Using Electroencephalography to locate target objects in the Left and Right visual fields in images at high speeds

Miguel Rocha e Costa, up202103370@up.pt

I. Introduction

In recent years, humanity witnessed remarkable advancements in the field of neurotechnology, specifically in the field of Brain-Computer Interfaces (BCI). This cutting-edge innovation holds immense potential to revolutionize how humans interact with computers, devices, and the world around us. By bridging the gap between the human brain and external systems, BCI technology enables direct communication and control, opening up a realm of possibilities that were once confined to science fiction.

BCI technology operates on the principle of decoding and interpreting brain activity to translate our thoughts, intentions, and commands into actionable outputs. Because of this, the emerging importance of BCI technology stems from its potential on many fronts, such as empowering individuals with motor disabilities [1], reshaping healthcare practices by improving the diagnosis of certain conditions, namely the designated Diseases of Consciousness (DOC) [2], enhance human-computer interaction in Virtual environments both for medical (e.g., neurorehabilitation) [3] and non-medical purposes (*e.g.,* gaming) [4], and advance our understanding of the human brain. The booming of the technology has led to the creation of several companies that try to penetrate these systems in the market at affordable prices, namely *OpenBCI*, *Neurosity*, *InteraXon Inc* (Muse), and others trying to solve more serious issues such as Elon Musk's *Neuralink*. Because of this, BCI technology has been recently designated as the technology of the year 2023 according to *Nature Electronics* [5].

Because it relies on the ongoing information of the brain, the first problem associated with the technology lies in choosing the data acquisition technique. Focusing on non-invasive methodologies, albeit many neuroimaging techniques have been successfully used to obtain the required brain activity, specifically Functional Magnetic Resonance Imaging (fMRI), Functional Near-Infrared Spectroscopy (fNIRS), and Positron Emission Tomography (PET), despite its simplicity, Electroencephalography (EEG) still is the most popular brain activity acquisition technique used in the field [6]. In fact, its simplicity is one of the reasons why alongside with EEG systems being portable, inexpensive, and despite having low spatial resolution compared to other techniques, it compensates for the fact that it can read brain waves with a very high temporal resolution, namely at the speed of cognition [6].

The use of EEG in BCI systems shares a broad variety of applications with the most popular one being the restoration of mobility in motor-impaired individuals. Numerous studies have surfaced, demonstrating the undeniable benefits of these technologies across various disability levels, including bionic limb control [7], spelling applications [8], environmental navigation, particularly in wheelchair control [9], and even in addressing vegetative states [2]. Under the right circumstances, EEG signals can capture potentials related to the processing visual stimuli, for example, to continuously select a letter in a list of rapidly presented letters (10 Hz) to ultimately write words and sentences. This type of BCI is referred to as the speller and the experimental paradigm used to achieve this result is denominated Rapid Serial Visual Presentation (RSVP) [10].

RSVP paradigms have been extended to research novel BCI uses such as in neuromarketing, where brain signals are used to understand the cognitive functions that underlie the customer's decision to purchase a service or product [11]. In this particular study, subjects were presented with several images used in the advertisement of the *Coca-Cola* brand. The authors then build a model that was able to cluster the images and generated a faster and larger response, inferred as the images that were better perceived and associated with the brand [11]. Another study used an RSVP-BCI to reduce the time that security spent analyzing X-ray images in the airport. Using similar methodologies, the BCI was able to capture the existence of a gun in the X-ray image of the luggage by accessing the attentional mechanism of the brain that detects target objects in images [12]. The most impressive feature of these types of BCI is that they are able to discern these changes in brain waves and perform these decisions at the borderline of human consciousness, meaning that it is able to do it faster than the user is able to consciously perform that decision. Because of this, these BCIs offer an important field of research that may be able to help people perform tasks more efficiently [13].

This study focuses on similar problems. To be more precise, the goal of this study is to exploit the spatial-temporal characteristics of the N2pc component of the EEG, which will be accordingly contextualized below, and build an algorithm to determine whether a specific target in several images was present in the Left or Right Visual field using only information from the Brain. To achieve this, a public dataset of an RSVP paradigm related to the elicitation of the N2pc was used [14]. This work not only aims to develop the algorithm based on the methodology of the paper [15] where the dataset is inserted to drive the realization of the thesis but also tries to expand

the analysis of the classification process.

II. Electroencephalography

In order to understand the way the N2pc component is computed, a few concepts of the recording technique, the EEG, must be first reviewed.

Electroencephalography (EEG) is a non-invasive neurophysiological technique used to measure and record the electrical activity of the brain. EEG recordings are obtained by placing electrodes on the scalp of an individual, which detects and measures the electrical signals generated by the summed contribution of the brain's pyramidal neurons oriented parallel to each other [16]. The signal captured by this technique is highly non-linear and non-stationary with spectral characteristics between 0.5 and 60 Hz and amplitudes ranging from 0.5 to 200 μV [17].

EEG is widely used in neuroscience and clinical settings to study brain function, analyze brain disorders, and investigate cognitive processes. It provides valuable insights into brain activity patterns during various states such as sleep, wakefulness, attention, and different cognitive tasks [16]. EEG is particularly useful for studying temporal dynamics, event-related potentials (ERPs), and oscillatory activity both in healthy and pathological brains [18].

The development of EEG (Electroencephalography) required a standardized approach for electrode placement and nomenclature. This led to the establishment of the 10-20 system, which defines electrode positions based on specific measurements of the skull circumference [19]. The system divides the scalp longitudinally into 10% and 20% sections from the Nasion to the Inion, passing through the vertex and transversally from the left to the right preauricular point, again passing through the vertex [19]. Letters are used to represent different lobes of the brain: O for occipital, P for parietal, T for temporal, F for frontal, and C for central. The midline electrodes are designated with the letter z, such as Cz for zero. The numbering of electrodes starts incrementally from the midline, with even numbers for the right hemisphere and odd numbers for the left hemisphere [19].

Hardware advancements provided the tools to develop higher-density systems of electrodes that somewhat mitigate the low-resolution problem of the EEG technique mentioned previously. These are also based on measurements of the scalp and comprise the 10-10 and 10-5 systems. A topological representation of the 10-10 system can be seen in Figure 1.

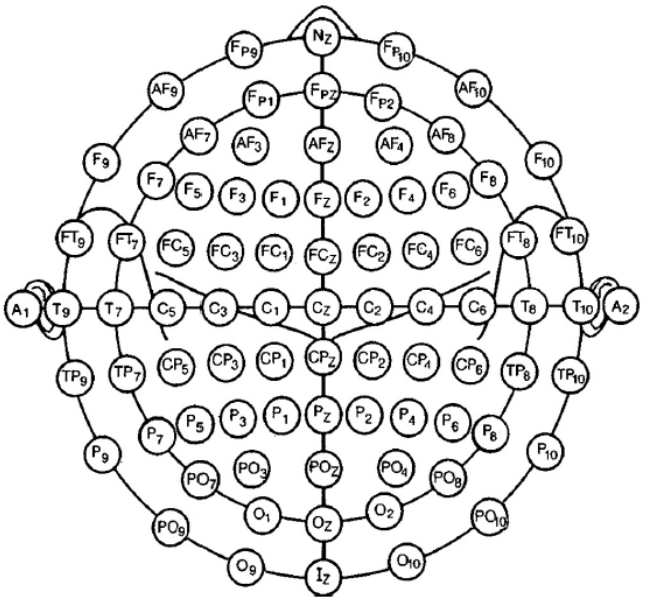

Fig. 1. Illustration of the 10-10 EEG system. Adapted from [19].

III. Rapid Serial Visual Presentation

The analysis of brain waves and consequently the BCI application heavily relies on the type of data that is being acquired, *i.e.,* the regions of the brain that are targeted in order to elicit the pretended signal and features. One very recently emerging type of application focuses on capturing brain waves during very fast presentation rates of images. This acquisition paradigm is denoted Rapid Serial Visual Presentation (RSVP) and consists of the process of sequentially presenting visual stimuli (e.g., letters or images) several times per second at a constant rate, typically used within BCI contexts to detect visual stimuli of interest [13]. These BCI systems can be coined as cortically-coupled computer vision [20].

Despite having come a long way, computer vision algorithms are yet to surpass the brain's abilities to analyze information and perform inferences of the environment such as to recognize objects. Thus, this type of paradigm exploits the brains built-in visual mechanisms to triage images in an accurate manner under a few hundred of milliseconds by extracting relevant information from circuitry related to the identification and detection of an object of interest with the user being barely aware of it [20]. RSVP-driven brain-computer interfaces (BCIs) are starting to be viewed as a practical strategy for improving the symbiosis between humans and machines and present promising opportunities for augmenting human capabilities during visual tasks [13]. In situations where humans are required to carefully review and analyze a large number of images in counterintelligence, policing, and healthcare applications, this paradigm can provide an alternative to significantly reduce triage time [13].

A. Event-Related Potentials

In order to identify target objects using brain signals during the RSVP method, the spatial-temporal characteristics carried in Event-Related Potentials (ERP) elicited during the *oddball* experiment are evaluated. Event-Related Potentials are a series of potentials that carry the whole volley of, in this case, brain processing of visual stimuli. ERPs are the result of averaging neural activity in response to repeated presentation of a specific stimulus. Compared to ongoing brain waves captured by EEG, ERPs typically have lower amplitudes, measuring around 10 μV [21]. During the study of the brain's response to stimuli, participants may exhibit spontaneous and involuntary neural activity at any given time during the recording. This neural activity can be influenced by factors such as attention, motor acts, and inner thoughts, resulting in random segments

of activity in the recorded signal that may overshadow the intended response. On the other hand, the neural activity evoked by a specific stimulus will consistently occur at the moment of its presentation. By averaging the EEG activity recorded in response to the same stimulus across multiple trials, the spontaneous and time-variable activity tends to cancel out, while isolating the signal that is phase-locked to the onset of the stimuli. Furthermore, averaging acts as a low-pass filter which helps in reducing the noise of the signal [22].

In the oddball paradigm, participants are presented with a series of stimuli that can be categorized into two classes: targets and non-targets. Targets are less frequent than non-targets, usually comprising around 5-10% of the total stimuli in the Rapid Serial Visual Presentation (RSVP) paradigm [13]. The targets are intentionally designed to be noticeably different from the non-targets, making them distinguishable to the participants. The task for participants is to classify these stimuli based on their respective categories.

In these situations, several potentials are generated after stimuli onset, each corresponding to a specific mental process. The most researched and used in the BCI field of these potentials is the P300 related to visual attention mechanisms in the brain [23]. However, several others, namely the N2pc are emerging and are being proven to be helpful in designing BCI technology.

1) N2pc: Similarly to P300, the N2pc is a component related to visual selective attention, appearing in the cortex contralaterally to where the object of interest appears in the visual field. In other words, if the object appears on the left visual field, the component will have a stronger contribution on the right side of the hemisphere and vice-versa [23; 15]. The N in N2pc is indicative of the negativity of the peak with amplitudes ranging the 2 μV within 170-300 after stimulus onset. Spatially, the N2pc reaches its maximum amplitude in electrode sites P7/P8 and PO7/PO8.

B. Classification models

The classification stage of the BCI pipeline handles the discrimination of separate tasks. Regarding BCI classification algorithms, the SVM and LDA are the most used in the field due to their high computational efficiency, ideal for scenarios where fast processing and classification are required [24]. However, with access to more computational power, other models have been recently emerging, namely the Random Forest, XGBoost, and AdaBoost. A brief description of their inner workings is given below.

1) Support Vector Machine: A Support Vector Machine (SVM) is a powerful machine learning algorithm used for classification and regression tasks. It is particularly effective in dealing with complex datasets with a clear margin of separation between classes [25]. In SVM, the algorithm tries to find an optimal hyperplane that separates the data into different classes. The hyperplane is chosen in such a way that it maximizes the margin, which is the distance between the hyperplane and the nearest data points of each class [25]. The data points closest to the hyperplane are called support vectors. The key idea behind SVM is to transform the input data into a higher-dimensional feature space using a kernel function. By doing so, the SVM can find a linear hyperplane that effectively separates the data even if it is not linearly separable in the original input space. Common kernel functions used in SVM include linear, polynomial, radial basis function (RBF), and sigmoid [25].

2) Linear discriminant Analysis: Linear Discriminant Analysis (LDA), on the other hand, is a statistical method used for dimensionality reduction and classification tasks. It is a supervised learning algorithm that aims to find a linear combination of features that maximizes the separation between different classes in the data [26]. LDA starts with data preparation, where the features are typically scaled to have zero mean and unit variance. This step ensures that all features contribute equally to the analysis. Next, LDA computes class statistics. It calculates the mean vector for each class, representing the average values of the features within that class. It also calculates the scatter matrix, which measures the spread of the data within each class. The scatter matrix can be computed using either the covariance matrix or the pooled within-class scatter matrix [26].

LDA also determines the between-class scatter matrix, which quantifies the spread between the class means. It is obtained by summing the individual scatter matrices weighted by the number of samples in each class. The eigendecomposition of the matrix product of the inverse of the within-class scatter matrix and the between-class scatter matrix is then performed [26]. This yields eigenvectors and eigenvalues. The eigenvectors corresponding to the largest eigenvalues represent the directions in the feature space that maximize class separability. The eigenvectors, known as discriminant functions, are used to transform the original feature space into a new feature space of lower dimensionality. These transformed features are called linear discriminants. The number of linear discriminants is equal to the number of classes minus one. In the transformed feature space, a classification model can be trained using linear discriminants as input features [26].

3) Random Forest: Random Forest is a technique that combines the strength of multiple decision trees to make accurate predictions. Instead of relying on a single decision tree, Random Forest creates an ensemble of decision trees [26]. Each decision tree in the Random Forest is trained on a different subset of the original data. These subsets, known as bootstrap samples, are created by randomly selecting data points with replacement [26]. This sampling process introduces variability and diversity among the trees. Additionally, at each split of a decision tree, only a random subset of features is considered as candidates. By doing this, Random Forest encourages each tree to focus on different aspects of the data. This diversification helps capture a broader range of patterns and improves the overall performance. The construction of each decision tree follows a recursive process called recursive partitioning [26]. The tree grows by making splits based on specific criteria like Gini impurity or information gain. The goal is to create nodes that separate the data points into increasingly pure subsets.

When it's time to make predictions, Random Forest combines the outputs of all the individual decision trees. For

classification tasks, it uses a voting scheme where each tree "votes" for a class, and the majority class wins. For regression tasks, the predictions of the individual trees are averaged to get the final prediction.

4) XGBoost: XGBoost, short for Extreme Gradient Boosting, is an advanced machine learning algorithm known for its exceptional performance and scalability. It is a boosting algorithm that combines the predictions of multiple weak learners to create a powerful ensemble model [26].

The key concept behind XGBoost is gradient boosting, which is a sequential process of building an ensemble of weak learners. It starts with a single weak learner, often a decision tree, and then iteratively adds more trees to refine the predictions. During each iteration, XGBoost focuses on the examples that were previously misclassified or had higher errors [26]. It assigns higher weights to these examples, directing subsequent weak learners to pay more attention to them. This gradient-based approach allows XGBoost to learn from its previous mistakes and prioritize areas where improvement is needed the most. To ensure that each new weak learner contributes to the overall ensemble without overshadowing the others, XGBoost employs a technique called regularization. Regularization introduces constraints on the complexity of individual trees, preventing overfitting and promoting generalization [26].

XGBoost also incorporates a clever optimization algorithm to efficiently search for the best tree structures. It uses second-order gradients to find the optimal splits, resulting in faster and more accurate tree construction [26].

5) AdaBoost: AdaBoost, also known as Adaptive Boosting, is an ensemble learning algorithm used for classification tasks. It combines multiple weak learners to create a strong predictive model. Unlike traditional ensemble methods that assign equal weights to all classifiers, AdaBoost assigns weights to each weak learner based on its performance. [26]

AdaBoost begins by initializing the weights of all training examples to be equal. It then iteratively trains a series of weak learners, such as decision stumps, on the training data. During each iteration, AdaBoost adjusts the weights of the training examples based on their classification accuracy. It assigns higher weights to misclassified examples, forcing subsequent weak learners to focus more on these challenging instances [26]. In each iteration, AdaBoost aims to find the weak learner that best classifies the weighted training examples. The algorithm evaluates different weak learners and selects the one with the lowest weighted error. The weak learner's performance determines its contribution to the final ensemble model [26].

After training all the weak learners, AdaBoost combines their predictions to make the final classification. Each weak learner's prediction is weighted based on its accuracy during training [26]. The ensemble prediction is typically determined by majority voting, where the class with the most weighted votes is selected as the final prediction. AdaBoost continues this iterative process until a predefined stopping criterion is met, such as reaching a specified number of weak learners or achieving a desired level of accuracy. The final ensemble of weak learners, with their weighted contributions, forms a strong classifier capable of accurate predictions [26].

IV. Methodology

In this section, the materials and the algorithm used are detailed, including a description of the data in the dataset and each step of the pipeline used.

A. RSVP Dataset

This research project involves the analysis of EEG data obtained from 11 individuals who are in good health [14]. The data was collected during an experiment where images were rapidly presented to the participants using the Rapid Serial Visual Presentation (RSVP) protocol in which several presentation speeds were used, in particular at 5, 6, and 10 Hz.

The dataset includes aerial photographs of London, covering an angle of 11.5 x 11.8 degrees in the visual field. The target images consist of airplanes that have been realistically superimposed, randomly rotated, and positioned. Non-target images, on the other hand, do not contain airplanes. The horizontal position of a target image is determined by the x-coordinate of the airplane's centroid. The database specifically includes lateral targets, which are defined as those with a centroid positioned at least +/-1.2 degrees from the center of the screen along the horizontal axis. Lateral targets account for approximately 60% of the total target images in the dataset.

The dataset provided consists of EEG signals saved in *.edf* format recorded from 8 channels according to the 10-20 system, specifically channels PO8, PO7, PO3, PO4, P7, P8, O1, and O2. Each participant at each presentation rate has two corresponding files. The last field in the file name, either "a" or "b," indicates the order of the recordings. The participants had a significant break between the two recordings.

The annotations provided in the dataset indicate the onset of each image presented during the task. Additionally, the annotations specify whether each image was a target (T=1) or not (T=0). If an image is a target, the x-coordinate of its centroid (ranging from 0 to 640 pixels) is also included in the annotation.

The EEG data were acquired at a sampling rate of 2048 Hz and were pre-processed by applying a band-pass filter between 0.15 and 28 Hz.

B. Algorithm pipeline

Due to N2pc being extremely small and thus a hard potential to detect when compared to the P300, the original paper focused on merging signals from a number of users via collaborative-BCI (cBCI) to improve the classification. However, since using the system with more people may prove to be difficult and impractical, this analysis focused on improving the single-user classification by trying a few other models and feature extraction parameters. Other than these changes, the pipeline used was identical to the paper. The presentation rates of 5Hz, 6Hz, and 10 Hz were assessed and the model's performance was compared using the Area Under the Curve due to the original paper using only this measurement to evaluate their model.

The language of choice to develop the pipeline was *python*. Because of this, the main package used was MNE [27] which

is an Open-source *Python* package for exploring, visualizing, and analyzing human neurophysiological data.

1) Pre-processing: The first task involved organizing the data of the dataset by concatenating all trials of each participant contained in files 'a' and 'b' corresponding to the two separate acquisitions. Upon being loaded to memory, the signals were baseline corrected with a window of -200 to 0 ms before stimuli onset. In other words, the mean amplitude value was subtracted from each time point of the whole trial period. This is important to normalize the ERP amplitudes across trials and subjects by removing the DC offset of the individual signals.

2) Visual angle calculation: Because only the x-coordinates of the target centroid were provided, a *function* that is able to determine if the image is in the Left (LVF) or Right (RVF) visual field attending to the threshold of 1.2 degrees from the center of the screen along the horizontal axis was designed. The *function* is based on the visual angle principle, also known as the object's angular size, and refers to the angle that an observed object subtends at the eye. It is typically measured in degrees of arc [28]. In simpler terms, it represents the apparent size of an object as perceived by an observer. Figure 2 depicts both a diagram and the formula used to achieve the calculation for each image. In the accompanying diagram, the eye of the observer is depicted looking at a frontal extent, indicated by the vertical arrow. This extent has a linear size referred to as S and is situated at a distance D from a point labeled O. Point O can be understood as representing the center of the eye's entrance pupil, which is positioned just a few millimeters in front of the lens [28]. The diagram illustrates three lines originating from endpoint A of the object, extending towards the eye. These lines symbolize the bundle of light rays that pass through the cornea, pupil, and lens, ultimately forming an optical image of endpoint A on the retina at a point denoted as a. Among these lines, the central line represents the chief ray. The same principles apply to object point B and its corresponding retinal image at point b. The visual angle, denoted as V, refers to the angle formed by the chief rays of points A and B. It quantifies the angular separation between these rays and provides a measure of the perceived size or extent of the object as observed by the viewer [28].

3) Feature extraction and classification: This work performed both feature extraction and classification in single-trial data. This means that each trial of each subject represents one row of the feature vector that is fed to the classifier. Because BCI performance is generally subject-dependent due to everyone's brains generating signals with slightly different characteristics across the same experimental paradigm, a classifier was specifically tailored for each individual.

The N2pc was computed by subtracting the contralateral and ipsilateral components on both LVF and RVF stimuli [15]. Particularly, the electrode pair differences PO7-PO8, P7-P8, PO3-PO4, and O1-O2 were computed. This means that for LVF stimuli, the ipsilateral electrodes PO7, P7, PO3, O1 were subtracted from the contralateral electrodes PO8, P8, PO4, O2. The resulting signals were then downsampled to 64 Hz to prevent overfitting issues [15]. By concatenating the samples in the period where according to the literature

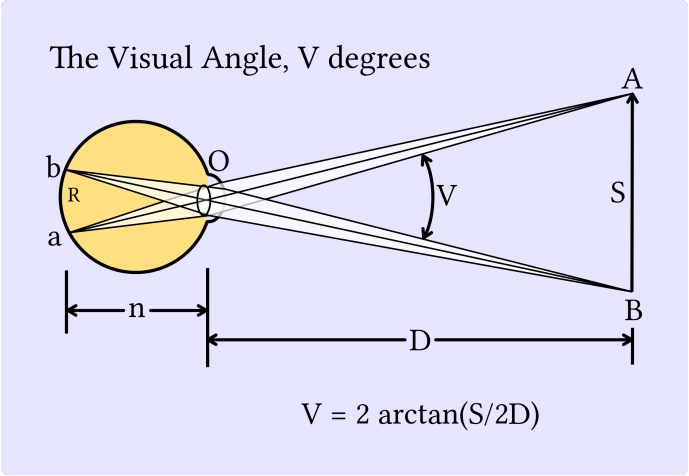

Fig. 2. Illustration of the calculation of visual angle. Adapted from [28].

the N2pc component emerges, *i.e.,* between 200 and 400 ms after stimuli onset, of all downsampled difference pairs, the resultant feature vector comprises 56 elements per trial.

Regarding the classification process, using a 75% training 25% testing split, the five models were validated using a 10-fold Straight Cross-Validation. In addition to the Support Vector Machine (SVM) with the linear kernel that was used in the original study, a Linear Discriminant Analysis (LDA), a Random Forest, an eXtreme Gradient Boosting (XGBoost), and an Adaptive Boosting (AdaBoost) were used. For the SVM model, all of the kernels mentioned above were used and compared in this analysis, namely linear, polynomial, radial basis function (RBF), and sigmoid. As for the LDA model, the Single Value Decomposition (SVD) and Least Square Solution (LSQR) solvers were used.

V. Results and Discussion

Firstly, a look at the grand-average N2pc component was computed across the 4 electrode pair differences was given. In ERP analysis, the grand-average represents the average peak across all target trials (LVF and RVF) and participants. The grand-average peaks with 5Hz presentation rate are depicted in Figure 3.

It is important to note that the standard representation of ERP components was used in Figure 3, *i.e.,* the graph is flipped along the x-axis, meaning that the negative y values are in the upper portion of the plot [15]. Furthermore, it was noted that the data in channels PO8 and PO7 are switched as the difference between contralateral and ipsilateral contributions resulted in a positive potential where the N2pc is supposed to be. However, since this is constant in all participants in those channels and classifiers are blind to those changes, the data was not flipped.

As for differences in the peaks within the 3 presentation rates used, just like in the original study, at 6Hz the grand-averaged peaks had slightly greater amplitudes when compared to 5 and 10Hz presentation rates. Similarly, at 10Hz, the amplitudes were the lowest of all presentation rates.

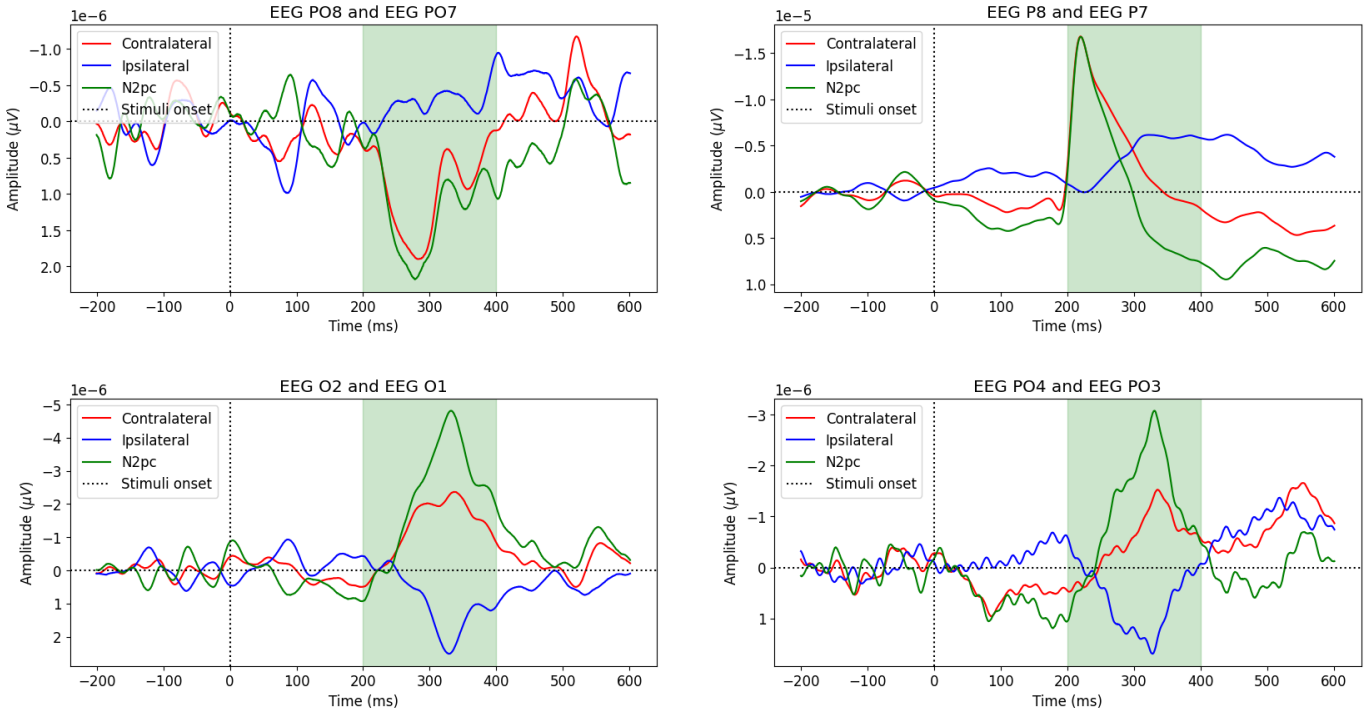

Fig. 3. Grand-average N2pc component in green alongside the Contralateral (red) and Ipsilateral (blue) components used to compute it. A green patch was drawn between 200 and 400 ms after stimuli onset to show the time period that was used to segment samples associated with the N2pc component that fit the classifier.

Several downsample frequencies were tested, namely 32, 64, and 128. However, the original value of 64 always worked better across all models.

Regarding the classification process, the AUC scores were used to compare the performance of the models. The AUC represents the probability that a randomly selected positive example will be ranked higher by the model than a randomly selected negative example. It ranges from 0 to 1, where a higher value indicates better performance. Table I holds the AUC scores for the SVM model at a presentation rate of 5Hz. Although the linear model tested here performed significantly worse than the linear SVM used in the paper, it is important to note that the individual AUC in Subjects 9 and 10 was greater than in the paper. However, in some participants, it performed significantly worse.

Due to the degree of 3 being the one that achieved the best results with the polynomial kernel, only the results with this degree are shown. Similarly to the linear kernel, this model performed worse on average with a lot more inter-subject variation in the AUC than initially seen in the paper. The RVF kernel provided a good improvement with four subjects achieving a better score than in the paper. However, the most competitive kernel was the Sigmoid achieving an AUC score of 0.74 on average with seven subjects performing better than in the original study.

With regards to the presentation rate of 6Hz with the AUC scores represented in Table II, the models performed worse on average with a bigger gap between the AUC scores achieved and the ones published. At this rate, the RBF and Sigmoid shared an equal result.

TABLE I
AUC SCORES OF SVM MODEL WITH THE FOUR KERNELS TESTED AT 5HZ PRESENTATION RATES. THE BOLD COLUMN REPRESENTS THE BEST TESTED MODEL WITH THE LAST COLUMN 'PAPER' REPRESENTING THE RESULTS PUBLISHED IN THE PAPER [15]

5 Hz					
Subject	Linear	Polynomial (degree 3)	RBF	**Sigmoid**	Paper
1	0.0	1.0	1.0	**1**	0.78
2	**0.55**	0.33	0.89	**0.3**	0.85
3	0.5	0.75	0.75	**0.86**	0.78
4	0.59	0.5	0.48	**0.3**	0.80
5	0.6	0.7	0.77	**0.81**	0.82
6	0.59	0.45	0.32	**0.69**	0.66
7	0.14	0.26	0.45	**0.82**	0.64
8	0.21	0.22	0.77	**0.71**	0.61
9	0.89	0.8	0.86	**0.89**	0.77
10	0.88	0.64	0.8	**0.88**	0.66
11	0.59	0.4	0.41	**0.61**	0.71
Average	0.50	0.55	0.68	**0.74**	0.77

With a presentation rate of 10Hz (Table III), the trend seen in the other two presentation rates continues with the Sigmoid, achieving the best result on average but still lower than in the paper. Thus, the most competitive result with the SVM was achieved with a presentation rate of 5Hz.

The LDA model (Table IV) with the SVD solver provided much more consistent inter-subject and on-average results than the SVM models. Notably, the performance improved with the increase of the presentation rate, achieving an AUC score of 0.74 at 10Hz. Conversely, the LSQR solver did not match the performance requirements with low AUC scores across the board.

TABLE II
AUC SCORES OF SVM MODEL WITH THE FOUR KERNELS TESTED AT 6HZ PRESENTATION RATES. THE BOLD COLUMN REPRESENTS THE BEST TESTED MODEL WITH THE LAST COLUMN 'PAPER' REPRESENTING THE RESULTS PUBLISHED IN THE PAPER [15]. THE MISSING VALUES REPRESENT MISSING DATA FOR THE SPECIFIC SUBJECT.

6 Hz					
Subject	Linear	Polynomial (degree 3)	RBF	**Sigmoid**	Paper
1	0.0	0.0	1.0	**1**	0.85
2	0.17	0.82	0.88	**0.89**	0.91
3	0.21	0.81	0.81	**0.81**	0.85
4	-	-	-	**-**	-
5	0.43	0.57	0.64	**0.68**	0.82
6	0.5	0.76	0.5	**0.43**	0.56
7	0.38	0.67	0.38	**0.46**	0.76
8	0.10	0.73	0.31	**0.18**	0.68
9	0.07	0.93	0.95	**0.95**	0.73
10	0.55	0.48	0.56	**0.63**	0.66
11	0.57	0.64	0.54	**0.57**	0.76
Average	0.29	0.64	0.66	**0.66**	0.76

TABLE III
AUC SCORES OF THE SVM MODEL WITH THE FOUR KERNELS TESTED AT 10HZ PRESENTATION RATES. THE BOLD COLUMN REPRESENTS THE BEST TESTED MODEL WITH THE LAST COLUMN 'PAPER' REPRESENTING THE RESULTS PUBLISHED IN THE PAPER [15]. THE MISSING VALUES REPRESENT MISSING DATA FOR THE SPECIFIC SUBJECT.

10 Hz					
Subject	Linear	Polynomial (degree 3)	RBF	**Sigmoid**	Paper
1	0.23	0.17	0.83	**0.9**	0.80
2	0.5	0.59	0.89	**0.83**	0.80
3	0.63	0.5	0.38	**0.5**	0.83
4	0.45	0.5	0.6	**0.65**	0.60
5	0.5	0.35	0.27	**0.73**	0.82
6	0.63	0.3	0.16	**0.54**	0.55
7	0.5	0.56	0.75	**0.53**	0.71
8	0.64	0.53	0.89	**0.42**	0.60
9	0.0	1.0	1.0	**1.0**	0.83
10	0.5	0.54	0.67	**0.83**	0.75
11	-	-	-	**-**	-
Average	0.46	0.50	0.66	**0.69**	0.76

TABLE IV
AUC SCORES OF THE LDA MODEL WITH THE TWO SOLVERS TESTED AT 5 HZ, 6 HZ, AND 10 HZ PRESENTATION RATES. THE BOLD COLUMN REPRESENTS THE BEST TESTED MODELS FOR EACH PRESENTATION RATE WITH THE LAST COLUMN 'PAPER' REPRESENTING THE RESULTS PUBLISHED IN THE PAPER [15]. THE MISSING VALUES REPRESENT MISSING DATA FOR THE SPECIFIC SUBJECT.

	LDA - 5 Hz		LDA - 6Hz		LDA - 10 Hz	
Subject	**SVD**	LSQR	**SVD**	LSQR	**SVD**	LSQR
1	**0.86**	1.0	**0.94**	0.71	**0.73**	0.9
2	**0.73**	0.33	**0.69**	0.47	**0.75**	0.42
3	**0.85**	0.75	**0.78**	0.63	**0.83**	0.71
4	**0.68**	0.5	**-**	-	**0.9**	0.53
5	**0.74**	0.7	**0.96**	0.27	**0.8**	0.57
6	**0.27**	0.45	**0.48**	0.79	**0.46**	0.63
7	**0.61**	0.26	**0.27**	0.44	**0.67**	0.66
8	**0.58**	0.22	**1**	0.29	**0.78**	0.67
9	**0.81**	0.8	**0.79**	0.51	**0.94**	0.75
10	**0.89**	0.64	**0.75**	0.58	**0.54**	0.38
11	**0.52**	0.4	**0.35**	0.35	**-**	-
Average	**0.69**	0.54	**0.70**	0.35	**0.74**	0.62

In Table V the Random Forest, XGBoost, and AdaBoost models are compared according to their AUC score. Across the board, these models performed more consistently with less pronounced inter-subject variances. The XGBoost model achieved the best average AUC scores during the 5Hz and 6Hz trials, both when compared to the other models used in this study and to the published results. The Random Forest came in second place, also performing slightly better than the original SVM used. However, at 10Hz presentation rates, the models used here did not surpass the original results with the AdaBoost coming the closest with an AUC score of 0.74.

Despite apparently not significantly improving the average AUC scores, it is interesting to note the inter-subject variances across all models. When compared to the original results, the AUC values presented here are more inconsistent across subjects with even the worst models achieving results better than in the paper for a specific subject. Moreover, some models were able to achieve AUC scores of 1.0 for certain subjects with some models producing higher scores across the majority of subjects (XGBoost). These changes in results with high variances between subjects, even when comparing the same linear model, can be explained by differences in the trial count. In the original paper, it is stated that 59 trials for LVF and 85 for RVF were used [15]. However, in the dataset, the average number of LVF trials is 34 and 37 for the RVF trials. This makes it hard to accurately compare the results. Despite this, if the linear SVM model performed much worse than the one in the paper, we could loosely infer that the performance improvement seen here with XGBoost might be even greater if the exact same data were provided.

A closer look at the performance of the best model, *i.e.,* XGBoost, is given in the pictures below. Figures 4 and 5 represent the training and testing accuracy, respectively at a 5Hz presentation rate. Figures 6 and 7 represent the training and testing accuracy, respectively at a 6Hz presentation rate. On the other hand, Figures 8 and 9 represent the training and testing accuracy, respectively at a 10Hz presentation rate.

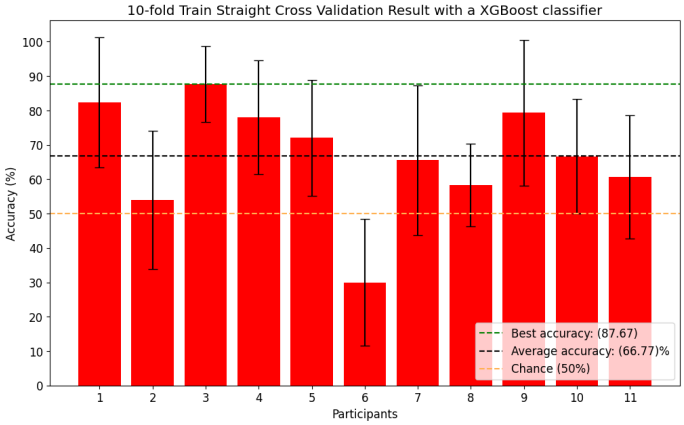

Fig. 4. Cross-validation accuracy at 5Hz presentation rate.

Similar to the AUC scores, the accuracy scores at 6Hz had higher values when compared to other presentation rates, with 66.80% in the training set and 71.08% in the testing set on average. Furthermore, many subjects achieved very respectable

TABLE V
AUC SCORES OF THE RANDOM FOREST, XGBOOST, AND ADABOOST MODELS TESTED AT 5 HZ, 6 HZ, AND 10 HZ PRESENTATION RATES. THE BOLD COLUMN REPRESENTS THE BEST TESTED MODELS FOR EACH PRESENTATION RATE WITH THE LAST COLUMN 'PAPER' REPRESENTING THE RESULTS PUBLISHED IN THE PAPER [15]. THE MISSING VALUES REPRESENT MISSING DATA FOR THE SPECIFIC SUBJECT.

	5 Hz			6Hz			10 Hz		
Subject	Random Forest	**XGBoost**	AdaBoost	Random Forest	**XGBoost**	AdaBoost	Random Forest	XGBoost	**AdaBoost**
1	1.0	**0.88**	1.0	1.0	**1.0**	1.0	0.73	0.67	**0.6**
2	0.95	**0.97**	0.81	0.93	**0.88**	0.89	0.75	0.78	**0.81**
3	0.85	**0.75**	0.65	0.89	**0.89**	0.89	0.5	0.50	**0.77**
4	0.50	**0.64**	0.58	-	**-**	-	0.3	0.75	**0.65**
5	0.47	**0.58**	0.56	0.52	**0.61**	0.68	0.62	0.63	**0.65**
6	0.60	**0.49**	0.83	0.67	**0.64**	0.74	0.33	0.25	**0.58**
7	0.85	**0.90**	0.84	0.80	**0.85**	0.75	0.81	0.78	**0.76**
8	0.79	**0.82**	0.72	0.96	**0.95**	0.85	0.90	0.67	**0.92**
9	0.93	**0.94**	0.93	0.98	**0.88**	0.93	1.0	0.97	**0.97**
10	0.93	**0.8**	0.9	0.49	**0.58**	0.33	0.83	0.75	**0.71**
11	0.60	**0.79**	0.43	0.54	**0.59**	0.53	-	-	**-**
Average	0.77	**0.78**	0.75	0.78	**0.79**	0.76	0.68	0.67	**0.74**

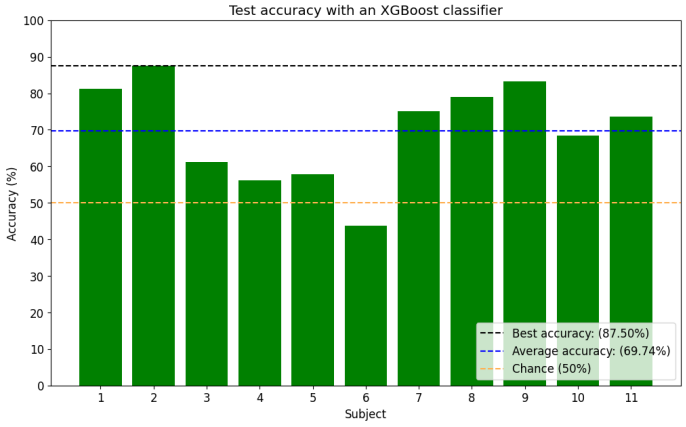

Fig. 5. Test accuracy at 5Hz presentation rate.

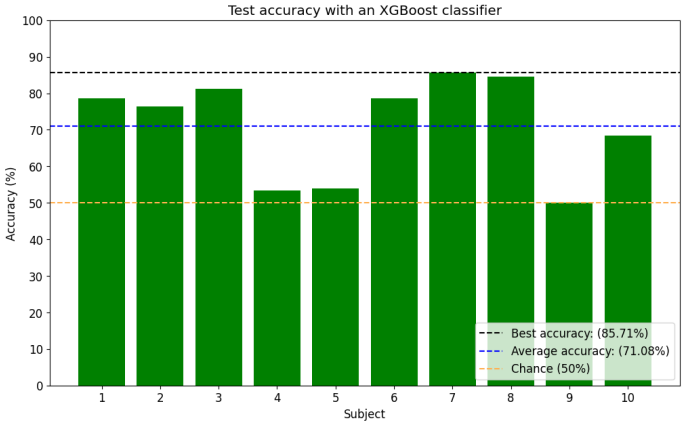

Fig. 7. Test accuracy at 6Hz presentation rate.

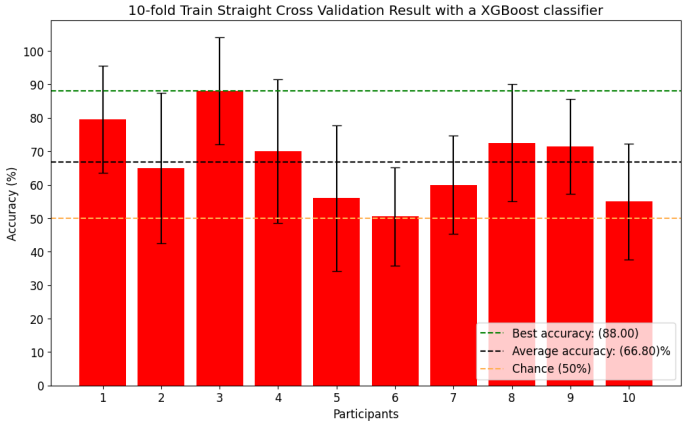

Fig. 6. Cross-validation accuracy at 6Hz presentation rate.

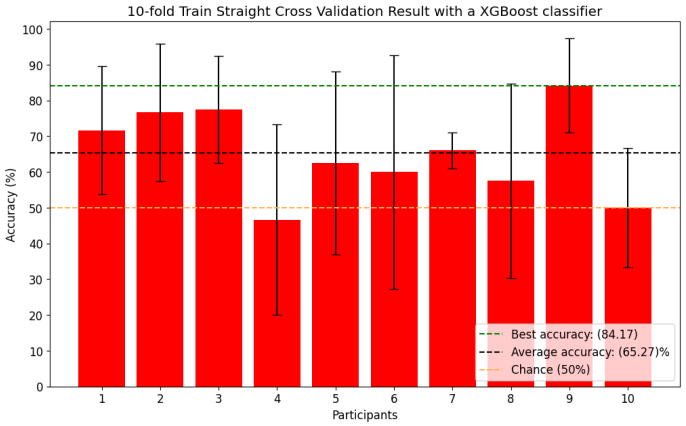

Fig. 8. Cross-validation accuracy at 10Hz presentation rate.

accuracies in the testing set with subjects 1,2,3, 6, 7, and 8 obtaining scores equal or superior to 80%.

VI. CONCLUSIONS

With BCI technology peaking in 2023, several applications that can exploit brain signals have emerged with the promise to aid human life. In this report, the brain inner model related to the identification and localization of target objects in images (N2pc) was exploited to build a pipeline for a BCI that will be further developed during the Master Thesis in Biomedical Engineering. The pipeline was constructed using an online dataset and the original study where the data is inserted.

Five binary classification models were used and compared to drive the discrimination between target objects present in

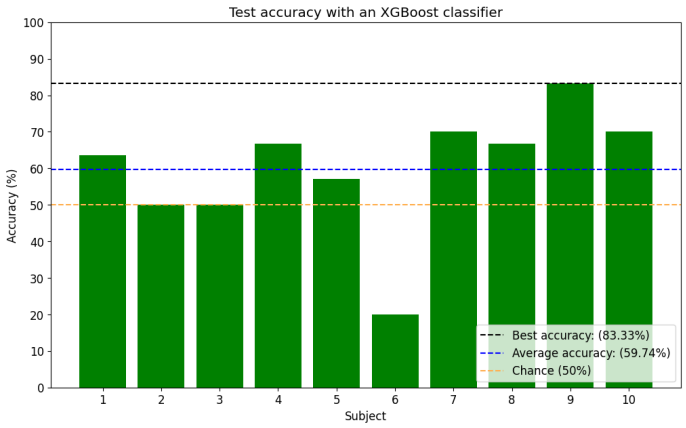

Fig. 9. Test accuracy at 10Hz presentation rate.

the Left and Right visual fields using only the N2pc potential. By analyzing the AUC scores associated with the training of the models, a few improvements were noticed when using the XGBoost model.

The aim of the thesis is to build a BCI system capable of localizing a target in images, not only horizontally but also vertically, to direct a virtual drone to the intended location in the environment. This report served as a great tool to drive the further development of the system by addressing some of the concepts of target localization using only EEG signals. However some promising results were achieved, many improvements could still be made. Because BCI systems are so subject-dependent, one approach that could improve the results is using a feature selection algorithm like the Mutual Based Information to feed the most discriminant features of each individual to the models.

References

[1] Robert Leeb, Luca Tonin, Martin Rohm, Lorenzo Desideri, Tom Carlson, and José Del R. Millán. Towards independence: A BCI telepresence robot for people with severe motor disabilities. *Proceedings of the IEEE*, 103(6):969–982, jun 2015. https://doi.org/10.1109/JPROC.2015.2419736 doi:10.1109/JPROC.2015.2419736.

[2] BCIs for DOC Patients: Assessment, Communication, and New Directions. pages 62–71, 2016. URL: https://link.springer.com/chapter/10.1007/978-3-319-40244-4$_7$.

[3] Athanasios Vourvopoulos and Sergi Bermúdezi Badia. Motor priming in virtual reality can augment motor-imagery training efficacy in restorative brain-computer interaction: A within-subject analysis. *Journal of NeuroEngineering and Rehabilitation*, 13(1):1–14, aug 2016. https://doi.org/10.1186/S12984-016-0173-2/TABLES/2 doi:10.1186/S12984-016-0173-2/TABLES/2.

[4] Thorsten O. Zander and Christian Kothe. Towards passive brain-computer interfaces: Applying brain-computer interface technology to human-machine systems in general. *Journal of Neural Engineering*, 8(2), 2011. https://doi.org/10.1088/1741-2560/8/2/025005 doi:10.1088/1741-2560/8/2/025005.

[5] Liam Drew. Decoding the business of brain–computer interfaces. *Nature Electronics 2023 6:2*, 6(2):90–95, feb 2023. URL: https://www.nature.com/articles/s41928-023-00929-9, https://doi.org/10.1038/s41928-023-00929-9 doi:10.1038/s41928-023-00929-9.

[6] Chang S. Nam, Inchul Choi, Amy Wadeson, and Mincheol Whang. Brain–Computer Interface: An Emerging Interaction Technology. In Chang S. Nam, Anton Nijholt, and Fabien Lotte, editors, *Brain–Computer Interfaces Handbook*, chapter 1, pages 11–52. CRC Press, Boca Raton, 1st edition, jan 2018. https://doi.org/10.1201/9781351231954-1 doi:10.1201/9781351231954-1.

[7] Robotic Arm with Brain – Computer Interfacing. *Procedia Technology*, 24:1089–1096, jan 2016. https://doi.org/10.1016/J.PROTCY.2016.05.241 doi:10.1016/J.PROTCY.2016.05.241.

[8] Brain–Computer Interface Spellers: A Review. *Brain Sciences 2018, Vol. 8, Page 57*, 8(4):57, mar 2018. URL: https://www.mdpi.com/2076-3425/8/4/57/htm https://www.mdpi.com/2076-3425/8/4/57, https://doi.org/10.3390/BRAINSCI8040057 doi:10.3390/BRAINSCI8040057.

[9] Review of real brain-controlled wheelchairs. *Journal of neural engineering*, 13(6), oct 2016. URL: https://pubmed.ncbi.nlm.nih.gov/27739401/, https://doi.org/10.1088/1741-2560/13/6/061001 doi:10.1088/1741-2560/13/6/061001.

[10] Kyungho Won, Moonyoung Kwon, Minkyu Ahn, and Sung Chan Jun. EEG Dataset for RSVP and P300 Speller Brain-Computer Interfaces. *Scientific Data 2022 9:1*, 9(1):1–11, jul 2022. URL: https://www.nature.com/articles/s41597-022-01509-w, https://doi.org/10.1038/s41597-022-01509-w doi:10.1038/s41597-022-01509-w.

[11] Ingon Chanpornpakdi, Motoi Noda, Toshihisa Tanaka, Yuval Harpaz, and Amir B. Geva. Clustering of advertising images using electroencephalogram. *Proceedings of 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC 2022*, pages 267–274, 2022. https://doi.org/10.23919/APSIPAASC55919.2022.9980161 doi:10.23919/APSIPAASC55919.2022.9980161.

[12] Evan Ackerman and Eliza Strickland. Are You Ready for Workplace Brain Scanning?: Leveraging brain data will make workers happier and more productive, backers say. *IEEE Spectrum*, 59(12):46–52, dec 2022. https://doi.org/10.1109/MSPEC.2022.9976479 doi:10.1109/MSPEC.2022.9976479.

[13] Stephanie Lees, Natalie Dayan, Hubert Cecotti, Paul McCullagh, Liam Maguire, Fabien Lotte, and Damien Coyle. A review of rapid serial visual presentation-based brain–computer interfaces. *Journal of Neural Engineering*, 15(2):021001, jan 2018. URL: https://iopscience.iop.org/article/10.1088/1741-2552/aa9817, https://doi.org/10.1088/1741-

2552/AA9817 doi:10.1088/1741-2552/AA9817.

[14] EEG Signals from an RSVP Task v1.0.0. URL: https://www.physionet.org/content/ltrsvp/1.0.0/.

[15] Ana Matran-Fernandez and Riccardo Poli. Towards the automated localisation of targets in rapid image-sifting by collaborative brain-computer interfaces. *PLOS ONE*, 12(5):e0178498, may 2017. https://doi.org/10.1371/JOURNAL.PONE.0178498 doi:10.1371/JOURNAL.PONE.0178498.

[16] M. Teplan. Fundamentals of EEG measurements. *Measurement Science Review*, 2(Section 2):1–11, 2002.

[17] Fabien Lotte, Chang S. Nam, and Anton Nijholt. Introduction: Evolution of Brain-Computer Interfaces. In Chang S. Nam, Anton Nijholt, and Fabien Lotte, editors, *Brain-Computer Interfaces Handbook: Technological and Theoretical Advance*, pages 1–11. Taylor Francis (CRC Press), Boca Raton, 1st edition, 2018. URL: https://hal.inria.fr/hal-01656743 https://hal.inria.fr/hal-01656743/document.

[18] Miguel Rocha e. Costa, Felipe Teixeira, and João Paulo Teixeira. Analysis of the Middle and Long Latency ERP Components in Schizophrenia. *Communications in Computer and Information Science*, 1488 CCIS:477–491, 2021. https://doi.org/10.1007/978-3-030-91885-$9_35doi : 10.1007/978 - 3 - 030 - 91885 - 9_35$.

[19] Margitta Seeck, Laurent Koessler, Thomas Bast, Frans Leijten, Christoph Michel, Christoph Baumgartner, Bin He, and Sándor Beniczky. The standardized EEG electrode array of the IFCN. *Clinical Neurophysiology*, 128(10):2070–2077, 2017. URL: http://dx.doi.org/10.1016/j.clinph.2017.06.254, https://doi.org/10.1016/j.clinph.2017.06.254 doi:10.1016/j.clinph.2017.06.254.

[20] Paul Sajda, Eric Pohlmeyer, Jun Wang, Barbara Hanna, Lucas C Parra, Shih-Fu Chang, P Sajda, E Pohlmeyer, J Wang, · S.-F Chang, S.-F Chang, B Hanna, L C Parra, D S Tan, and A Nijholt. Cortically-Coupled Computer Vision. *Human-Computer Interaction Series*, pages 133–148, 2010. https://doi.org/10.1007/978-1-84996-272-$8_9doi : 10.1007/978 - 1 - 84996 - 272 - 8_9$.

[21] Christoph S. Herrmann, Stefan Rach, Johannes Vosskuhl, and Daniel Strüber. Time-frequency analysis of event-related potentials: a brief tutorial. *Brain topography*, 27(4):438–450, 2014. URL: https://pubmed.ncbi.nlm.nih.gov/24194116/, https://doi.org/10.1007/S10548-013-0327-5 doi:10.1007/S10548-013-0327-5.

[22] Emily S. Kappenman and Steven J. Luck. The Oxford Handbook of Event-Related Potential Components. *The Oxford Handbook of Event-Related Potential Components*, pages 1–664, dec 2011. URL: https://academic.oup.com/edited-volume/34558, https://doi.org/10.1093/OXFORDHB/9780195374148.001.0001 doi:10.1093/OXFORDHB/9780195374148.001.0001.

[23] Yoshiyuki Kashiwase, Kazumichi Matsumiya, Ichiro Kuriki, and Satoshi Shioiri. Temporal Dynamics of Visual Attention Measured with Event-Related Potentials. *PLOS ONE*, 8(8):e70922, 2013. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0070 https://doi.org/10.1371/JOURNAL.PONE.0070922 doi:10.1371/JOURNAL.PONE.0070922.

[24] Mamunur Rashid, Norizam Sulaiman, Anwar P. P. Abdul Majeed, Rabiu Muazu Musa, Ahmad Fakhri Ahmad, Bifta Sama Bari, and Sabira Khatun. Current Status, Challenges, and Possible Solutions of EEG-Based Brain-Computer Interface: A Comprehensive Review. *Frontiers in Neurorobotics*, 14(June):1–35, 2020. URL: https://pubmed.ncbi.nlm.nih.gov/32581758/, https://doi.org/10.3389/fnbot.2020.00025 doi:10.3389/fnbot.2020.00025.

[25] J. Platt and Nikos Karampatziakis. Probabilistic Outputs for SVMs and Comparisons to Regularized Likelihood Methods. 2007.

[26] Zhi-Hua Zhou. Machine Learning. 2021. URL: https://link.springer.com/10.1007/978-981-15-1967-3, https://doi.org/10.1007/978-981-15-1967-3 doi:10.1007/978-981-15-1967-3.

[27] MNE — MNE 1.4.0 documentation. URL: https://mne.tools/stable/index.html.

[28] Calculation of Visual Angle. URL: http://www.yorku.ca/eye/visangle.htm.