neural and physiological signal modalities to improve the classification performance of the classifier. While each single modality (EEG, pupil diameter and dwell time) contains the reorienting information on its own, combining the information across modalities significantly improves the classification performance both in the fixed and in the free condition. While the use of a hybrid classifier to classify targets vs. non target stimuli is still rare, the performance of our classifier is comparable to those of previous target detection studies typically done outside of a VR headset [19, 33, 34]. Our results therefore suggest that the hybrid HDCA classify may potentially serve as a basis for the development of attention-based BCI applications that can perform well in realistic scenarios and not only in well-controlled experimental environments.

4.4. Limitations/future directions

While the current study has shed light on some of the questions surrounding the dynamics of attention reorienting signals in naturalistic environments, many of them still remain unanswered. One of the major limitations to our study design is despite the subjects' ability to move their head, the movement is still limited to one plane of motion. With the use of HMD VR goggles, a study in which subjects are free to move in all planes of motion in a 'visual search' task may answer further questions regarding the orienting of attention in realistic scenarios [35, 36]. In addition, while the current study attempted to divide the subjects' gaze direction into distinct events, gaze movements in realistic scenarios have been shown to be more complex, with saccade and fixation events constantly interleaving in time [37, 38]. Lastly, while

the hybrid HDCA classifier demonstrates good performance in the current work, further studies are required to investigate the possibility of applying it in a closed-loop system in order to serve as a basis for the development of a real-time BCI application.

5. Conclusion

In this study, we explored the relationship between gaze events and attention reorienting signals in a more naturalistic environment. We determined that dwell time contributes most significantly to both the ocular and neural reorienting signals. However, the distribution of the reorienting signals across the remaining gaze events, namely peripheral, FS and RS, are different across the two modalities. Specifically, the pupil reorienting signals show a forward shift in time in comparison to the EEG reorienting signals, consistent with the theory in which the cortical regions of the ventral attention network (e.g. ACC and PCC) modulates the activity of the subcortical regions associated with the reorienting process (e.g. the LC-NE system). Nevertheless, when applying the hybrid classifier which combines the EEG, pupil dilation and dwell time signals together, it was able to capture and integrate the reorienting signals across different modalities and classify target vs. distractor stimuli with high accuracy. We expect the results of this study will provide the basis for the development of an attention-based BCI system that can operate in more naturalistic environment in the future.

Data availability statement

The data that support the findings of this study will be openly available following an embargo at the following URL/DOI: https://github.com/LIINC/LIINC_VR_Reorienting. Data will be available from 30 November 2021[39].

ORCID iDs

Sharath Koorathota ● https://orcid.org/0000-0003-2014-424X
Qi Wang ● https://orcid.org/0000-0001-8656-1439
Paul Sajda ● https://orcid.org/0000-0002-9738-1342

References

[1] Corbetta M, Patel G and Shulman G L 2008 The reorienting system of the human brain: from environment to theory of mind *Neuron* **58** 306–24

[2] Gabay G, Pertzov Y and Henik A 2011 Orienting of attention, pupil size and the norepinephrine system *Atten. Percept. Psychophys.* **73** 123–9

[3] Bradley M M 2009 Natural selective attention: orienting and emotion *Psychophysiology* **46** 1–11

[4] Pritchard W S 1981 Psychophysiology of P300 *Psychol. Bull.* **89** 506–40

[5] Polich J 2007 Updating P300: an integrative theory of P3a and P3b *Clin. Neurophysiol.* **118** 2128–48

[6] Nieuwenhuis S, Aston-Jones G and Cohen J D 2005 Decision making, the P3 and the locus coeruleus-norepinephrine system *Psychol. Bull.* **131** 510–32

[7] Aston-Jones G and Cohen J D 2005 An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance *Annu. Rev. Neurosci.* **28** 403–50

[8] Joshi S and Gold J I 2020 Pupil size as a window on neural substrates of cognition *Trends Cogn. Sci.* **24** 466–80

[9] Liu Y, Rodenkirch C, Moskowitz N, Schriver B and Wang Q 2017 Dynamic lateralization of pupil dilation evoked by locus coeruleus activation results from sympathetic, not parasympathetic, contributions *Cell Rep.* **20** 3099–112

[10] Arvaneh M, Robertson I H and Ward T E 2018 A P300-based brain-computer interface for improving attention *Front. Hum. Neurosci.* **12** 524

[11] Li Y, Long J, Yu T, Yu Z, Wang C, Zhang H and Guan C 2010 An EEG-based bci system for 2-D cursor control by combining Mu/Beta rhythm and P300 potential *IEEE Trans. Biomed. Eng.* **57** 2495–505

[12] Capotosto P, Corbetta M, Romani G L and Babiloni C 2012 Electrophysiological correlates of stimulus-driven reorienting deficits after interference with right parietal cortex during a spatial attention task: a TMS-EEG study *J. Cogn. Neurosci.* **24** 2363–71

[13] Halder S, Hammer E M, Kleih S C, Bogdan M, Rosenstiel W, Birbaumer N and Kübler A 2013 Prediction of auditory and visual P300 brain-computer interface aptitude *PLoS One* **8** e53513

[14] Jin J, Zhang H, Daly I, Wang X and Cichocki A 2017 An improved P300 pattern in BCI to catch user's attention *J. Neural Eng.* **14** 036001

[15] Jangraw D C, Johri A, Gribetz M and Sajda P 2014 NEDE: an open-source scripting suite for developing experiments in 3D virtual environments *J. Neurosci. Methods.* **235** 245–51

[16] Fei-Fei L, Fergus R and Perona P 2007 Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories *Comput. Vis. Image Underst.* **106** 59–70

[17] Kothe C A 2007 Labstreaming layer (LSL): a system for unified collection of measurement time series in research experiments

[18] Delorme A and Makeig S 2004 Eeglab: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis *J. Neurosci. Methods* **134** 9–21

[19] Jangraw D C, Wang J, Lance B J, Chang S-F and Sajda P 2014 Neurally and ocularly informed graph-based models for searching 3D environments *J. Neural Eng.* **11** 046003

[20] Freedman E G 2008 Coordination of the eyes and head during visual orienting *Exp. Brain Res.* **190** 369–87

[21] Fang Y, Nakashima R, Matsumiya K, Kuriki I and Shioiri S 2015 Eye-head coordination for visual cognitive processing *PLoS One* **10** e0121035

[22] Kirino E, Belger A, Goldman-Rakic P and McCarthy G 2000 Prefrontal activation evoked by infrequent target and novel stimuli in a visual target detection task: an event-related functional magnetic resonance imaging study *J. Neurosci.* **20** 6612–8

[23] Bledowski C, Prvulovic D, Hoechstetter K, Scherg M, Wibral M, Goebel R and Linden D E J 2004 Localizing P300 generators in visual target and distractor processing: a combined event-related potential and functional magnetic resonance imaging study *J. Neurosci.* **24** 9353–60

[24] Linden D E, Prvulovic D, Formisano E, Völlinger M, Zanella F E, Goebel R and Dierks T 1999 The functional neuroanatomy of target detection: an fMRI study of visual and auditory oddball tasks *Cereb. Cortex* **9** 815–23

[25] Hong L, Walz J M and Sajda P 2014 Your eyes give you away: prestimulus changes in pupil diameter correlate with poststimulus task-related EEG dynamics *PLoS One* **9** e91321

[26] Brisson J, Mainville M, Mailloux D, Beaulieu C, Serres J and Sirois S 2013 Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers *Behav. Res. Methods* **45** 1322–31

[27] Mathôt S, Fabius J, Van Heusden E and Van der Stigchel S 2018 Safe and sensible preprocessing and baseline correction of pupil-size data *Behav. Res. Methods* **50** 94–106

[28] Rafal R D, Calabresi P A, Brennan C W and Sciolto T K 1989 Saccade preparation inhibits reorienting to recently attended locations *J. Exp. Psychol.: Hum. Percept. Perform.* **15** 673

[29] Unema P J A, Pannasch S, Joos M and Velichkovsky B M 2005 Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration *Vis. Cogn.* **12** 473–94

[30] Aydarkhanov R, Uscumlic M, Chavarriaga R, Gheorghe L and Millan J D R 2021 Closed-loop EEG study on visual recognition during driving *J. Neural Eng.* **18** 026010

[31] Suefusa K and Tanaka T 2017 A comparison study of visually stimulated brain–computer and eye-tracking interfaces *J. Neural Eng.* **14** 036009

[32] Dinuzzo M, Mascali D, Moraschi M, Bussu G, Maugeri L, Mangini F, Fratini M and Giove F 2019 Brain networks underlying eye's pupil dynamics *Front. Neurosci.* **13** 965

[33] Kim M, Kim B H and Jo S 2015 Quantitative evaluation of a low-cost noninvasive hybrid interface based on EEG and eye movement *IEEE Trans. Neural Syst. Rehabil. Eng.* **23** 159–68

[34] Ma X, Yao Z, Wang Y, Pei W and Chen H 2018 Combining brain-computer interface and eye tracking for high-speed text entry in virtual reality *23rd Int. Conf. on Intelligent User Interfaces* pp 263–7

[35] Olk B, Dinu A, Zielinski D J and Kopper R 2018 Measuring visual search and distraction in immersive virtual reality *R. Soc. Open Sci.* **5** 172331

[36] Kollenberg T, Neumann A, Schneider D, Tews T-K, Hermann T, Ritter H, Dierker A and Koesling H 2010 Visual search in the (un)real world: how head-mounted displays affect eye movements, head movements and target detection *ETRA '10: Proc. 2010th Symp. on Eye-Tracking Research & Applications* pp 121–4

[37] Clay V, König P and König S 2019 Eye tracking in virtual reality *J. Eye Mov. Res.* **12** 1

[38] Sidenmark L and Gellersen H 2019 Eye & head: synergistic eye and head movement for gaze pointing and selection *UIST '19: Proc. 32nd Annual Symp. on User Interface Software and Technology* pp 1161–74

[39] Lapborisuth P 2021 (https://doi.org/10.1088/1741-2552/ac4593)