Brainwave-Augmented Eye Tracker: High-Frequency SSVEPs Improves Camera-Based Eye Tracking Accuracy

Brainwave-Augmented Eye Tracker: High-Frequency SSVEPs Improves Camera-Based Eye Tracking Accuracy

Alexandre Armengol-Urpi
Massachusetts Institute of Technology
Cambridge, MA, USA
armengol@mit.edu

Andres F. Salazar-Gomez
Massachusetts Institute of Technology
Cambridge, MA, USA
salacho@mit.edu

Sanjay E. Sarma
Massachusetts Institute of Technology
Cambridge, MA, USA
sesarma@mit.edu

ABSTRACT

In this work, we leverage neural mechanisms of visual attention to improve the accuracy of a commercial eye tracker through the analysis of electroencephalography (EEG) waves. Gaze targets were rendered in a computer screen with imperceptible flickering stimuli ($\geq 40Hz$) that elicited attention-modulated steady-state visual evoked potentials (SSVEPs). Our hybrid system combines EEG and eye-tracking modalities to overcome accuracy limitations of the gaze-tracker alone. We integrate EEG and gaze data to efficiently exploit their complementary strengths driving a Bayesian probabilistic decoder that estimates the target gazed by the user. Our system's performance was analyzed across the screen with varying target sizes, spacings and dataset epoch lengths, using data from 10 subjects. Overall, our hybrid approach improves the classification accuracy of the eye tracker alone for all target parameters and dataset epoch lengths in 11 units on average. The system shows a larger impact at peripheral screen regions where performance enhancement is maximal, reaching improvements of over 45 units. The findings of this work demonstrate that the intrinsic accuracy limitations of camera-based eye-trackers can be corrected with the integration of EEG data, and opens opportunities for gaze tracking applications with higher target granularity.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Graphical user interfaces**; *HCI theory, concepts and models.*

KEYWORDS

HCI, BCI, gaze detection, eye-tracking, SSVEP, EEG

ACM Reference Format:
Alexandre Armengol-Urpi, Andres F. Salazar-Gomez, and Sanjay E. Sarma. 2022. Brainwave-Augmented Eye Tracker: High-Frequency SSVEPs Improves Camera-Based Eye Tracking Accuracy. In *27th International Conference on Intelligent User Interfaces (IUI '22), March 22–25, 2022, Helsinki, Finland.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3490099.3511151

IUI '22, March 22–25, 2022, Helsinki, Finland

ACM ISBN 978-1-4503-9144-3/22/03.
https://doi.org/10.1145/3490099.3511151

1 INTRODUCTION

Camera-based eye tracking is a technique that consists of monitoring a person's eye movements in order to estimate where the individual is looking at any given time. Due to its easily deployable nature and relatively cheap cost, eye tracking has become a popular method to enhance interaction with a computer. Gaze monitoring enables myriad applications that range from direct gaze input [34, 44, 80] to cognition and perception studies [13, 78], market research [75] and even healthcare [7].

The performance of gaze-enabled computer interfaces as well as the validity of research outcomes based on eye-tracking analysis are strictly dependent on the quality of gaze data. Hence, it is of vital importance to account for the limitations of the eye-tracking system when designing an application or a research experiment [23, 30]. The quality of gaze data obtained by an eye-tracker is given in terms of accuracy and precision in visual angle units, and they are usually provided by the eye-tracker manufacturers. Accuracy refers to the absolute difference between the true and the estimated gaze direction. Precision is computed as the standard deviation of the measured gaze points during a fixation. Past studies have shown that tracking quality values reported by manufacturers usually differ from the numbers obtained by researchers [29, 63, 64] and that these quality measures vary greatly across different tracking conditions and users.

Factors that affect the quality of the measured gaze data are very broad, and they include, among others, head movements, bad calibrations, astigmatism and camera resolution [6, 32], as well as inherent factors such as natural eye jitters during fixations [36]. Moreover, accuracy and precision are not constant across the visual field, but depend strongly on the visual angle between the fixation point and the camera [23]. Therefore, in the context of eye-tracking in a computer display, gaze data quality will depend on the region in the screen the user is gazing at; peripheral areas will show lower accuracy and precision than central regions.

Accuracy and precision of an eye-tracking system will establish the limitations of an application or a research experiment with respect to the sizes and spacings between targets or regions of interest (ROIs). Ideally, targets and ROIs should be large and far enough apart so as to not be affected by the accuracy limitations of the eye tracker and the intrinsic jitter of eye movements. This, however, heavily restricts the amount of targets available and impairs the granularity of the system. In studies where the stimuli are closely spaced, lack of accuracy can be critical in the proper analysis of gaze data.

There has been countless efforts in the field to improve gaze data quality of a given system [6, 32] or to mitigate such limitations with creative strategies, such as using fish-eye lens to magnify the region

the user is looking at [4], correcting inaccuracies with manual mouse adjustments or voice commands [53, 76], using gaze gestures [19] or exploiting Fitts's Law to predict eye movements [22, 80]. These approaches, however, generally undermine the workflow of the interface, making it unnatural and less intuitive, or affect the original appearance of the displayed elements. Also, many of these solutions suffer from the so called "Midas Touch" problem, which is an effect that originates from the fact that humans use vision to alternate between scanning the environment and gazing at features that deserve attention [36, 37]. This can cause interactive elements to unintentionally activate as the eye is simply traversing the scene, which can be annoying and less seamless for the user. Ultimately, even in applications that simply consist of recording gaze data with no interactions, the core problem is still that eye-trackers suffer from a systematic error even right after rigorous calibration [32, 56]. And according to [49], even if an eye-tracker was perfectly accurate, the practical accuracy of the system would be restricted to about 0.5-1 degrees due to the size of the fovea, the area of high visual acuity.

Therefore, it is reasonable to think that the performance of eye-tracking systems would benefit from joining efforts with other modalities that are less sensitive to the many factors that affect gaze data quality or the workflow of the interaction. In this work, we leverage neural mechanisms of selective visual attention to boost the accuracy of gaze data through the analysis of evoked brainwaves in response to embedded visual stimuli.

Selective visual attention is the tendency of visual processing to optimally allocate neurocognitive resources to stimuli that are relevant for our behavioral goals [15, 54]. Visual attention increases neural responses to stimuli, which can be exploited to determine which region in the visual field is receiving a person's focus. One of the most thoroughly demonstrated processes of this phenomenon is the enhancement of Steady-State Visually Evoked Potentials (SSVEPs) when a person attends a flickering stimuli [43, 55]. SSVEP are oscillatory waves generated at the visual cortex that occur in response to visual stimulation at specific frequencies [26, 28]. When a person shifts the attention to a particular flickering stimuli, the amplitudes of the evoked SSVEPs increase with respect to when the stimuli is ignored [55, 57]. Interestingly, SSVEPs are evoked even if the flickers are too fast for the user to consciously perceive them, which usually happens for frequencies larger than 40Hz, the so called critical flicker frequency (or critical fusion rate) [68]. Since SSVEPs can be generated at up to ~90Hz [26], there is a frequency band in which SSVEPs will occur and flicker will be invisible to humans. This invisible flickering frequency band is the one we are interested in for the purposes of this work.

In this paper, we propose the concept of embedding invisible visual flicker in the computer display to exploit visual attention neural processes and combine it with gaze data using a probabilistic Bayesian approach. The fact that such neural mechanisms are not sensitive to the factors that limit data quality of traditional gaze-trackers makes this approach a potential solution to improve the accuracy of camera-based eye-tracking systems. Moreover, the original appearance of the displayed scene is minimally altered because of the invisibly flickering visual stimuli. To the best of our knowledge, this is the first time that accuracy of an eye-tracker system is enhanced using electroencephalography (EEG) analysis.

2 RELATED WORK

The symbiotic combination of eye tracking and EEG modalities has been largely explored for a wide range of different applications. For instance, it has been proposed as a solution for target selection in the Human-Computer Interface (HCI) context [65, 66, 77, 79, 81], where gaze data provides spatial information of user's attention and the EEG channel acts as the event or object selection trigger.

EEG and gaze hybrid approaches have also been pursued to enhance cursor control. In [33], the precision of a continuous cursor control is improved using EEG information compared to pure gaze-control, and in [17] the combination of EEG and gaze data led to an increased completion accuracy and reduced task completion time without explicit gaze behavior required by the user.

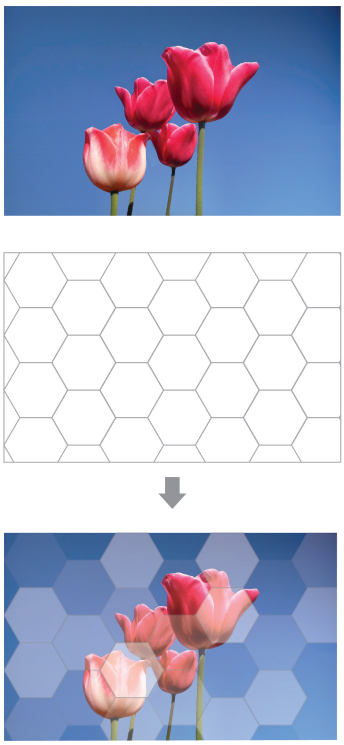
Combining eye-tracking with SSVEP has also been explored. In [10], the authors combine gaze data with low-frequency SSVEPs for an interface with more accurate communication and control. They developed a 4-target menu navigation system for users with and without brain injury. In [20], they merged eye-tracking and SSVEP inputs to enhance the classification accuracy of a target selection interface. They demonstrated speed up in operation and performance over existing gaze hybrid BCI systems. The authors in [70] proved that by combining SSVEP with gaze data they could achieve improved user control with respect to pure eye-tracking or pure SSVEP interfaces.

Merging EEG and eye-tracking modalities has also been exploited beyond the field of HCIs. For example, eye fixation-related potentials (EFRP) are key to study certain cognitive processes during free natural exploration of visual stimuli. EFRPs allow synchronizing the recording of neurocognitive processes that occur during visual exploration by correlating perceptual events with a sequence of eye movements and/or fixations [35, 39]. Similar approaches combining these two data modalities have been followed to investigate physiological decision processes during marketing studies [41], for emotion recognition [82], for the identification of children with autism spectrum disorder [40] or even for quadcopter or limb control [42, 52].

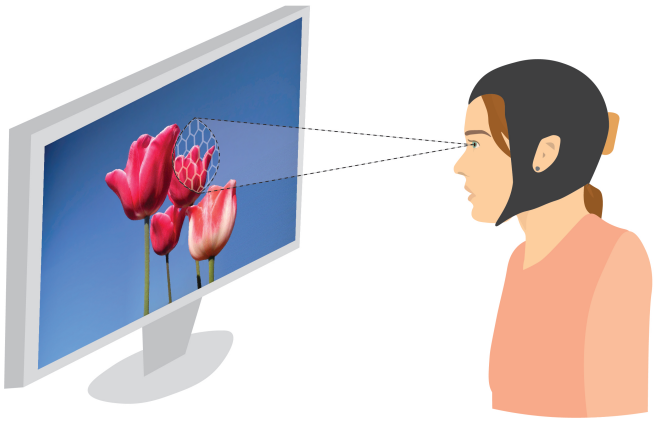
The utilization of high-frequency SSVEPs has also been recently used in the field of Brain-Computer Interfaces (BCIs). In order to make these interfaces more user-friendly and deployable, minimizing eye fatigue and user discomfort created by the flickering stimuli is essential. Therefore, SSVEP-based BCIs with frequencies above the critical fusion rate are increasingly being adopted by researchers [3, 25, 38, 67].

3 CONCEPT

The concept we propose consists of leveraging the neural mechanisms of visual attention to improve the accuracy of a commercial eye-tracker alone. To do so, we need to create invisible visual flicker on the screen that will evoke the corresponding SSVEPs when the user attends a particular element. These flickers are generated by modulating the luminance at high frequencies so they are not consciously perceived by the user. This can be implemented by partitioning the screen into equally sized fragments or tiles, whose luminance is individually modulated. The size and spacing between fragments will delimit the accuracy bounds of the system. See Figure 1a for a schematic representation of the concept.



(a) The screen is partitioned into equally sized hexagonal fragments.



(b) Visualization of a proposed implementation of our concept, where the flickering mask is only applied at the approximate region where the user is gazing. The fast sampling rate of the eye trackers is exploited. Then, leveraging the EEG system, a more precise point of regard can be determined.

Figure 1: Schematic representation of the concept we propose. The screen is segmented into different equally sized tile-like fragments. Each region luminance is invisibly modulated to encode information which can be decoded through the EEG.

The encoding element to discriminate between targets or screen fragments can be either frequency or phase, since it is well known that elicited SSVEPs synchronize in both frequency and phase to the visual stimuli the user is attending [27, 45, 51, 72, 74]. For simplicity, in this work we adopted the frequency encoding approach.

Modulating the luminance of the whole screen simultaneously is an inefficient strategy, both computationally and in terms of misusing the limited amount of available frequencies for stimulation. Hence, for a real world application, we propose to leverage the fast sampling rate of eye trackers (usually above 60Hz) to determine the approximate region in the screen the user is gazing at, and then apply the flickering mask only in that region to pinpoint the exact spot of attention (see Figure 1b). However, in this paper, we did not implement this adaptive solution since we were more interested in studying the accuracy improvements of our system once the subject is already focusing on a specific screen region.

4 METHODS

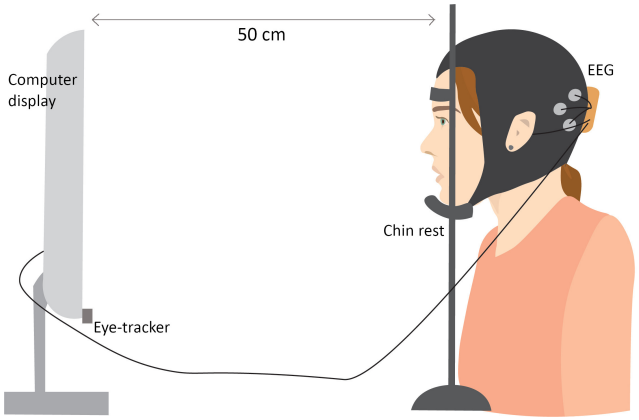
4.1 Participants

10 healthy BCI naive subjects, age 25-36 (mean=28.4, SD=3.0) with normal or corrected-to-normal vision participated in this study. Subjects were seated in front of a monitor and placed their heads on a chin rest so that the distance and visual angles to the stimuli were kept constant. This also guaranteed that subjects' head was always inside the virtual head box required by the eye-tracker manufacturer for a proper gaze monitoring. The chin rest was located 50 cm away from the monitor, as seen in Figure 2a. The study was approved by the institution's committee on the use of humans as experimental subjects.

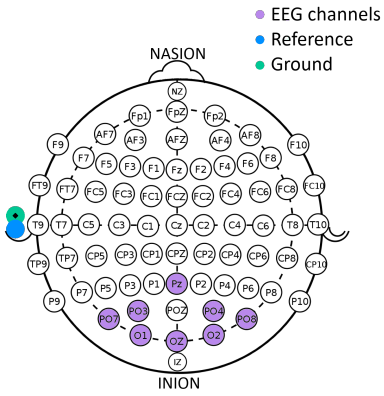
4.2 Equipment

4.2.1 Eye Tracker. Eye gaze was monitored using a Tobii Pro Nano [58] with 60Hz sampling rate, which was attached at the center of the lower part of the monitor frame. Tobii reports an accuracy of 0.3° and a precision of 0.1° RMS at optimal conditions [1]. Tobii specifies that the optimal screen size should be of up to 24". Since we used a larger monitor, we restricted the active area to match the size of a 24" display.

4.2.2 EEG equipment. EEG data was acquired using an Enobio system from Neuroelectrics [60] with 8 wet electrodes located in the parietal and occipital regions using the 10-20 distribution (PO3, PO4, PO7, PO8, Pz, O1, Oz, O2) at a sampling rate of 500Hz (see Figure 2b). Both the reference and ground electrodes were placed at the left mastoid. Conductive gel was used to improve the quality of the brainwaves recording.



(a) Subjects placed the head on a chin rest situated 50cm from the screen. The eye tracker was attached to the lower part of the display.



(b) EEG electrodes placement according to the international 10-20 system.

Figure 2: Experimental setup.

4.2.3 Computer display. We used a 27" LCD monitor with 240Hz of refresh rate. The active area to render stimuli was reduced to a 24" display to match Tobii recommendations of optimal display size. Hence, we defined an active rectangle that corresponded to a 24" monitor with aspect ratio 16:9 with dimensions 530 x 300 mm.

4.3 Data Acquisition

A software application was developed in Matlab to generate the visual stimuli and collect the gaze data. We utilized the Tobii SDK [73] to record the gaze data and the Neuroelectrics Instrument Controller (NIC2) [59] to record the EEG data. Gaze and EEG datastreams were synchronized using Lab Streaming Layer (LSL) [48] markers sent to NIC2 through Matlab.

Every subject was calibrated using Tobii's calibration routine available in the SDK. For best results, we utilized 9 calibration points spanning across the active tracking area, which is the maximum number of calibration dots recommended by Tobii. More than 9 calibration points doesn't help improve the results, according to the manufacturer. After calibration, a calibration validation was performed to estimate the performance of the eye tracker. For the validation, a new set of stimuli points was presented to the participant, and we computed values for accuracy and precision based on the position of the gaze data in relation to the stimuli points. We repeated the full calibration procedure if accuracy and precision RMS of either right or left eye were above 1.5° and 0.5° respectively.

4.4 Stimuli Design

We chose to use a grid of hexagon-shaped stimuli as an exemplification of what could be a real use case of the screen partitioning approach (see Section 3). The reasoning behind this is that a hexagon grid, as opposed to other shapes such as squares, can maintain a constant distance between the centers of all adjacent hexagons. This allows for a consistent inter-stimulus distance across all stimuli. A similar approach was taken in [61], where the effects of competing stimuli on the evoked SSVEP were studied using a grid of hexagons.

The stimuli consists of a grid of 7 hexagons, one in the center and the rest circling around, aligned with the 6 sides of the central one (see Figure 3). We studied grids of different sizes and distances between hexagons. Hexagon diameters (d in Figure 3) were [1°, 2°, 3°, 4°] visual angles and the offsets between hexagons (s in Figure 3) were [0°, 0.5°, 1°, 2°]. We generated stimuli using all combinations of sizes and offsets, resulting in a total of 16 different grids.

Hexagons were colored white and each one flashed at a different frequency, ranging from 40Hz to 46Hz in steps of 1Hz. To generate the different frequencies, we modulated the luminance using the sampled sinusoidal method [50], which allows to create stimuli signals at any frequency up to half the refresh rate of the computer monitor. The expression that represents the modulated luminance is the following:

$$L(f_{st}, k) = 0.5sin(2\pi f_{st}(k/RR)) + 0.5 \quad (1)$$

where f_{st} is the frequency of the visual stimulus, k is an integer corresponding to the frame index and RR is the refresh rate of the display. Hence, $L(f_{st}, k)$ represents the luminance value of a sampled sine of frequency f_{st} with a sampling rate of RRHz. The dynamic range of the wave is from 0 to 1, where 0 represents dark and 1 maximum luminance.

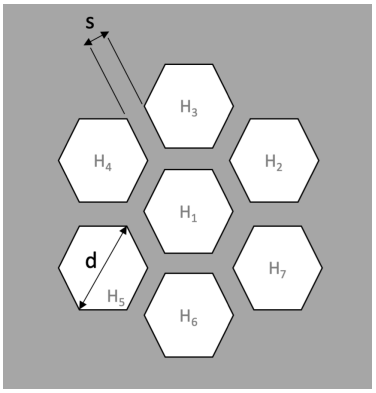
To study the effect of gaze location on the eye-tracking data quality, we rendered the stimuli in different positions across the screen. Figure 4 shows a schematic of the coordinates where the central hexagon of the grid was placed. The stimulation software was written in Matlab using the Psychophysics Toolbox [9], which interfaced directly with Tobii SDK and LSL.

4.5 Experimental Procedure

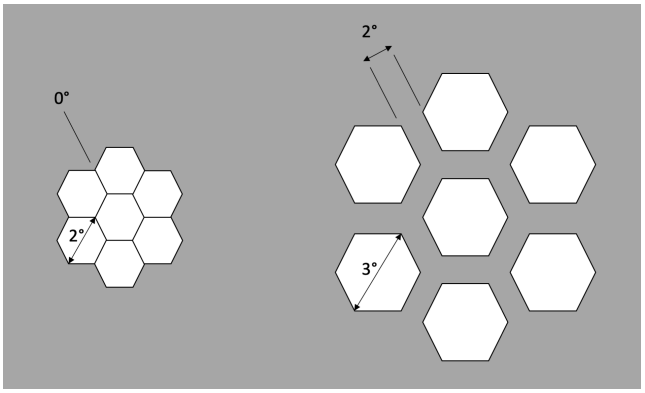
The experiment was divided in 4 main blocks: calibration, gaze model training, EEG model training and main experiment.

4.5.1 Calibration. The session started with the calibration of the eye-tracker as explained in Section 4.3. Then, in order to keep a record of the quality of the EEG recording, baseline alpha brainwaves were monitored during 10 seconds while the subject closed their eyes.

4.5.2 Gaze model training. The goal of this block was to gain a prior knowledge about the behavior of the eye-tracker across the screen. This prior would later be used in the Bayesian inference



(a) General representation of stimuli, defined by two parameters. Each hexagon was assigned an index as shown. H_1 was the target gazed by the user during the whole experiment.



(b) Two examples of stimuli used in the experiments (not in scale)

Figure 3: Grid of hexagons used for visual stimuli. In 3a we show the two parameters that completely define a stimulus. d is the length of the hexagons diagonal. s is the spacing or offset between hexagons in the grid. Both dimensions are in visual angles. d ranged from 1 to 4 in steps of 1 visual angles. s took the values 0, 0.5, 1 and 2 visual angles.

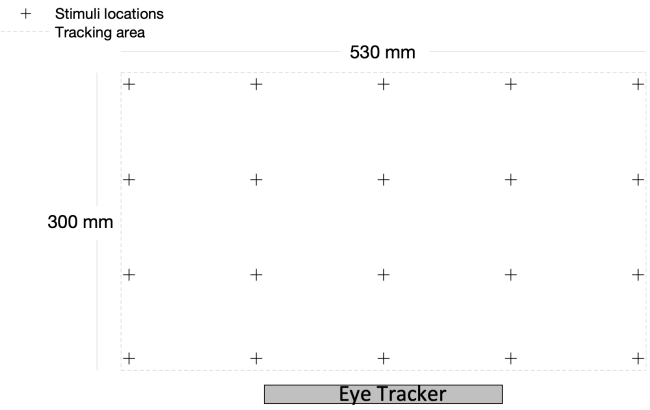


Figure 4: Coordinates where the stimuli (hexagon grids) were rendered during the main experiment. The eye-tracker was placed below the screen.

model to estimate the hexagon targeted by the subject together with the EEG data. Hence, this block consisted of sequentially displaying a total of 32 dots that covered the whole screen (see Figure 7). The subject received the instruction to gaze at each point during 5 seconds while gaze data was being recorded. Subjects were asked to not blink nor move during the 5 seconds. There was a rest period of 3 seconds between trials where subjects could blink. The total time required to complete this block was 5 minutes.

4.5.3 EEG model training. This session consisted on obtaining a prior knowledge about the targeted hexagon classification accuracy performed by the EEG system. EEG data was recorded while subjects gazed at the central hexagon of the grid. We chose the target to be the central hexagon since this would represent the worse-case scenario in terms of accuracy and it mimics a real use case where all hexagons would be surrounded by other targets. Each one of the 16 possible grid configurations (4 hexagon diameters and 4 spacings) was shown in seven 5-second trials. The flashing frequency of the central hexagon was different every trial, so it flashed at the 7 different frequencies (40-46Hz) throughout the 7 trials. The frequencies of the surrounding hexagons were chosen randomly every trial from the 6 remaining ones. We repeated this for all 16 grid configurations. The location of the grid in the screen was randomly chosen every trial. This way, we could obtain an average performance of the EEG system for each configuration. Subjects were asked to avoid blinking during the 5 seconds of stimulation. There was a rest time of 3 seconds between trials and an unlimited rest time every 16 trials. The time required to complete this block was around 15 minutes.

4.5.4 Main experiment. The main experiment consisted of recording simultaneously EEG and gaze data while subjects gazed at the central hexagon of the grid. All 16 grid configurations were sequentially shown during 5 seconds at every predefined screen location once (see Section 4.4 and Figure 4). Flashing frequencies of the 7 hexagons were randomly assigned every trial as well as grid parameters and locations. Subjects were asked to avoid blinking during the 5 second-trials, but they could rest and blink during the inter-trial 3 seconds. Subjects never received any type of real-time feedback during the experiment. There was an unlimited rest time every 16 trials. The total time required to complete this block was around 50 minutes.

4.6 Target estimation from eye-tracker

Epochs of different lengths were extracted using a moving window with steps of 200ms from the 5-second trials, starting 400ms after target onset to account for the time it took to the user to move the point of regard to the target. Epoch lengths were 0.5 to 5 seconds in steps of 0.5 seconds.

Tobii delivers binocular data. Hence, to obtain the estimated gaze point, we computed the midpoint between the left and right eye gaze points, since the calibration of the Tobii Pro Nano is binocular, and it's a common procedure among the eye-tracking research community [14, 31].

From each epoch of gaze data, we extracted the point of fixation by averaging all datapoints. Given a point of fixation, we determined that the targeted hexagon by the user was the one whose center was closest to the point of regard. Hence, using the eye-tracking system we can estimate the targeted hexagon for each data epoch. See Figure 5 for a graphical example of this process.

4.7 Target estimation from EEG

Epochs with the same start and end time as in the gaze data were obtained for the EEG data. Each epoch was filtered with a 4-order Butterworth band-pass filtered between 5 and 100Hz and a 60Hz notch filter.

Since the 7 hexagons flash at different frequencies, components at each frequency will be present on the elicited SSVEPs. Therefore, we need to extract the most dominant evoked response to determine which hexagon the subject is attending. To do so, we used canonical correlation analysis (CCA), which is a multivariate statistical method widely used in the detection of SSVEPs [5, 46]. From each epoch we estimated the hexagon attended by the user by selecting the frequency with highest correlation with the EEG channels.

4.8 Bayesian Integration of EEG and Gaze data

In order to efficiently combine the complementary strengths of the two sources of information, we chose to follow a probabilistic Bayesian approach. Hence, we want to estimate the posterior probability of the user gazing at the hexagon H_i given all the information available. According to Bayes' rule:

$$P(H_i \mid E_k, T_j, p, s, d, f_i) = \frac{P(E \mid H_i, T_j, p, s, d, f_i) P(H_i \mid T_j, p, s, d, f_i)}{P(E_k \mid T_j, p, s, d, f_i)} \tag{2}$$

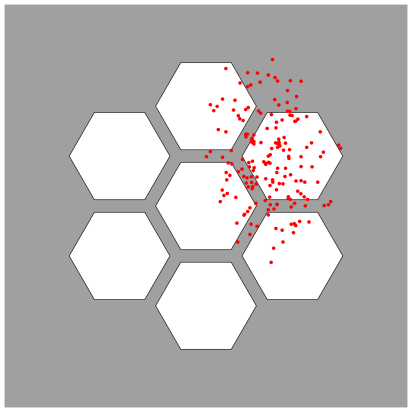
where:

H_i = hexagon gazed at by user; $i = 1, 2, \ldots, 7$
E_k = gazed hexagon estimated by EEG; $k = 1, 2, \ldots, 7$
T_j = gazed hexagon estimated by eye-tracker; $j = 1, 2, \ldots, 7$
p = position of hexagon grid in the screen
s = spacing between hexagons in the grid
d = diameter of hexagons
f_i = flashing frequency of hexagon H_i

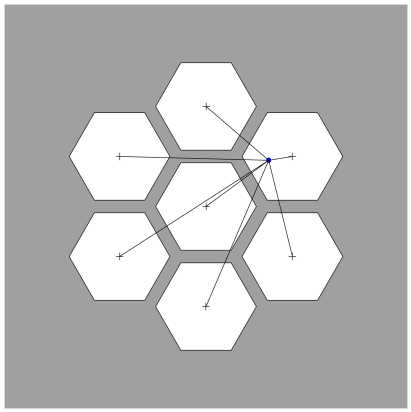
Applying Bayes' rule again on $P(H_i \mid T_j, p, s, d, f_i)$ we find:

$$P(H_i \mid E_k, T_j, p, s, d, f_i) = \frac{P(E_k \mid H_i, T_j, p, s, d, f_i) P(T_j \mid H_i, p, s, d, f_i) P(H_i \mid p, s, d, f_i)}{P(E_k \mid T_j, p, s, d, f_i) P(T_j \mid p, s, d, f_i)} \tag{3}$$

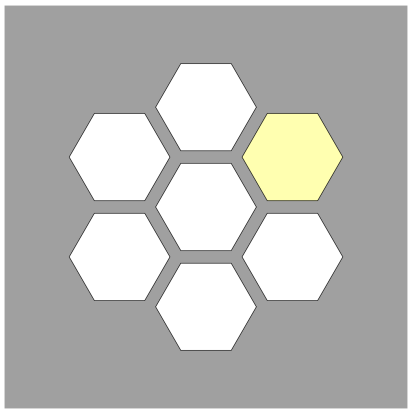
Given this expression, we can realize some simplifications. Firstly, since the 7 possible frequencies of the system are close enough, we



(a) Gaze points from a 3-second epoch.



(b) Fixation point computed from the average of the gaze points in 5a. Distances from the fixation point to hexagon centers are shown.



(c) Estimated targeted hexagon by user highlighted in yellow.

Figure 5: Visualization of the process of estimating the hexagon that the user is gazing at from the gaze points in an epoch.

assume for simplicity that the power of the elicited SSVEPs is not dependent on the stimulating frequency, i.e. $P(E_k \mid H_i, T_j, p, s, d, f_i) \approx P(E_k \mid H_i, T_j, p, s, d)$. Moreover, the eye-tracking system is independent from the flashing frequencies as well, i.e. $P(T_j \mid H_i, p, s, d, f_i) = P(T_j \mid H_i, p, s, d)$. Hence, we can remove the dependency on frequency for the whole Bayesian system.

Secondly, regardless of the screen location of the stimuli, the targeted hexagon will always fall in the center of the users' fovea. Therefore, even though stimuli at the screen corners will be marginally further away to the subject's eyes than stimuli at the screen center, the effect on the evoked SSVEP will be minimal. Consequently, the dependency of the EEG model on the stimuli location can be also removed, i.e. $P(E_k \mid H_i, T_j, p, s, d) \approx P(E_k \mid H_i, T_j, s, d)$.

Thirdly, to mimic a real world application, we assume that all 7 targets are equally likely for the user to gaze, i.e. $P(H_i \mid p, s, d) = P(H_i) = 1/7$ for all i.

Fourthly, since the workload that we could ask from subjects was limited, there was not enough data gathered during the training blocks to obtain the joint conditional probabilities $P(E_k \mid H_i, T_j, p, s, d, f_i)$. Hence, we did the approximation of assuming E_k and T_j are independent, so $P(E_k \mid H_i, T_j, s, d) \approx P(E_k \mid H_i, s, d)$.

Finally, the denominator can be seen as a normalization term, which can be omitted if the objective is to simply compare the posterior probabilities.

Given these considerations, the expression for the posterior probability of the Bayesian system results as follows:

$$P(H_i \mid E_k, T_j, p, s, d) \propto P(E_k \mid H_i, s, d)\ P(T_j \mid H_i, p, s, d) \quad for \quad i, j, k = 1, 2, ..., 7 \tag{4}$$

where $P(H_i \mid E_k, T_j, p, s, d)$ is the *posterior* probability of the user gazing at hexagon H_i given the information from the EEG and eye-tracking systems as well as the properties of the hexagon grid. The $1/7$ corresponding to $P(H_i)$ has been omitted due to the proportional relationship. The 7 hexagons of the grid were indexed as shown in in Figure 3a.

4.8.1 Computing likelihoods of eye-tracking system. In order to compute the posterior likelihood of the Bayesian system, we need to calculate $P(T_j \mid H_i, p, s, d)$, which is the probability that the eye-tracking system outputs T_j as the targeted hexagon given the hexagon grid properties and that the gazed hexagon by user is H_i.

For this, we will take advantage of the data obtained during the gaze model training of the experimental procedure (see Section 4.5.2). This allowed us to obtain a prior knowledge of the eye-tracker behavior at several locations across the screen. Assuming that the captured gaze points are normally distributed, we can fit a two-dimensional Gaussian to the data points at each location. Each Gaussian presents a mean (μ_x, μ_y) with offsets O_x, O_y with respect to the target and a covariance Σ. See Figure 6 for a graphical example.

Since we had only previously collected training gaze data at limited screen locations, there was the need to estimate the behavior of the eye-tracker (mean offsets O_y and O_x and covariance Σ) at arbitrary screen coordinates. Hence, we interpolated the properties of the Gaussian distributions doing a weighted-averaging of the K-nearest-neighbor Gaussians. This approach guarantees that the

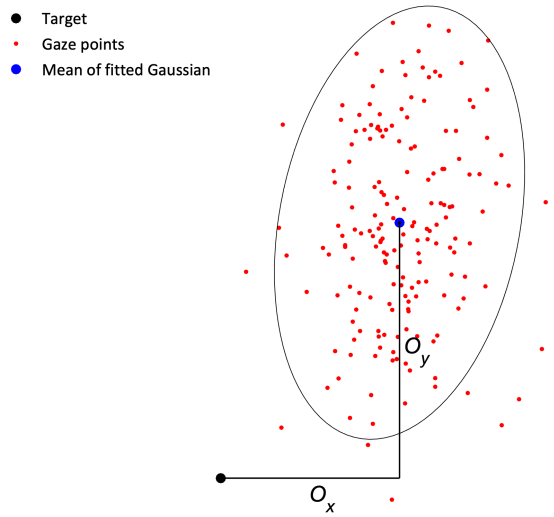


Figure 6: A 2D Gaussian can be fitted to the gaze points delivered by the eye-tracker. The Gaussian presents an offset O_x, O_y with respect to the target. An iso-contour of the fitted Gaussian distribution is represented here with a 95% error ellipse.

estimated covariance matrix remains positive definite, since the sum of positive definite matrices is also positive definite. Therefore, given an arbitrary screen location gazed by the user, we can estimate the offsets O_y and O_x and the covariance matrix Σ that will fully define the predicted Gaussian distribution of the gaze points delivered by the tracker (see Equations 5, 6, 7 and Figure 7).

$$\hat{\Sigma} = \frac{\sum_{n=1}^{K} \Sigma_n / d_n}{\sum_{n=1}^{K} 1/d_n} \tag{5}$$

$$\hat{O}_x = \frac{\sum_{n=1}^{K} O_{xn} / d_n}{\sum_{n=1}^{K} 1/d_n} \tag{6}$$

$$\hat{O}_y = \frac{\sum_{n=1}^{K} O_{y_n} / d_n}{\sum_{n=1}^{K} 1/d_n} \tag{7}$$

where:

$\hat{\Sigma}, \hat{O}_x, \hat{O}_y$ = estimated covariance matrix and offsets of the Gaussian distribution of gaze points at an arbitrary screen location

$\Sigma_n, O_{xn}, O_{y_n}$ = covariance matrix and offsets of the fitted Gaussian to collected gaze points at neighbor location n

d_n = distance between the queried arbitrary screen location and the neighbor target of the fitted Gaussian n

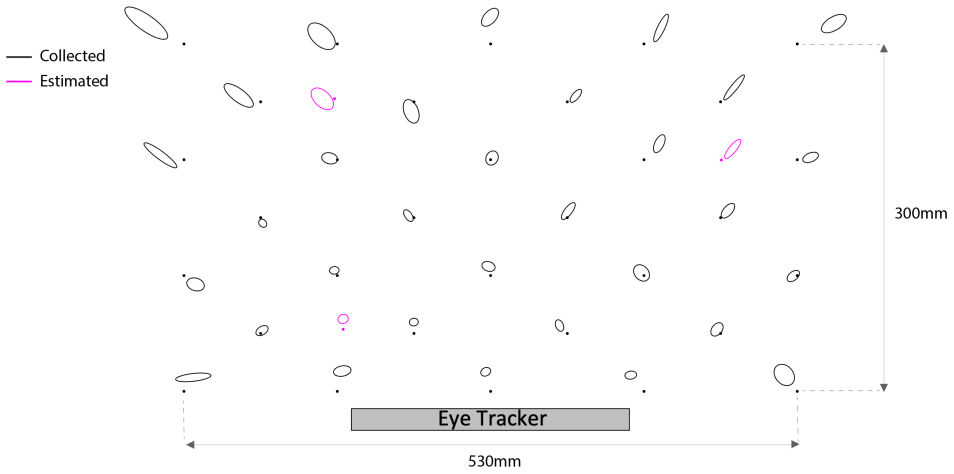


Figure 7: Gaze targets (dots) and Gaussians (95% error ellipses) associated to each target. Black points represent the targets shown to subjects during gaze model training. Black ellipses represent fitted Gaussians to the collected gaze points during training. Pink points and ellipses represent arbitrary examples of queried targets and estimated Gaussians using weighted KNN. Data for this plot corresponds to subject 5.

Finally, to compute $P(T_j \mid H_i, p, s, d)$ we first apply the method described above to estimate a Gaussian using the center of the hexagon H_i as the targeted point in the screen. Then, using Monte Carlo sampling we obtain an approximate probability of the hexagon outputted by the eye-tracking system being T_j, given the estimated Gaussian distribution of gaze points and the known hexagon grid dimensions.

4.8.2 Computing likelihoods of EEG system. In this section, we explain the computation of $P(E_k \mid H_i, s, d)$, which is the probability that the EEG system outputs E_k as the targeted hexagon given the hexagon grid properties and that the gazed hexagon by user is H_i. During the EEG training block experiment (Section 4.5.3), the user was asked to gaze at the central hexagon (H_1) for all the different hexagon grid configurations. Hence, we can compute $P(E_k \mid H_1, s, d)$ for all k. We found out that $P(E_k \mid H_1, s, d)$ for $k \neq 1$ are very similar for all $k \neq 1$ (equally likely). Hence, we will assume $P(E_k \mid H_1, s, d) = \frac{1-P(E_1|H_1,s,d)}{6}$ for $k \neq 1$. Similarly, we will assume $P(E_k \mid H_i, s, d) = P(E_1 \mid H_1, s, d)$ for $k = i$, since we did not collect data to compute $P(E_k \mid H_i, s, d)$ for $i \neq 1$ (user always targeted H_1).

4.8.3 Inference of the targeted hexagon. To estimate the hexagon gazed at by the user, we need to compute the posterior likelihood for all 7 hexagons H_i and choose the target that shows highest probability.

Hence, for a given epoch of EEG and gaze data, we obtain the estimates from both systems individually, E_k and T_j, and then we compute the posterior probabilities assuming the user is gazing each hexagon H_i, $i = 1, 2, \ldots, 7$. Finally, the estimated target by the Bayesian system is as follows:

$$\hat{H} = \arg\max_{H_i} P(H_i \mid E_k, T_j, p, s, d) \tag{8}$$

5 RESULTS

5.1 Likelihoods of EEG prior model

We applied the EEG target estimation on the training data obtained during the EEG model training session (see Section 4.5.3). This allowed us to obtain the likelihoods of correctly or wrongly classifying the gazed target for each hexagon grid configuration and for different epoch lengths. The probabilities for all target parameters (diameters and spacings) and for 3-second epochs, averaged across all subjects are shown in Figure 8. Same process was followed to obtain probabilities for different epoch lengths.

As seen, the probabilities of correctly classifying the target increase with target size d and spacing s. The reasoning for the first one is that the more region of the visual field is stimulated by the attended stimuli, the stronger is the corresponding evoked cortical response and the easier it is to discriminate it from the other non-attended stimuli responses [62, 72]. The reasoning for the latter is that there is a limited amount of neural resources to process visual stimuli, as has been shown by previous studies [11]. Hence, multiple stimuli in the visual field will compete for neural representation, which may suppress the dominant frequency response and hinder the detection of the attended stimulus depending on the inter-stimuli distance [24, 61].

Since the EEG system can perform differently on each subject (see standard error in Figure 8), we used each subject's individual probabilities to compute the posterior probabilities of the Bayesian system. The higher the performance of the EEG system is for a specific subject, the more the Bayesian system will rely on it to estimate the gazed target, and vice versa.

5.2 Performance of the Hybrid Bayesian System

In this section, we report the performance of the hybrid system for different target parameters, screen positions and epoch lengths,

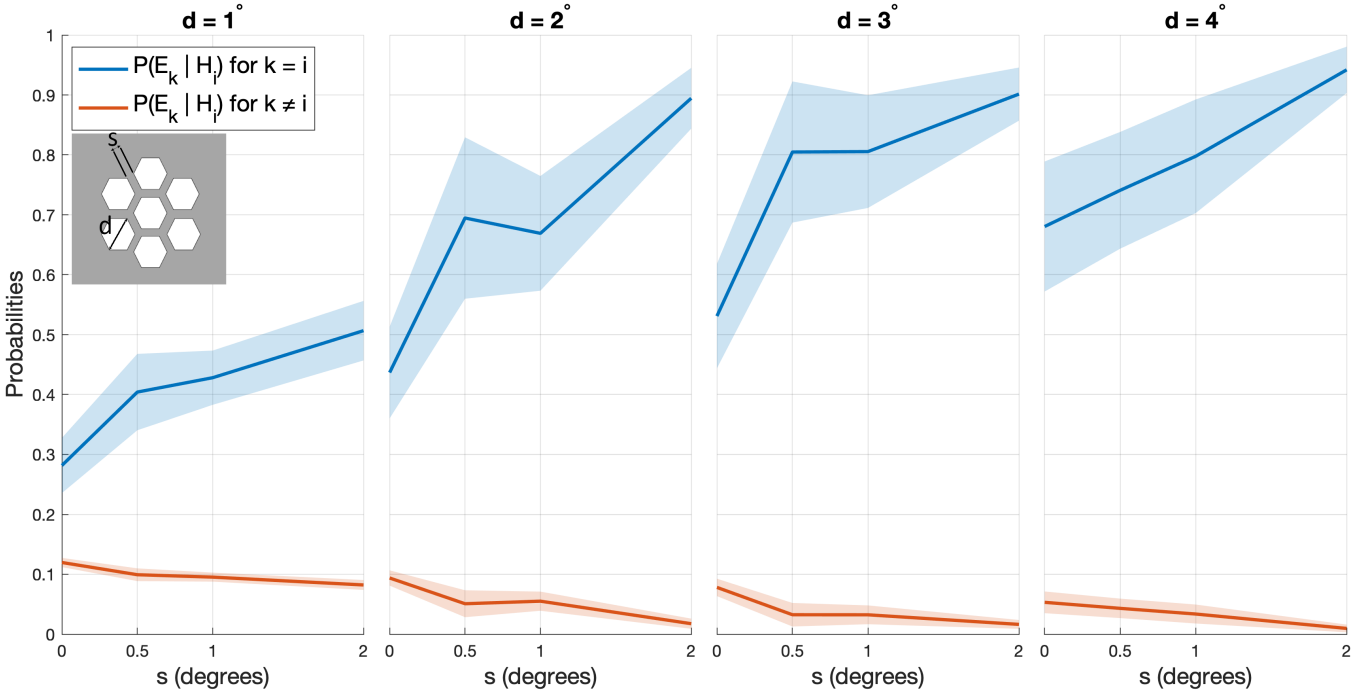


Figure 8: Probabilities that the EEG system outputs a target gazed by the user (blue) and a target not gazed by the user (orange), as a function of target diameter d and target spacing s. These values were obtained using 3-second epochs. Solid lines represent the probabilities averaged across all subjects. Colored shadows indicate standard error.

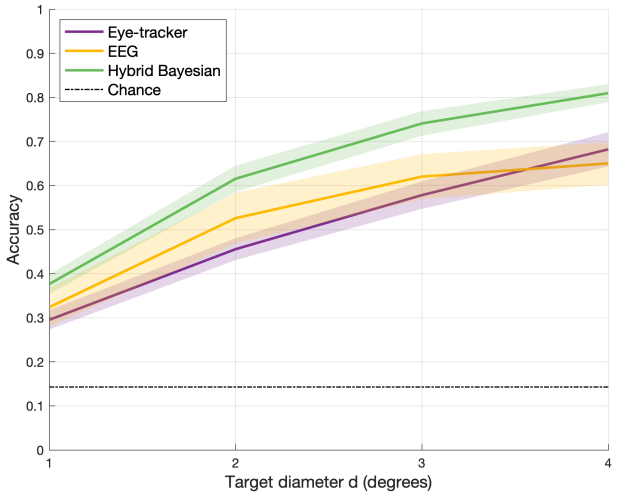
while comparing it to the accuracy of each system alone. Accuracy values are obtained by computing the ratio of correctly over wrongly classified targets across all epochs. Accuracy of the eye-tracking system alone is computed by selecting the target closest to the averaged gaze point obtained from each epoch, as explained in Section 4.6. Accuracy of the EEG system alone is computed by selecting the target whose frequency presents the highest correlation to the EEG channels using CCA method, as explained in Section 4.7.

5.2.1 Varying hexagon grid parameters. Here we examine the systems' performance for different target diameters and spacings. In order to simplify the analysis, we chose a constant epoch length of 3 seconds for the plots. We consider this length can show representative results for the different approaches. A report of the performance as a function of epoch lengths is done in Section 5.2.3.

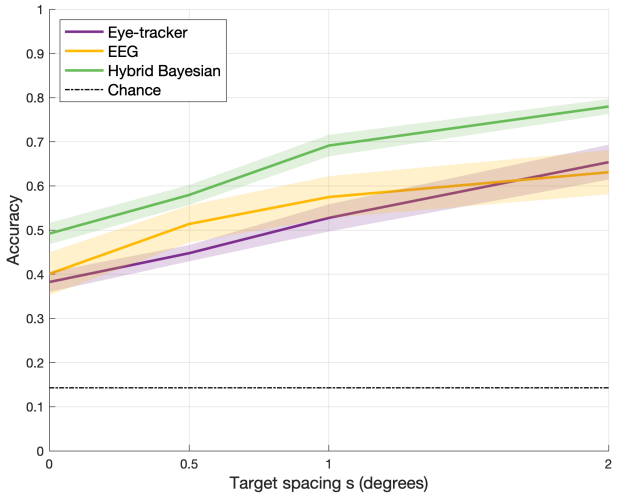
In Figure 9 we show the accuracies for all 3 systems as a function of targets diameter and spacing. Values are obtained by computing the accuracies across all screen positions. We can see how the hybrid Bayesian system outperforms the independent EEG and eye-tracker systems for all target parameters. A statistical analysis was conducted to determine whether the accuracy differences between the hybrid and eye-tracking-only systems are significant. We used the one-sample t-test for μ, the difference of the means of two populations $\mu_1 - \mu_2$, with null hypothesis that the data comes from a normal distribution with zero mean and unknown variance. The μ_1 data corresponds to a vector containing the accuracy values of all subjects for a particular hexagon grid parameter using the hybrid approach, and μ_2 to the accuracies from only the eye-tracker system. Hence, rejecting the null hypothesis means that the performance accuracies between systems (hybrid and eye-tracker-only) are different because μ is not equal to zero. The test was repeated for all hexagon grid parameters (diameters and spacings) and all showed statistical significance ($p < 0.05$). These results suggest that there are statistically significant differences between both approaches.

5.2.2 Performance across screen positions. We analyzed the system performance as a function of the screen regions. In order to visualize what locations our hybrid system had a greater impact on, we computed the accuracy difference between hybrid and eye-tracking systems across all target parameters at each screen coordinate and for each subject. In Figure 10 we show a 2D heat map with the percentage increase of accuracy across the screen. We can see that the largest performance enhancement takes place at the peripheral screen regions. This was expected since the systematic error in eye-tracking systems becomes more pronounced for larger angles between the fixation point and the camera [29]. Some screen corners show accuracy increments of over 1000% since the eye-tracking system performed poorly and the percentage increase boosts. The most central regions show null or slightly negative improvement.

5.2.3 Varying epoch length. Speed is an important element to take into account in gaze monitoring systems and computer interfaces. Hence, we studied the effect of the epoch length on the systems' performance. Shorter epoch lengths implies that a system can respond faster to a change of condition, in this case, to a change of



(a) Accuracies as a function of target diameter.



(b) Accuracies as a function of target spacings.

Figure 9: Performance of the three systems as a function of hexagon grid parameters. Values were computed through all screen positions, using epochs of 3 seconds. Solid lines represent accuracies averaged across all subjects, shadows represent standard errors.

region or element attended. Figure 11 reports the performance of all three systems as a function of epoch lengths for all the target parameters, averaged across all subjects. On the one hand, we can see that the performance of the eye-tracker is independent from the epoch length, since its accuracy is constant in all plots. This was expected since the sampling rate of the eye-tracker is 60Hz, which is too fast for the considered epoch lengths to have an effect on the tracker performance. On the other hand, the EEG system is indeed dependent on epoch length, since longer EEG data enhances the SNR of the SSVEPs. As a result, our hybrid system shows an increasing accuracy with epoch length as well. Interestingly, the performance of the hybrid system is equal or better than any of the two individual systems for all epoch lengths and target parameters, confirming the successful Bayesian integration of the two data modalities. Even in the shortest epoch lengths, where the performance of the EEG system is very poor in comparison to the eye-tracker, the Bayesian system is robust enough to at least match or even outperform the eye-tracker, which is the result of exploiting the likelihoods of the prior models obtained during the training session.

5.2.4 Contribution of each individual system. In this section, we explore the contribution of each system to the output of the hybrid approach for different conditions. We want to analyze whether the accuracy enhancement of the hybrid system is due to the prior model from a single system alone, or whether it is the result of a symbiotic association of both models. To do so, we modified the probability terms of the EEG and eye-tracking systems to include a new parameter α_E and α_T. These parameters regulate the extent to which EEG or eye-tracking data is weighted in the hybrid Bayesian system:

$$P_{mod}(E_k \mid H_i, s, d) = \frac{\alpha_E}{7} + (1 - \alpha_E)P(E_k \mid H_i, s, d) \quad (9)$$

$$P_{mod}(T_j \mid H_i, p, s, d) = \frac{\alpha_T}{7} + (1 - \alpha_T)P(T_j \mid H_i, p, s, d) \quad (10)$$

When α_E is 1, the likelihood given by the EEG system is always $1/7$, so it becomes a random classifier and the Bayesian system relies exclusively on eye-tracking data. As α_E decreases, the hybrid system increasingly gives weight to EEG data. The reasoning is parallel for α_T and the eye-tracking system. Hence, the posterior probability of the Bayesian system with the modified prior terms is as follows:

$$P_{mod}(H_i \mid E_k, T_j, p, s, d) \propto P_{mod}(E_k \mid H_i, s, d)\ P_{mod}(T_j \mid H_i, p, s, d) \quad (11)$$

To quantify how much each system's prior contributes to the hybrid output, we vary the parameter α_i associated to that system from 0 to 1, while maintaining the other system fully engaged ($\alpha_k = 0$). If the hybrid accuracy doesn't change with α_i, it implies that the Bayesian system is not relying on that system's data to compute the estimated gazed target, hence it is dispensable.

First, we study the contribution of each system's priors for medium length epochs (3 seconds). A relevant initial analysis is to compute the overall hybrid system accuracy, aggregating all target parameters and screen positions accuracies. The result is shown in Figure 12. We can see that both systems contribute similarly to the overall accuracy, since the variation from $\alpha_i = 0$ to $\alpha_i = 1$ results on comparable accuracy change.

More interesting findings can be observed by doing similar analysis for specific target parameters, screen positions or epoch lengths. In Figure 13, we show the same study done for different target diameters and spacings. We can see that the accuracy improvement when each system is incorporated in the Bayesian model is different depending on the system, the target diameter or spacing. For example, when targets are very small ($d = 1$ and $d = 2$), the performance enhancement when the EEG system chimes in (dark green) is larger than when the eye-tracking system does. This may be related to the fact that the Gaussians used to compute the probabilities of the eye-tracking system are not as informative for smaller targets

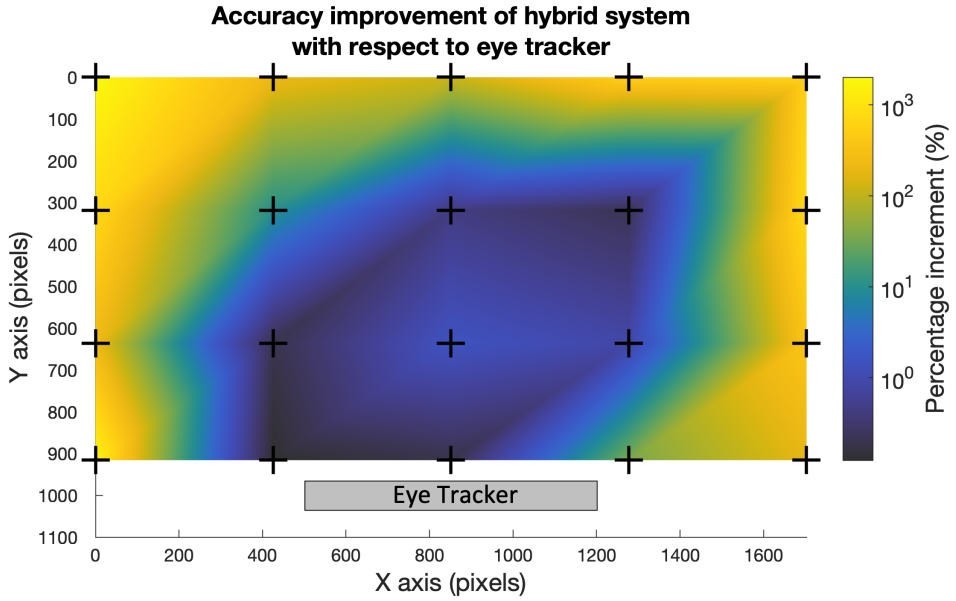


Figure 10: Accuracy percentage increase of hybrid Bayesian system with respect to the eye-tracker as a function of screen position. Accuracy values were computed across all target parameters using 3-second epochs. Cross markers indicate the locations where stimuli was shown and where accuracy values were computed. The coloring for the rest of the screen were obtained by interpolating the values at the markers. Scale is logarithmic. Plots show the averaged values across all subjects.

due to the precision and systematic errors. For larger targets, the accuracy improvement is comparable for both systems. We can view this effect more clearly if we plot the accuracy change (from $\alpha_i = 1$ to $\alpha_i = 0$) for each system and target parameters. This is shown in Figure 14.

We can also do the same analysis for different screen locations. Figure 15 shows the results for four different representative locations, two central and two peripheral. We can clearly see that the behavior is opposite for peripheral and central screen positions. In the former (two plots on the left), the effect of transforming the eye-tracking system from a random classifier ($\alpha_T = 1$) to an intelligent one that exploits the prior probabilities ($\alpha_T = 0$) has no effect on the accuracy of the hybrid system, since it is constant for all α_T (light green). On the contrary, the contribution of the EEG prior likelihoods has a large impact on the Bayesian system performance for these screen regions, since accuracy improves as α_E decreases (dark green). The behavior is the opposite for central screen regions (two plots on the right). In Figure 16 we plot the difference of hybrid system accuracy between $\alpha_i = 0$ and $\alpha_i = 1$ as a function of screen position to see what regions each system has a larger effect when they engage in the gazed target inference. We can see that the observed for two pair of points, generalizes to the whole screen. This implies that for peripheral screen regions, on average, the hybrid system can perform similarly exploiting the EEG data alone, and for central regions the EEG data is less crucial.

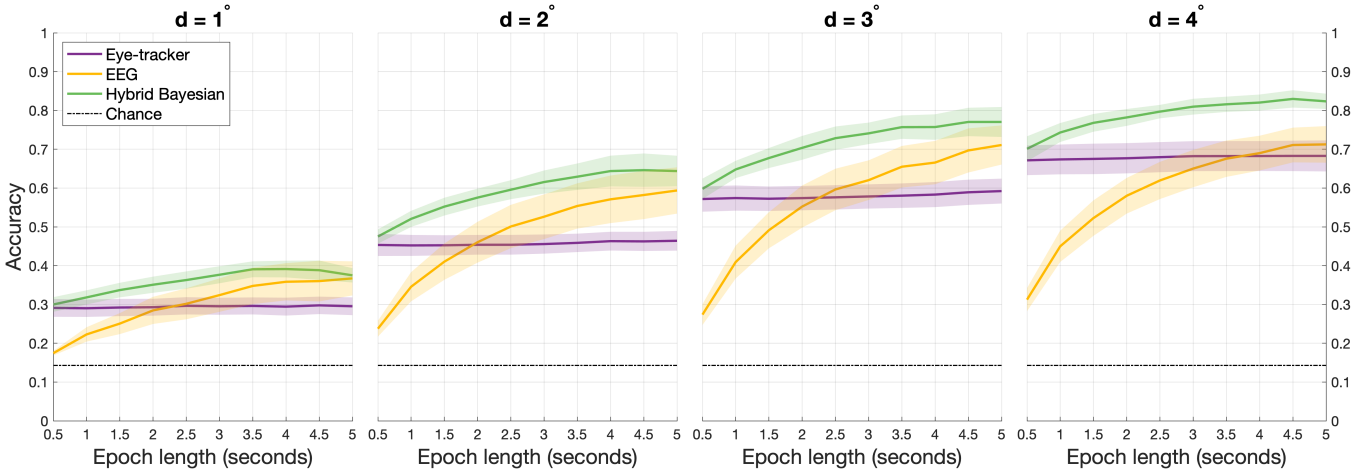
Finally, we performed the same analysis as a function of epoch length. In Figure 17 we show the results for 4 different epoch lengths (0.5, 2, 3.5 and 5 seconds). We can see that for the shortest epoch length, the hybrid system resorts more to the eye-tracking data, since the accuracy increment when α_T goes from 1 to 0 is much larger than when α_E does. This makes sense since the performance of the EEG system for short epoch lengths is low. As the epoch length increases, the system progressively weights in the EEG data. For longer epoch lengths, the behavior is reversed, and the EEG system contributes more than the eye-tracker to the overall accuracy. This can be better visualized by plotting the accuracy increment when each individual system chimes in as a function of the epoch length, as shown in Figure 18.

6 DISCUSSION

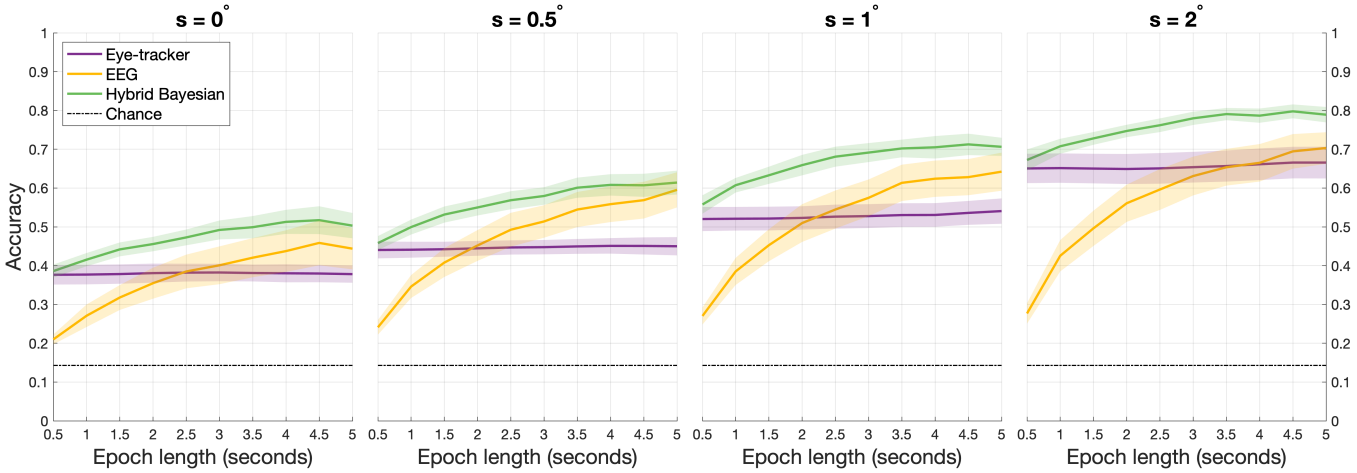
In this section we summarize the findings of this work, explain the reasoning behind the results and present the limitations with proposed solutions.

6.1 Findings

6.1.1 Accuracy limitations of the eye-tracker. The eye-tracker manufacturer reports an accuracy of 0.3° of visual angle (at optimal conditions). The smallest distance between the center of the target hexagon and a non-target hexagon throughout all the experiments was ~0.4° (for $d = 1°$ and $s = 0°$). Hence, theoretically, and according to the manufacturer, even for the hexagon grid with smallest targets and spacings, the eye-tracker should be able to detect the target gazed by the user without error (at optimal conditions). However, we can see that this is definitely not the case, since the average accuracy when detecting the gazed target using only the eye-tracker for $d = 1°$ is below 30%. Note that this value is obtained averaging the accuracies of all four spacings s, so for the specific case of $d = 1°$ and $s = 0°$ this accuracy would be even lower. Even for the largest



(a) System accuracies as a function of epoch length for different target diameters. Values for each target diameter were computed across all target spacings and screen positions.



(b) System accuracies as a function of epoch length for different target spacing. Values for each target spacing were computed across all target diameters and screen positions.

Figure 11: System accuracies as a function of epoch length for different target parameters. Values computed for all screen positions and averaged across all subjects. Solid lines represent averaged values, shadows indicate standard errors.

targets and spacings, the accuracy of the eye-tracker alone does not exceed 70%. Therefore, we can conclude on the one hand that the data quality values reported by manufacturers can differ considerably from the realistic results, and on the other hand that there is room for improvement exploiting data from other modalities such as the EEG.

6.1.2 Performance of the Bayesian hybrid system. Our Bayesian integration of the two data modalities resulted in an improvement of accuracy for all target sizes, spacings and epoch lengths. Averaging across all target parameters, screen locations and epoch lengths, the hybrid system showed an an accuracy increase of around 11 units with respect to the eye-tracker alone. In terms of percentage improvement, this corresponds to about 23%.

The system shows a large performance variation across the screen that we will address below, reaching an accuracy improvement of over 45 units at certain screen coordinates (or over 1000% in percentage increase). Overall, based on the numbers and graphs shown throughout this paper and supported by the results of the statistical analyses conducted, we can conclude that our approach outperforms the accuracy of the eye-tracker.

Due to the limited amount of training time, the probabilistic approach we followed is a simplified version of what could have been done if all the required data was available. Given the amount of data gathered, it was not possible to compute the conditional probabilities $P(E_k \mid H_i, T_j, p, s, d, f_i)$, which could have improved the performance of the Bayesian system since the output of the eye-tracking system can provide information to alter the output probability of the EEG system. Instead, in this work we assumed that

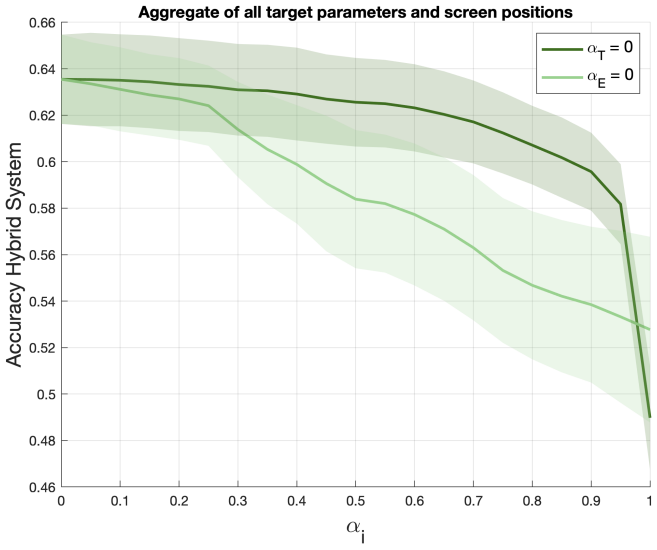


Figure 12: Accuracy of the hybrid system as a function of α_i, where $i = E, T$. When $\alpha_i = 0$, the corresponding system is fully engaged in the Bayesian computation. For $\alpha_i = 1$, the system is random and does not have an effect on the gazed target inference. Dark green plot corresponds to varying the weight of the EEG prior, while maintaining the eye-tracking system fully engaged ($\alpha_T = 0$). Light green plot corresponds to varying the weight of the eye-tracking prior, while maintaining the EEG system fully engaged ($\alpha_E = 0$). Plots obtained using 3-second epochs data. Solid lines represent the average across all subjects, shadows indicate standard error.

the outputs of the EEG and eye-tracking systems are independent. In future work, we aim to explore the performance of the system taking into account such conditional probabilities and the dependency of both variables.

6.1.3 Contribution of the prior models. Gaze data obtained during the training session allowed us to develop an eye-tracker model that gave us a prior knowledge about the tracker behavior across the screen. The Gaussians obtained at each screen location reflect the accuracy limitations of the tracker (offset between target point and Gaussian mean) as well as its precision (see Figure 7). Both metrics get worse with higher visual angles between the target and the eye-tracker. Similarly, an EEG prior model was developed from the EEG data acquired during the training session of each subject. This allowed us to obtain probabilities of correctly or wrongly classifying the gazed target for all target parameters. The likelihoods of correctly classifying the gazed target increase with the target diameter d and spacing s.

The extent to which each prior model contributes to performance of the hybrid system was quantified by transforming it to a random classifier ($\alpha_i = 1$) and increasingly weighting it in the hybrid system until it is fully engaged ($\alpha_i = 0$). The difference between the accuracies in the two extremes gave us a way to quantify the contribution of such system to the whole model.

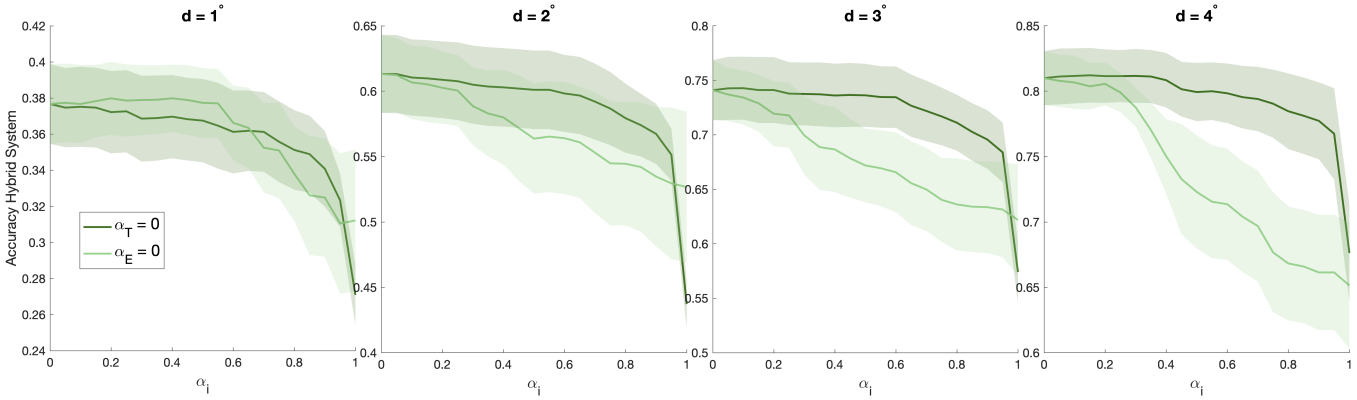
6.1.4 Large variation across the screen. An important outcome of this work is the huge dependency of the different systems' performance on the screen region. Already from the prior model of the eye-tracking system one could anticipate that the behavior would vary largely with screen coordinates. We can see this from two perspectives. On the one hand, the accuracy improvement of the hybrid system with respect to the naive eye-tracker system is much more pronounced in the peripheral screen regions than in the central ones (see Figure 10). Improvements ranging from 100% to 1000% were achieved at the external areas, whereas a zero or slightly negative improvements occurred at the central regions. On the other hand, the contribution of the eye-tracking prior model to the hybrid system accuracy was negligible or negative at the peripheral regions and very significant at the central regions –up to 80% of accuracy increase with respect to when only the EEG prior model is active (see Figure 16). The behavior of the EEG prior model was opposite. These results convey that if the application seeks to minimize accuracy errors, on average, the EEG modality will be essential in the peripheral regions, but it will be less decisive in the central regions.

6.1.5 Effect of the epoch length. An expected finding was the non-dependency of the eye-tracker performance on the epoch length. The fast sampling rate of the gaze tracker makes this system very robust to the epoch length variation. A change of the eye-tracker performance with epoch length would imply that the system suffers from an accuracy drift with time, a common problem of eye-trackers due to environment vibrations. Conversely, the EEG system was indeed dependent on epoch length, since longer EEG data enhances the SNR of the SSVEPs. We can see progressive accuracy increase with epoch length, showing an asymptotic behavior at the longest lengths. As a consequence, the hybrid system shows an increasing accuracy with epoch length as well, and with equal or better performance than any of the two systems alone for all epoch lengths and target parameters. This result demonstrates the successful application of the probabilistic approach. For the shortest epoch lengths, where the EEG cannot add much value, the hybrid system opts for resorting to the eye-tracker data. As epoch length increases, the likelihoods of correctly classifying the target with the EEG system also escalate, and therefore the hybrid system starts weighting in the EEG data. Overall, we can see similar trends of the hybrid system accuracy as a function of epoch length for all target parameters. Hence, we can conclude that our Bayesian system is robust to short epoch lengths compared to other EEG-based interfaces, where the system performance is hindered if the system speed is too large.

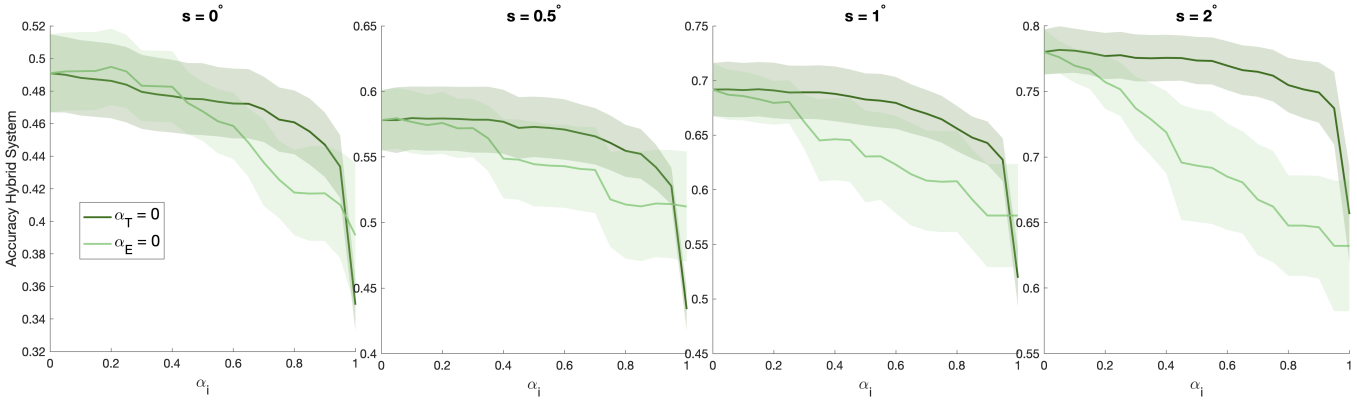
6.2 Limitations

In this section, we describe several limitations of our research, specially focusing on the barriers that separate the current work from an optimal implementation of the concept we propose in Section 3.

6.2.1 Appearance. In order for our concept to be seamlessly implemented and incorporated into any virtual scene for eye-tracking purposes, the visual flicker must be invisible. For this reason, we used high-frequency flashing ($> 40Hz$). However, the critical fusion frequency (i.e. the frequency above which a flicker is perceived as



(a) Accuracy of the hybrid system as a function of α_i, for different target diameters. For visualization purposes, y-axis scale is not the same across the plots.



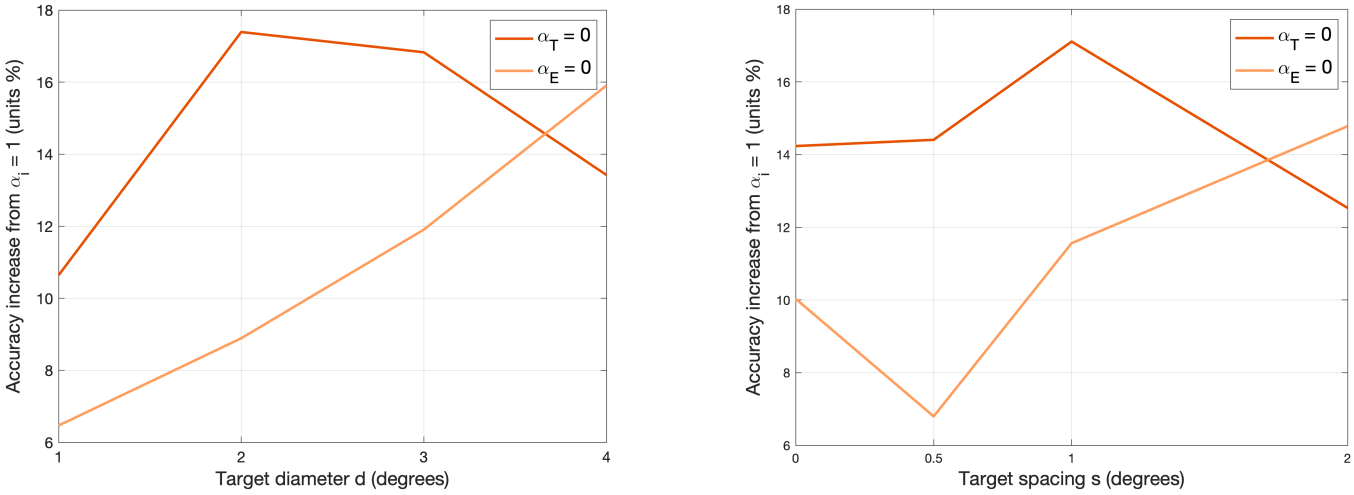
(b) Accuracy of the hybrid system as a function of α_i, for different target spacings. For visualization purposes, y-axis scale is not the same across the plots.

Figure 13: Accuracy of the hybrid system as a function of α_i, for different target diameters and spacings. The accuracy improvement when each system is incorporated the Bayesian model is different depending on the target diameter or spacing. Plots computed using 3-second epochs data. Solid lines represent the average across all subjects, shadows indicate standard error.

steady by the naked human eye) can vary, depending on several factors related to the stimuli itself (including size, luminance and location in the visual field) and the user (including age and stimuli habituation -exposure time) [2, 12, 18, 68]. Therefore, it is necessary to previously scrutinize all these variables to determine what frequencies are more suitable to use in each case, as discussed in the Future Work section.

Another critical factor is the dependency on the display refresh rate for the stimuli generation, and the subsequent emergence of visible "beatings" when rendering frequencies that are not a sub-multiple of the refresh rate. These appear when using the sampled sinusoidal technique to generate arbitrary frequencies being constrained by the refresh rate, as opposed to using an analog device such as an LED. This effect translates into an undesired perception of slowly oscillating luminance because the beatings frequency is much below the critical fusion rate. We can think of several solutions for this limitation. First, utilizing phase encoding of targets instead of frequency encoding would lead to reducing drastically the number of frequencies to use. This would permit to employ only frequencies that are sub-multiple of the refresh rate which do not lead to the beating effect. Second, reducing the dynamic range of the luminance also reduces the power of the beatings which make them harder to perceive, as reported in [3]. However, decreased signal power comes with the caveat of reduced SSVEP amplitudes, which may lead to lower SNR, lower accuracy or requiring longer detection times. Finally, displays with variable refresh rate are increasingly receiving more attention specially in the gaming industry, since they solve the screen stuttering and tearing problem by an adaptable synchronization of the display refresh rate with the GPU's render rate [69]. We think that a smart implementation of such technology may potentially allow to avoid the undesired beating effect.

6.2.2 Speed. A fundamental aspect of eye-tracking systems is the speed at which eye movements can be monitored. The sampling rate of camera-based eye-trackers is usually high (≥ 60), which allows



(a) Accuracy improvements as a function of target diameter.

(b) Accuracy improvements as a function of target spacing.

Figure 14: Accuracy improvement (difference) when each individual system shifts from a random classifier ($\alpha_i = 1$) to an intelligent one that exploits the prior probabilities ($\alpha_i = 0$), while the other system is fully engaged ($\alpha_k = 0$).

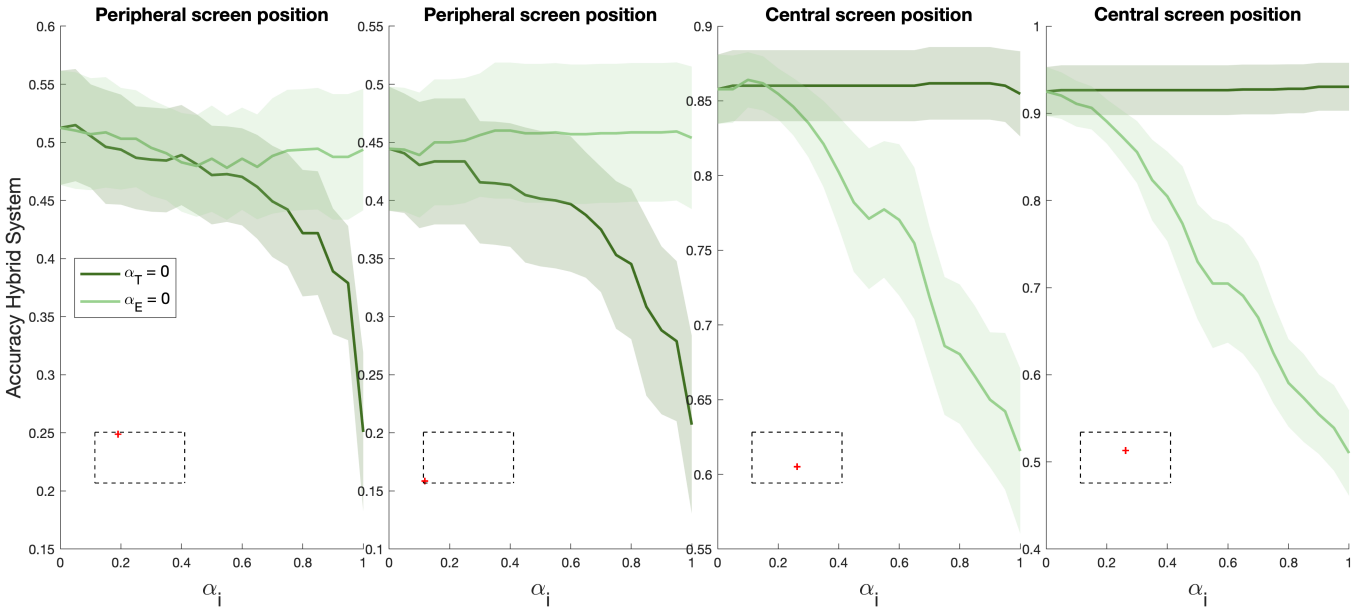
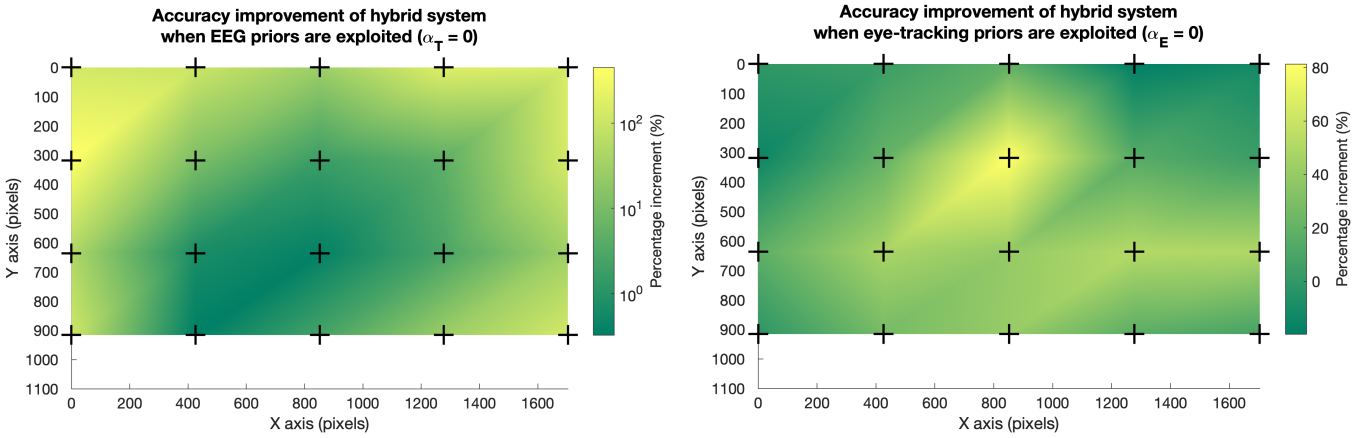


Figure 15: Accuracy of the hybrid system as a function of α_i, for different screen positions. The accuracy improvement when each system is incorporated in the Bayesian model is different depending on the screen region. Plots computed using 3-second epochs data. For visualization purposes, y-axis scale is not the same across the plots. The screen coordinates each plot corresponds to are shown with a red cross in the bottom left region of each graph. Solid lines represent the average across all subjects, shadows indicate standard error.

for a fast gaze monitoring. Many research studies take advantage of this to analyze the fast eye saccades for cognitive [47], medical [71] or reading studies [8], among others. Such type of applications would not benefit from our system due to the speed limitations of the EEG system, which originate from two different sources. Firstly, there is an inherent speed limit associated to the time required for the visual pathway to process the stimuli and elicit the SSVEP response, which ranges from 80 to 160 ms [16, 21]. And secondly, a latency is introduced due to the time it takes for the SSVEP SNR to be high enough to discriminate the targeted frequency. That is why the accuracy of the EEG system increases with epoch length. Therefore, our hybrid system would be most beneficial in applications where



(a) Percentage difference of hybrid system accuracy between $\alpha_E = 0$ and $\alpha_E = 1$, while $\alpha_T = 0$.

(b) Percentage difference of hybrid system accuracy between $\alpha_T = 0$ and $\alpha_T = 1$, while $\alpha_E = 0$.

Figure 16: Accuracy improvement (percentage) when each individual system shifts from a random classifier ($\alpha_i = 1$) to an intelligent one that exploits the prior probabilities ($\alpha_i = 0$), while the other system is fully engaged ($\alpha_k = 0$), as a function of screen position.

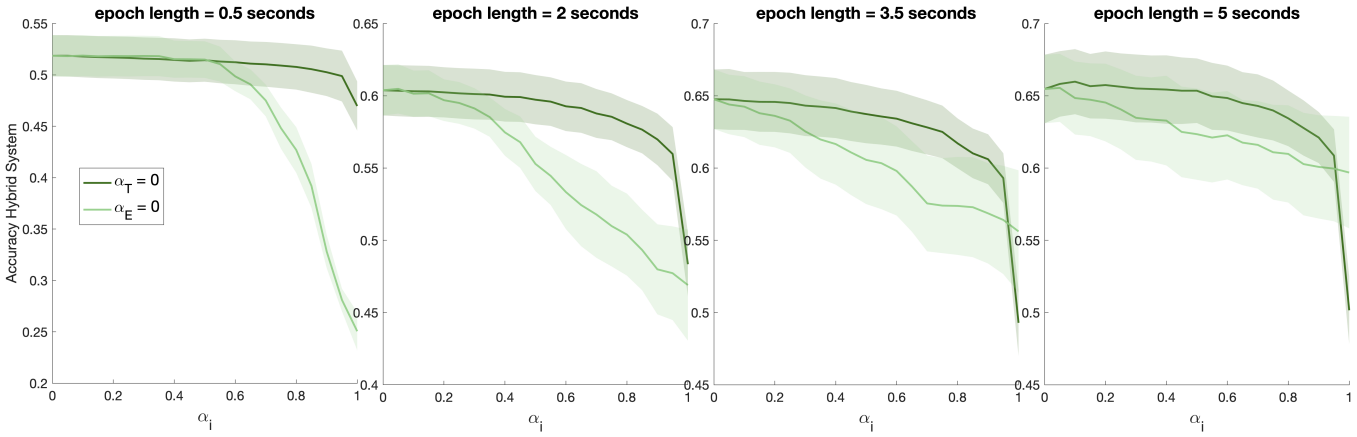


Figure 17: Accuracy of the hybrid system as a function of α_i, for different epoch lengths, averaged across all target parameters and screen positions. Solid lines represent the average across all subjects, shadows indicate standard error.

accuracy prevails over speed, such as direct gaze input interfaces, fixation-based attention monitoring or analysis of region of interest (ROI).

7 CONCLUSION AND FUTURE WORK

7.1 Conclusion

In this work, we presented a new concept to improve the accuracy of camera-based eye trackers by embedding invisibly flickering stimuli into the virtual scene. This way, we were able to exploit neural mechanisms of visual attention that are more robust to the several factors that affect eye-tracking data quality. Visual attention to a specific flickering target amplifies the corresponding elicited response at the visual cortex, which we took advantage of to gain more information about which was the targeted region. This concept was proved by designing a 7-target figure of 7 coupled hexagons with varying sizes and spacings. We integrated the EEG and eye-tracking data using a probabilistic Bayesian model that leveraged the likelihoods of each individual data source to output the correct target. The hybrid system was evaluated at different screen locations, target parameters and epoch lengths. Our system improved the performance of both the eye-tracking and EEG systems alone for all target diameters and spacings, reaching increments of over 40% for some target parameters and epoch lengths. On average, the largest improvement of the Bayesian system took place at the peripheral screen regions due to the gaze data quality decay. In terms of system speed, the performance of our system is equal or better than any of the two individual systems for all epoch lengths and target parameters, confirming the fruitful integration of the two data modalities and the proper exploitation of

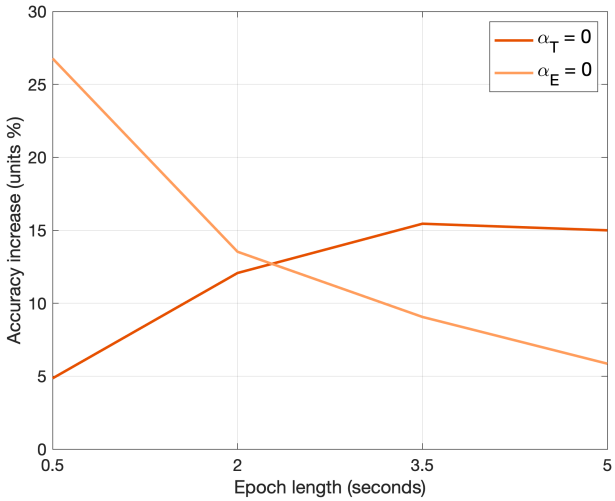


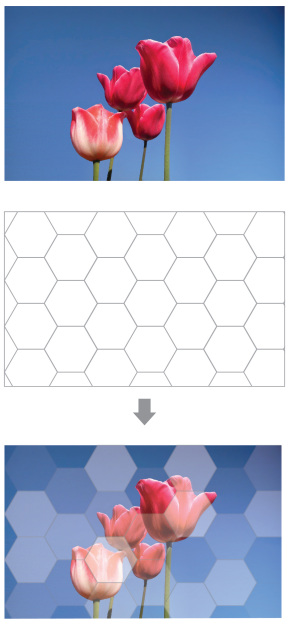
Figure 18: Accuracy improvement (difference) when each individual system shifts from a random classifier ($\alpha_i = 1$) to an intelligent one that exploits the prior probabilities ($\alpha_i = 0$), while the other system is fully engaged ($\alpha_k = 0$), as a function of epoch length.

the prior model likelihoods. For high speeds (short epoch lengths), our Bayesian system is robust enough to match or even outperform the eye-tracker, without being hindered by the lower performance of the EEG system. The contribution of the prior likelihoods of each individual system to the hybrid system accuracy was analyzed . This proved that our hybrid system relayed on both data source modalities rather than just exploiting one of them. The findings of this work demonstrate that the intrinsic accuracy limitations of camera-based eye-trackers can be corrected with the integration of EEG data.

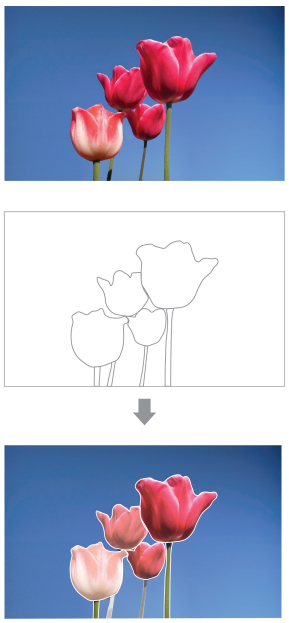
7.2 Future Work

Our future plans for this work span different directions. Firstly, we would like to gather the required training data to compute the conditional probabilities for the complete Bayesian model, as explained above. We intend to compare the performance of the Bayesian system taking into account such conditional probabilities with respect to the current results.

Secondly, we would like to address the appearance limitations described above by implementing phase encoding of targets, as well as introducing frequencies at a higher band of the SSVEP spectrum so the critical fusion rate is well surpassed regardless of the stimuli characteristics. As mentioned in the Limitations section, the critical fusion rate can vary depending on the size of the stimuli, color, subject's age, etc. For that reason we propose that future real-world applications of this work further explore what stimulus frequency range should be used given the characteristics (i.e. size, colors) of the possible regions of interest of that particular application. Moreover, post-experiment questionnaires should be carried out to assess the subjects' perception of the flickering stimuli and the optimal frequencies for the particular application's tasks and users.



(a) Approach implemented in this work, with equally sized screen partitions.



(b) Approach proposed for future work, where the screen is partitioned into segments defined by the objects or ROIs appearing in the screen at each frame.

Figure 19: Current and proposed future implementations of our concept.

Third, we aim to carry out an implementation of this concept on a real virtual scene or application, and test its performance beyond the predefined and uninformative hexagons presented in this study. In use cases where there are clearly delimited objects or ROIs for the user to select or gaze at, we propose to embed the visual flickers in each of these elements (see Figure 19b). In the current work we focused only on the equally-sized screen partition approach since it is more generalizable and allowed for a well-framed study of our hybrid system's accuracy improvements. Once confirmed the effectiveness of our system, other application-specific approaches such as the one shown in Figure 19b can be considered for future work.

And finally, we would like to explore the utilization of feature-based attention (FBA), which is the ability to commit selective attention to non-spatial features, like colors, motions or orientations. We believe FBA could be effective in discriminating semi-occluded, superimposed or motion objects in the virtual scene, and it can be an interesting modality to integrate with our current system.

ACKNOWLEDGMENTS

This research was partially supported by the Obra Social la Caixa Fellowship for Post-Graduate Studies. We thank all participants of the study for their time and interest in this research. We also thank the personnel at the MIT Clinical Research Center for their help and implication during the experiments with subjects. Finally, we would like to thank Maria Tomas for her efforts in creating the illustrations of this work.

REFERENCES

[1] Tobii Pro Nano 2018. [Accessed April 2021]. *Eye Tracker Data Quality Test Report: Accuracy, precision and detected gaze under optimal conditions—controlled environment. Available at: www.tobiipro.com.*

[2] Duane A Anderson, Jane Huntington, and Ernst Simonson. 1966. Critical fusion frequency as a function of exposure time. *JOSA* 56, 11 (1966), 1607–1611.

[3] Alexandre Armengol-Urpi and Sanjay E Sarma. 2018. Sublime: a hands-free virtual reality menu navigation system using a high-frequency SSVEP-based brain-computer interface. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. 1–8.

[4] Michael Ashmore, Andrew T Duchowski, and Garth Shoemaker. 2005. Efficient eye pointing with a fisheye lens. In *Proceedings of Graphics interface 2005*. Citeseer, 203–210.

[5] Guangyu Bin, Xiaorong Gao, Zheng Yan, Bo Hong, and Shangkai Gao. 2009. An online multi-channel SSVEP-based brain–computer interface using a canonical correlation analysis method. *Journal of neural engineering* 6, 4 (2009), 046002.

[6] Pieter Blignaut, Kenneth Holmqvist, Marcus Nyström, and Richard Dewhurst. 2014. Improving the accuracy of video-based eye tracking in real time through post-calibration regression. In *Current Trends in Eye Tracking Research*. Springer, 77–100.

[7] Zillah Boraston and Sarah-Jayne Blakemore. 2007. The application of eye-tracking technology in the study of autism. *The Journal of physiology* 581, 3 (2007), 893–898.

[8] H Bouma and AH De Voogd. 1974. On the control of eye saccades in reading. *Vision Research* 14, 4 (1974), 273–284.

[9] David H Brainard. 1997. The psychophysics toolbox. *Spatial vision* 10, 4 (1997), 433–436.

[10] Chris Brennan, Paul McCullagh, Gaye Lightbody, Leo Galway, Sally McClean, Piotr Stawicki, Felix Gembler, Ivan Volosyak, Elaine Armstrong, and Eileen Thompson. 2020. Performance of a steady-state visual evoked potential and eye gaze hybrid brain-computer interface on participants with and without a brain injury. *IEEE Transactions on Human-Machine Systems* 50, 4 (2020), 277–286.

[11] Donald Eric Broadbent. 2013. *Perception and communication*. Elsevier.

[12] Josef Brozek and Ancel Keys. 1945. Changes in flicker-fusion frequency with age. *Journal of Consulting Psychology* 9, 2 (1945), 87.

[13] Hannah Faye Chua, Julie E Boland, and Richard E Nisbett. 2005. Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences* 102, 35 (2005), 12629–12633.

[14] Yongqin Cui and Jan M Hondzinski. 2006. Gaze tracking accuracy in humans: Two eyes are better than one. *Neuroscience letters* 396, 3 (2006), 257–262.

[15] Robert Desimone and John Duncan. 1995. Neural mechanisms of selective visual attention. *Annual review of neuroscience* 18, 1 (1995), 193–222.

[16] Francesco Di Russo and Donatella Spinelli. 1999. Electrophysiological evidence for an early attentional mechanism in visual processing in humans. *Vision research* 39, 18 (1999), 2975–2985.

[17] Xujiong Dong, Haofei Wang, Zhaokang Chen, and Bertram E Shi. 2015. Hybrid brain computer interface via Bayesian integration of EEG and eye gaze. In *2015 7th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 150–153.

[18] WA Douthwaite, JA Halliwell, AM Lomas, WK Yan Muk, and JN Topliss. 1985. Critical fusion frequency in the central visual field. *Ophthalmic and Physiological Optics* 5, 1 (1985), 15–21.

[19] Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the computer using gaze gestures. In *IFIP Conference on Human-Computer Interaction*. Springer, 475–488.

[20] Andéol Évain, Ferran Argelaguet, Géry Casiez, Nicolas Roussel, and Anatole Lécuyer. 2016. Design and evaluation of fusion approach for combining brain and gaze inputs for target selection. *Frontiers in neuroscience* 10 (2016), 454.

[21] Benedetto Falsini and Vittorio Porciatti. 1996. The temporal frequency response function of pattern ERG and VEP: changes in optic neuritis. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 100, 5 (1996), 428–435.

[22] Stephen Farrell and Shumin Zhai. 2005. System and method for selectively expanding or contracting a portion of a display using eye-gaze tracking. US Patent App. 10/648,120.

[23] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In *Proceedings of the 2017 Chi conference on human factors in computing systems*. 1118–1130.

[24] Sandra Fuchs, Søren K Andersen, Thomas Gruber, and Matthias M Müller. 2008. Attentional bias of competitive interactions in neuronal networks of early visual processing in the human brain. *NeuroImage* 41, 3 (2008), 1086–1101.

[25] Felix W Gembler, Aya Rezeika, Mihaly Benda, and Ivan Volosyak. 2020. Five shades of grey: exploring quintary m-sequences for more user-friendly c-VEP-based BCIs. *Computational Intelligence and Neuroscience* 2020 (2020).

[26] Christoph S Herrmann. 2001. Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental brain research* 137, 3-4 (2001), 346–353.

[27] Steven A Hillyard and Lourdes Anllo-Vento. 1998. Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences* 95, 3 (1998), 781–787.

[28] Steven A Hillyard, Hermann Hinrichs, Claus Tempelmann, Stephen T Morgan, Jonathan C Hansen, Henning Scheich, and Hans-Jochen Heinze. 1997. Combining steady-state visual evoked potentials and fMRI to localize brain activity during selective attention. *Human brain mapping* 5, 4 (1997), 287–292.

[29] Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

[30] Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. 2012. Eye tracker data quality: what it is and how to measure it. In *Proceedings of the symposium on eye tracking research and applications*. 45–52.

[31] Ignace TC Hooge, Gijs A Holleman, Nina C Haukes, and Roy S Hessels. 2019. Gaze tracking accuracy in humans: One eye is sometimes better than two. *Behavior Research Methods* 51, 6 (2019), 2712–2721.

[32] Anthony J Hornof and Tim Halverson. 2002. Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers* 34, 4 (2002), 592–604.

[33] Baihan Huang, Anthony HP Lo, and Bertram E Shi. 2013. Integrating EEG information improves performance of gaze based cursor control. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 415–418.

[34] Thomas E Hutchinson, K Preston White, Worthy N Martin, Kelly C Reichert, and Lisa A Frey. 1989. Human-computer interaction using eye-gaze input. *IEEE Transactions on systems, man, and cybernetics* 19, 6 (1989), 1527–1534.

[35] Florian Hutzler, Mario Braun, Melissa L-H Võ, Verena Engl, Markus Hofmann, Michael Dambacher, Helmut Leder, and Arthur M Jacobs. 2007. Welcome to the real world: Validating fixation-related brain potentials for ecologically valid settings. *Brain research* 1172 (2007), 124–129.

[36] Robert JK Jacob. 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. *ACM Transactions on Information Systems (TOIS)* 9, 2 (1991), 152–169.

[37] Robert JK Jacob and Keith S Karn. 2003. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye*. Elsevier, 573–605.

[38] Lu Jiang, Yijun Wang, Weihua Pei, and Hongda Chen. 2019. A four-class phase-coded SSVEP BCI at 60Hz using refresh rate. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 6331–6334.

[39] Juan E Kamienkowski, Matias J Ison, Rodrigo Quian Quiroga, and Mariano Sigman. 2012. Fixation-related potentials in visual search: a combined EEG and

eye tracking study. *Journal of vision* 12, 7 (2012), 4–4.
[40] Jiannan Kang, Xiaoya Han, Jiajia Song, Zikang Niu, and Xiaoli Li. 2020. The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data. *Computers in biology and medicine* 120 (2020), 103722.
[41] Rami N Khushaba, Chelsea Wise, Sarath Kodagoda, Jordan Louviere, Barbara E Kahn, and Claudia Townsend. 2013. Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (EEG) and eye tracking. *Expert systems with applications* 40, 9 (2013), 3803–3812.
[42] Byung Hyung Kim, Minho Kim, and Sungho Jo. 2014. Quadcopter flight control using a low-cost hybrid interface with EEG-based classification and eye tracking. *Computers in biology and medicine* 51 (2014), 82–92.
[43] Yee Joon Kim, Marcia Grabowecky, Ken A Paller, Krishnakumar Muthu, and Satoru Suzuki. 2007. Attention induces synchronization-based response gain in steady-state visual evoked potentials. *Nature neuroscience* 10, 1 (2007), 117–125.
[44] Manu Kumar, Andreas Paepcke, and Terry Winograd. 2007. Eyepoint: practical pointing and selection using gaze and keyboard. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 421–430.
[45] Po-Lei Lee, Jyun-Jie Sie, Yu-Ju Liu, Chi-Hsun Wu, Ming-Huan Lee, Chih-Hung Shu, Po-Hung Li, Chia-Wei Sun, and Kuo-Kai Shyu. 2010. An SSVEP-actuated brain computer interface using phase-tagged flickering sequences: a cursor system. *Annals of biomedical engineering* 38, 7 (2010), 2383–2397.
[46] Zhonglin Lin, Changshui Zhang, Wei Wu, and Xiaorong Gao. 2006. Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. *IEEE transactions on biomedical engineering* 53, 12 (2006), 2610–2614.
[47] Simon P Liversedge and John M Findlay. 2000. Saccadic eye movements and cognition. *Trends in cognitive sciences* 4, 1 (2000), 6–14.
[48] Lab Streaming Layer (LSL). [Accessed September 2021]. *https://labstreaminglayer.readthedocs.io/index.html*.
[49] Päivi Majaranta and Kari-Jouko Räihä. 2007. Text entry by gaze: Utilizing eye-tracking. *Text entry systems: Mobility, accessibility, universality* (2007), 175–187.
[50] Nikolay V Manyakov, Nikolay Chumerin, Arne Robben, Adrien Combaz, Marijn Van Vliet, and Marc M Van Hulle. 2013. Sampled sinusoidal stimulation profile and multichannel fuzzy logic classification for monitor-based phase-coded SSVEP brain–computer interfacing. *Journal of neural engineering* 10, 3 (2013), 036011.
[51] Xiaoqian Mao, Wei Li, Hong Hu, Jing Jin, and Genshe Chen. 2020. Improve the Classification Efficiency of High-Frequency Phase-Tagged SSVEP by a Recursive Bayesian-Based Approach. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28, 3 (2020), 561–572.
[52] David P McMullen, Guy Hotson, Kapil D Katyal, Brock A Wester, Matthew S Fifer, Timothy G McGee, Andrew Harris, Matthew S Johannes, R Jacob Vogelstein, Alan D Ravitz, et al. 2013. Demonstration of a semi-autonomous hybrid brain–machine interface using human intracranial EEG, eye tracking, and computer vision to control a robotic upper limb prosthetic. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22, 4 (2013), 784–796.
[53] Darius Miniotas, Oleg Špakov, Ivan Tugoy, and I Scott MacKenzie. 2006. Speech-augmented eye gaze interaction with small closely spaced targets. In *Proceedings of the 2006 symposium on Eye tracking research & applications*. 67–72.
[54] Tirin Moore and Marc Zirnsak. 2017. Neural mechanisms of selective visual attention. *Annual review of psychology* 68 (2017), 47–72.
[55] ST Morgan, JC Hansen, and SA Hillyard. 1996. Selective attention to stimulus location modulates the steady-state visual evoked potential. *Proceedings of the National Academy of Sciences* 93, 10 (1996), 4770–4774.
[56] James D Morgante, Rahman Zolfaghari, and Scott P Johnson. 2012. A critical test of temporal and spatial accuracy of the Tobii T60XL eye tracker. *Infancy* 17, 1 (2012), 9–32.
[57] Matthias M Müller, Terence W Picton, Pedro Valdes-Sosa, Jorge Riera, Wolfgang A Teder-Sälejärvi, and Steven A Hillyard. 1998. Effects of spatial selective attention on the steady-state visual evoked potential in the 20–28 Hz range. *Cognitive brain research* 6, 4 (1998), 249–261.
[58] Tobii Pro Nano. [Accessed April 2021]. *https://www.tobiipro.com/product-listing/nano/*.
[59] Neuroelectrics. [Accessed September 2021]. .
[60] Neuroelectrics. [Accessed September 2021]. *Neuroelectrics Enobio, https://www.neuroelectrics.com/solutions/enobio*.
[61] Kian B Ng, Andrew P Bradley, and Ross Cunnington. 2011. Effect of competing stimuli on SSVEP-based BCI. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 6307–6310.
[62] Anthony M Norcia, L Gregory Appelbaum, Justin M Ales, Benoit R Cottereau, and Bruno Rossion. 2015. The steady-state visual evoked potential in vision research: A review. *Journal of vision* 15, 6 (2015), 4–4.
[63] Marcus Nyström, Richard Andersson, Kenneth Holmqvist, and Joost Van De Weijer. 2013. The influence of calibration method and eye physiology on eyetracking data quality. *Behavior research methods* 45, 1 (2013), 272–288.
[64] Kristien Ooms, Lien Dupont, Lieselot Lapon, and Stanislav Popelka. 2015. Accuracy and precision of fixation locations recorded with the low-cost Eye Tribe tracker in different experimental setups. *Journal of eye movement research* 8, 1 (2015).
[65] Felix Putze, Jutta Hild, Rainer Kärgel, Christian Herff, Alexander Redmann, Jürgen Beyerer, and Tanja Schultz. 2013. Locating user attention using eye tracking and EEG for spatio-temporal event selection. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. 129–136.
[66] Felix Putze, Johannes Popp, Jutta Hild, Jürgen Beyerer, and Tanja Schultz. 2016. Intervention-free selection using EEG and eye tracking. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. 153–160.
[67] Takeshi Sakurada, Toshihiro Kawase, Tomoaki Komatsu, and Kenji Kansaku. 2015. Use of high-frequency visual stimuli above the critical flicker frequency in a SSVEP-based BMI. *Clinical neurophysiology* 126, 10 (2015), 1972–1978.
[68] Ernst Simonson and Josef Brozek. 1952. Flicker fusion frequency: background and applications. *Physiological reviews* 32, 3 (1952), 349–378.
[69] Gerrit A Slavenburg, Marcel Janssens, Luis Lucas, Robert Jan Schutten, and Tom Verbeure. 2020. 46-1: Invited Paper: Variable Refresh Rate Displays. In *SID Symposium Digest of Technical Papers*, Vol. 51. Wiley Online Library, 669–672.
[70] Piotr Stawicki, Felix Gembler, Aya Rezeika, and Ivan Volosyak. 2017. A novel hybrid mental spelling application based on eye tracking and SSVEP-based BCI. *Brain sciences* 7, 4 (2017), 35.
[71] Els Stuyven, Koen Van der Goten, André Vandierendonck, Kristl Claeys, and Luc Crevits. 2000. The effect of cognitive load on saccadic eye movements. *Acta psychologica* 104, 1 (2000), 69–85.
[72] Erich E Sutter. 1992. The brain response interface: communication through visually-induced electrical brain responses. *Journal of Microcomputer Applications* 15, 1 (1992), 31–45.
[73] Tobii. [Accessed September 2021]. *Tobii SDK, https://developer.tobiipro.com/*.
[74] Yijun Wang, Xiaorong Gao, Bo Hong, Chuan Jia, and Shangkai Gao. 2008. Brain-computer interfaces based on visual evoked potentials. *IEEE Engineering in medicine and biology magazine* 27, 5 (2008), 64–71.
[75] Michel Wedel and Rik Pieters. 2000. Eye fixations on advertisements and memory for brands: A model and findings. *Marketing science* 19, 4 (2000), 297–312.
[76] Masatake Yamato, Katsuro Inoue, Akito Monden, Koji Torii, and Ken-ichi Matsumoto. 2000. Button selection for general GUIs using eye and hand together. In *Proceedings of the working conference on advanced visual interfaces*. 270–273.
[77] Xinyi Yong, Mehrdad Fatourechi, Rabab K Ward, and Gary E Birch. 2011. The design of a point-and-click system by integrating a self-paced brain–computer interface with an Eye-tracker. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 1, 4 (2011), 590–602.
[78] Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. 2016. Measuring cognitive load using eye tracking technology in visual computing. In *Proceedings of the sixth workshop on beyond time and errors on novel evaluation methods for visualization*. 78–85.
[79] Thorsten O Zander, Matti Gaertner, Christian Kothe, and Roman Vilimek. 2010. Combining eye gaze input with a brain–computer interface for touchless human–computer interaction. *Intl. Journal of Human–Computer Interaction* 27, 1 (2010), 38–51.
[80] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and gaze input cascaded (MAGIC) pointing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 246–253.
[81] Darisy G Zhao, Anatoly N Vasilyev, Bogdan L Kozyrskiy, Eugeny V Melnichuk, Andrey V Isachenko, Boris M Velichkovsky, and Sergei L Shishkin. 2021. A passive BCI for monitoring the intentionality of the gaze-based moving object selection. *Journal of Neural Engineering* 18, 2 (2021), 026001.
[82] Wei-Long Zheng, Bo-Nan Dong, and Bao-Liang Lu. 2014. Multimodal emotion recognition using EEG and eye tracking data. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 5040–5043.