# Action Anticipation from EEG Signals for Human-Robot Interaction

Rodrigo Miguel Vieira
*Departamento de Engenharia Eletrotécnica e de Computadores*
*Insituto Superior Técnico*
Lisbon, Portugal
rodrigo.m.vieira@tecnico.ulisboa.pt

***Abstract*—The number of context in which humans interact with robots is ever-growing. To achieve safe, reliable, and meaningful interaction between agents, the ability to anticipate each other's actions is key. A body of literature suggest electroencephalography (EEG) signals may provide a window into the processes that precede action in humans, allowing movement to potentially be predicted before motion begins. This thesis aims to provide a comparison between different approaches to the task of action anticipation from EEG signals by leveraging the versatility of Convolutional Neural Networks (CNNs). Additionally, the use of gaze features, also shown in previous research to exhibit specific patterns during action anticipation, is explored as a complement to EEG to reach this goal. A novel metric to quantify time advantage provided by an action anticipation system, while providing insight into classifier confidence level, is also proposed and evaluated in this work. Results indicate shallow, fully-convolutional architectures are able to achieve an accuracy of 88.00% (standard deviation of 13.81%) in action anticipation epoch classification, consistently classifying epochs as preceding action with a time advantage of 120 milliseconds, with minimal EEG signal processing, in an end-to-end approach. Leveraging gaze features, using a hybrid CNN classifier, can result in an accuracy of 93.14% (standard deviation of 15.18%) in this task. Ultimately, this thesis demonstrates the viability of the use of EEG signals for action anticipation in a physical human-robot interaction setting, describing and testing classifiers capable of anticipating human action from these readings in an accurate and timely manner.**

***Index Terms*—Brain-Computer Interfaces, Collaborative Interaction, Vision for Robotics, Cooperation and Coordination, Neural Systems.**

## I. Introduction

The ubiquitous presence of technology in modern life has sparked an interest in understanding how humans and machines can interact in productive, safe, and fulfilling ways.

Robots provide a new frontier for human-machine interaction. In recent decades, robots have become more common and more capable of performing a wide range of tasks: automated guided vehicles move packages around warehouses; industrial robots assemble complex machinery, such as cars; educational robots provide children with an introduction to programming and logic; service robots spare users the work of performing tedious chores; and these are but a few examples from the universe of potential applications.

As scenarios of human-robot interaction grow in number, so does the need to address the technical, ethical, economical, and sociological questions behind them. A constant area of research in this field is the search for new ways for humans and robots to collaborate in order to complete a task, to which the ability to anticipate human action is crucial.

Electroencephalography (EEG) provides us with a way of detecting action intention before its onset. EEG is a functional imaging technique with high temporal resolution, and has been shown to detect changes in brain waves preceding action by over 1 second [1]. Mobile EEG headsets and caps are now more common than ever; they are also relatively inexpensive, non-invasive, and considerably easier to set up than when first introduced.

As a result, EEG signals offer a promising source of data for achieving action anticipation in human-robot interaction applications. Modern Deep Learning techniques, such as Convolutional Neural Networks, can support multi-modal data, making EEG signals a potentially interesting complement to the current Computer Vision based methods.

### A. Problem statement

Strictly defining what constitutes Human-Robot Interaction (HRI) as the successive manipulation of the same object by a human and a robot, and having established action anticipation as playing a fundamental role in cooperation between agents, the motivation behind a system capable of accurately, and in a timely manner, predicting human limb motion becomes clear.

According to previous works, brain regions responsible for motor action exhibit distinct responses prior to motion onset, both during imagined and executed movement [1]–[4]. Additionally, pupil dilation and other gaze features have also been shown to provide useful information to predict human action [5].

Thus, given the prevalence of human-robot interaction scenarios, and having identified EEG signals as potentially providing useful information for action anticipation a physical HRI context, this thesis aims to answer the following questions:

- How do end-to-end EEG-based action anticipation approaches compare to those that employ feature extraction based on signal characteristics identified in literature?
- Does the use of gaze information significantly improve the performance of a classifier for this task, compared to using only EEG?

- Can classification performance be maintained using imagined, rather than executed, motion?
- What kind of classification performance can be achieved in the task of action anticipation using EEG signals, and how much time advantage does it grant?

### B. Contributions

The main objective of this thesis is to serve as a basis for the comparison of Machine Learning methodologies that use EEG signals to anticipate action, by providing a comparison between different Deep Learning, Convolutional Neural Network (CNN)-based approaches to the task of action anticipation from EEG signals; the resulting action anticipation model should also be capable of anticipating action within a short amount of time (within 100 milliseconds of action onset). By answering the stated research questions, this work provides insight into how these signals may be integrated into a robot's control and decision system to improve human-robot interaction. This document should, thus, provide a foundation to research aiming to use EEG to coordinate action between humans and robots, as well as an additional data-point in the growing field of research that is EEG-based action anticipation.

The main contributions of this thesis are:

- Describing and implementing an experimental human-robot interaction protocol for action anticipation research using EEG;
- Proposing a novel metric that combines classification time advantage with a measure of its consistency;
- Obtaining a dataset for action anticipation that combines time-labelled EEG signals and gaze information;
- Proposing, testing, and comparing the performance of sole EEG with hybrid EEG + gaze classifiers.

## II. Background

### A. EEG overview

The human brain possesses billions of neurons in constant communication with each other. The electrochemical impulses sent between the terminals and synapses of communicating neurons can be detected at the scalp through the use of EEG.

Through electroencephalography, it is possible to measure the involvement of specific regions of the brain when performing some brain function, such as face processing, memory recollection, or coordinating motor execution, to name a few.

Analysis of scalp level electrical signals led to the discovery of brain waves: oscillatory voltages originating from specific regions of the brain, which can be associated with a number of different states, actions, and emotions.

With regards to motor intention execution, the most notable of these brain waves is the $\mu$ rhythm, measured between 8 to 13 hertz, and usually detected over sensorimotor cortex. When humans perform or imagine movement, groups of neurons in the sensorimotor region desynchronize, leading to a disruption in the $\mu$ rhythm dubbed Event-Related Desynchronization (ERD). By analyzing the EEG signals from this region, a computer can then detect motor action intention, regardless of whether it was actually executed or simply imagined (the latter phenomenon is the basis for Motor Imagery (MI) Brain Computer Interface (BCI) paradigms).

### B. Action anticipation from EEG signals

In 1965, Kornhuber & Deecke published their findings regarding the detection of a readiness potential preceding voluntary movement [6]. This *bereitschaftspotential* (BP), referred to using the original German term, is made up of two main components: early BP, beginning approximately 1.5 seconds before movement onset, has a very low amplitude; late BP starts 0.5 seconds before action and exhibits greater positivity, making it easier to detect. This potential can be detected in electrodes placed over the Supplementary Motor Area and Primary Motor Area.

Since the BP presents a very low amplitude, in an environment where noise, interference and movement artifacts are prevalent, state-of-the-art action anticipation systems based on EEG signals tend to focus on analysis of ERD in the $\mu$ rhythm in across central electrodes.

Before movement onset, it is also possible to identify Event-Related Spectral Perturbation (ERSP) starting at roughly the same time as the BP, even in people suffering from certain neurological conditions, such as Parkinson's disease [7]. These phenomena consist of contralateral-dominant $\alpha$ ERD and $\beta$ Event-Related Synchronization (ERS) before movement onset.

### C. Analysis of Electroencephalography Signals

While the applications of different BCI paradigms may differ, as well as specific system implementation details, certain considerations related to EEG signal analysis must always be made, informing the methodology proposed in section III.

Designing an EEG signal analysis and classification pipeline requires a mindful examination of the following aspects: spatial filtering and source localization; noise and artifact removal; and the windowing/epoching performed on the signal.

There are several ways of performing spatial filtering, ultimately obtaining a more robust, lower dimensionality dataset, such as by employing a **Common Average Reference (CAR)**, where the average potential of all electrode channels is subtracted from each channel to average the model error and prevent biasing towards a specific region [8], and **Independent Component Analysis** (ICA) for source localization and in-band noise and artifact removal [9].

Along with the desired EEG signals, EEG sensors will detect a host of other sources: power grid interference in the 50-60 hertz range, electrooculography (EOG) and electromiography (EMG) activity, from eye and muscle movements, and EEG signals from distant parts of the brain, known as volume conduction [10]. In order to extract useful information from this data, it is thus necessary to pre-process the signals in order to reduce noise, by applying band-pass filters, or through ICA, to deal with in-band noise and artifacts.

Certain features which are important for event detection, such as bandpowers during MI trials, require an analysis of frequency-domain features, rendering the direct application

of Machine Learning algorithms for time series classification inadequate. This motivates the use of windowing/epoching techniques, diving the signal into time windows: these should, at the very least, be shorter than the minimum interval between stimuli, but there are lower bounds to epoch length, due to edge effects and relevant slow-cortical potentials. To diminish these effects, the epoch length chosen for action anticipation systems is usually in the 250 to 500 millisecond range.

### D. Machine Learning for action anticipation

Machine Learning consists of the design, training, and evaluation of models capable of representing patterns "hidden" within data, dispensing the need for explicit development of algorithms and models to solve specific problems.

One common application of Machine Learning methods is to solve the task of data classification, extracting patterns from data and attempting to label them as belonging to some target class. When the target classes are defined *a priori*, and the classification of a set of data points is known, we are in the presence of a supervised learning classification problem.

The challenge of action anticipation from EEG can be cast as a Machine Learning classification problem: in this case, we want to label a specific segment of data (an epoch) as either preceding action or not. The Machine Learning model is trained on classified epochs, with the goal of learning which features can be used to distinguish action anticipation epochs from resting epochs.

*1) Convolutional Neural Networks:* Convolutional Neural Networks (CNNs) were inspired by human vision processing, introducing specialized layers capable of interpreting and classifying images and other large inputs with much fewer training parameters than a fully-connected Neural Network using each pixel value as an input feature. An added benefit of a simpler model is also the prevention of overfitting, which may occur when an overly complex model fully learns its training dataset, but becomes less suited to the classification of data outside this set.

A key difference in the structure of a CNN relative to other Neural Networks is that its neurons organize in a three-dimensional scheme: height by width by depth[1]. CNN consist of a sequence of three types of layers: convolutional layers, pooling layers, and fully-connected layers.

**Convolutional layers** determine the output after applying a convolution to their input with a trainable kernel (also called filter). This kernel is a small matrix of weights spread along the input depth, filtering the input image into an activation map. For each kernel in the layer, an output will be produced, which is then stacked to form the output volume. An important assumption is made here: a feature that is useful in some region of the image is assumed to be useful anywhere else in the image. It is this assumption that allows each depth slice of the input to simply be convolved with a kernel, massively reducing the number of free parameters, as all neurons in the same depth layer will share the same weights. To each kernel is also associated a bias, with the same spatial dimensions as the output.

As a result of this operation, the number of convolutional layers an image can pass through is limited, since its size will decrease with each such feedforward pass. When the final output is produced, the error is backpropagated through the network and used to update the kernel values with some variant of gradient descent. A particularly capable implementation of this method was proposed by Kingma et al. in 2014 with "Adam" [11], a method which makes use of exponentially decaying influence of previous first and second order gradients to update the model parameters.

We may also define three hyperparameters to optimize the performance of our CNN: depth, or the number of kernels in each layer; stride, which defines how the kernel is successively placed along the input image during convolution; and zero-padding, which pertains to the padding of the input border with 0's, allowing for output dimensionality to be controlled.

**Pooling layers** aim to reduce the dimensionality of the representation within the neural network. To do this, a specific operation, such as calculating the maximum, or average, is applied to a neighbourhood of pixels. Due to the destructive nature of this operation, and considering convolutional layers also reduce the dimensions of their input, use of pooling must be done with care. Use of pooling layers summarizes features and reduces the computational burden on deeper layers of the network.

**Fully-connected layers** are responsible for producing decisions based on the activation maps generated by convolutional and pooling layers. Within these layers, neurons have access to all activations in the previous layer. The final layer, constituted of as many neurons as there are target labels, outputs the final classification result.

Generally, the structure of a CNN is as follows: the input image is fed to a stack of convolutional layers, followed by ReLU activation, and a pooling layer; this layer sequence may be repeated several times. The resulting activations are then used as input to one or several fully-connected layer, and the final output is generated by a fully-connected output layer with sigmoid or *softmax* activation.

### E. Related work

*1) Research on action anticipation from EEG signals:* Several solutions to the problem of anticipating human action from EEG signals have been proposed. Planelles et al. [3] propose a methodology for detection and anticipation of arm movement using the sum of band powers in three frequency bands involved in ERSP. Utilizing very basic features, and traditional Machine Learning classifiers, such as Support Vector Machines, this approach is capable of achieving a high classification True Positive rate, at 70%, but also presents a high false positive rate, at 28%, suggesting the use of a more sophisticated pipeline may yield better results and more reliable classification.

[1]Unlike usual terminology for Neural Networks, this "depth" is the third dimension of the activation volume, or output, of a convolutional layer, rather than the number of layers in the model.

Lew et al. [2] base their approach on detection of the BP to anticipate self-paced reaching motion. To achieve this, the methodology relies on auxiliary EOG and EMG sensors, which are leveraged during pre-processing to help remove movement artifacts without affecting the low-frequency BP. While this approach reaches high levels of performance, with a True Positive rate of $81 \pm 11\%$, maintained even in stroke patients, the careful pre-processing performed, as well as the need for additional sensors, is not compatible with real-time decision scenarios in HRI.

The work of Buerkle et al. [4] employs a modern Deep Learning Classifier: the Long Short-Term Memory Recurrent Neural Network (LSTM-RNN). This classifier is trained on a (visually) selected EEG channel to detect action anticipation epochs in an End-to-End approach, achieving high accuracy ($85$ to $92\%$), and a time advantage of up to 513 milliseconds. The need to employ a visual channel selection, as well as other intermediate steps that require arbitrary evaluations by the system designer, are the main drawbacks of this approach.

*2) Convolutional Neural Networks in EEG action decoding:* Due to their powerful, automatic feature extraction from raw data, CNNs are particularly interesting for the task of EEG signal decoding and classification. Generally, EEG signal decoding using these Deep Learning classifiers can be done using three different ways to input data [12]: using the raw EEG signal, in all its channels, after applying minimal pre-processing; performing some transformation to the EEG signal, such as time-frequency decomposition; or by transforming and selecting only some of the EEG channels.

*3) Action Anticipation using Gaze Features:* When interacting in a shared environment, humans interact through both verbal, and non-verbal cues, such as through their gaze, or certain movements, implicitly communicating their ensuing actions to those around them. This has been shown by Cannon et al. [13], who demonstrate action anticipation involves the representation of self-produced action, resulting in prospective gaze when performing motor tasks. As a result, human gaze encodes information that can be used to anticipate action in a collaborative environment.

A particularly context-independent prospective gaze feature is pupil diameter, as well as features derived from it. According to Kahneman [14], changes in pupil diameter reflect cognitive load due to links between ocular muscles and neurotransmitters, and changes in pupil diameter have been shown to be a valid measure of cognitive load in response preparation [15]. Based on these findings, Naber et al. [16] demonstrate analysis of pupil diameter allows visual task performance to be anticipated by as much as 500 ms.

Combining pupil diameter information with EEG is not a novel idea: Rozado et al. [5] explore features derived from pupil diameter readings during MI-BCI as a way to improve performance, finding trial mean diameter to significantly distinguish action anticipation and baseline trials.

## III. Methodology

This section describes the methodology applied to develop an action anticipation classification pipeline from EEG signals and gaze information. The chapter is divided into four subsections: pre-processing of data in subsection III-A; classification model design and implementation in subsection III-B; a novel time advantage metric is proposed, and its evaluation is described in III-C; data analysis and validation steps in subsection III-D.

### A. Time-series pre-processing

The time series action anticipation classification pipeline can be divided into two stages: data pre-processing, to prepare data for classification through noise filtering and feature generation/extraction, and classifier training and evaluation.

To prepare EEG data for classification, the following steps were followed: a one-pass, zero-phase, non-causal highpass filter was applied at 0.1 hertz to remove baseline drift; a one-pass, zero-phase, non-causal notch filter was applied at 50 hertz to remove grid interference; Common Average Referencing was applied, in order to prevent the signal from being biased towards specific areas over the scalp; Independent Component Analysis was then performed on the signal to remove muscle artifact components.

The signal was then epoched, selecting the intervals spanning $[0, 1]$ (containing action anticipation) and $[-1, 0]$ (no action anticipation) seconds relative to arrow cue appearing on screen. The epochs were then processed using a Short-Time Fourier Transform, preserving temporal information. The STFT was applied to each channel separately, using a sine window of 64 samples in width, with a time step of 16 samples between windows.

The use of CNN also motivates the introduction of a second approach, with no feature extraction step. Since Deep Learning classifiers are capable of identifying and extracting useful features, these can potentially select useful features on their own (end-to-end approach). As such, a parallel pipeline was implemented and tested, without performing STFT on the epoched signal before classification.

Prior to classification, vision pupillometry data was also pre-processed. Since the eye-tracker used in our experiments only transmits data from one eye at a time, an assumption was made that pupil diameter is the same for both eyes, and thus a unique value for this measurement could be obtained by combining data from both eyes. The resulting time series was also heavily corrupted by noise and artifacts. To mitigate this issue, the following steps were taken: the top envelope of the diameter measure was estimated based on local (within a 10 sample range) maxima, and any outliers (at least two standard deviations from mean) were dropped from the time series; then, the signal was recreated at a constant 50 Hz frequency using Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) interpolation; and the resulting signal was filtered using an Exponentially Weighted Moving Average with an $\alpha$ value of 0.1.

The pre-processed signal was then divided into epochs as done with EEG signals. In addition to this, the three most relevant pupillometry features identified by Rozado et al. [5] during their study on BCI-system performance improvement using pupil diameter information were also computed and appended to the signal: the diameter mean, derivative, and maximum for each epoch.

### *B. Classification models and implementations*

The classification models described, trained, and evaluated in this thesis were implemented in Python (version 3.11.4) using the *tensorflow* package (v2.12.0) (*keras* backend) for the Deep Learning classifiers. Pre-processing was done using the *mne-python* (v1.4.2) and *scipy* packages.

The following classifiers were evaluated (names used to refer to them henceforth and in figures in quotation marks): "MLP" - Multilayer Perceptron; "CNN" - Deep CNN; "CNN_shallow" - Shallow, fully convolutional CNN. In addition to these, EEGNet [12] was also tested using raw EEG information, under the denomination "CNN_EEGNet".

The **MLP** classifier was tested to offer a baseline to compare the more sophisticated, State-of-the-Art CNNs against. This classifier features two fully-connected layers with ReLU activation, with $U$ units each, and an output unit using sigmoid activation. Use of dropout layer between each fully-connected layer was also tested.

The **CNN** classifier features a convolutional block and a fully-connected block. The convolutional block is constituted by a stack of two 2-D convolutional layer and maximum pooling ($2 \times 2$) layer sequences; each convolutional layer features $F$ filters, $k \times k$ kernel size, and ReLU activation. The fully-connected block has two fully-connected layers with $U$ units and ReLU activation, followed by an output unit with sigmoid activation. Use of a dropout layer between convolutional layers, and between fully-connected layers, was once again tested.

The **CNN_shallow** classifier is fully-convolutional, based on the shallow architecture featured in the work of Schirrmeister et al. [17]. This classifier features a single convolutional layer, $F$ filters, $k \times k$ kernel, followed by a maximum pooling layer ($10 \times 10$), ending with an output unit using sigmoid activation.

Finally, the **CNN_EEGNet** classifier is implemented as proposed by Lawhern et al. [12], with parameters $F1$ and $D$ evaluated and ajusted to optimize performance. This classifier is a deep CNN, tailored to the analysis and classification of EEG signals, featuring Batch Normalization, as well as depthwise and separable 2-D convolutions, and presenting State-of-the-Art performance in BCI single-trial classification.

The "Adam" optimizer [11] was used with a binary cross-entropy loss. Each classifier was trained for a single condition (MOVE/NOMOVE) and subject. The trained models were saved, and then evaluated on a validation set of EEG epochs, randomly sampled from all epochs (20% validation split). Model performance was evaluated based on 4 different metrics: accuracy, precision, recall, and F1-score (harmonic mean of precision and recall), over 30 total runs for each condition. For each model parameter, a performance analysis was conducted, in order to select the most adequate values.

*1) Action anticipation from Gaze:* Action anticipation classification using pupil diameter was tested using the MLP, CNN, and CNN_shallow classifiers. Since this is a single time feature, classification models have to be slightly adjusted: CNN-based classifiers used a 1-dimensional convolution, rather than 2-dimensional. Classifier architectures remained otherwise unchanged, with model parameters once again selected based on classification performance metrics.

*2) Hybrid classification:* To understand the potential of introducing gaze features, namely pupil diameter, as a complementary signal to EEG for action anticipation, the deep learning classifiers, MLP, CNN, and CNN_shallow, were extended to allow for the parallel classification of EEG and gaze data.

To do this, CNN-based classifiers were redesigned to feature two parallel convolutional blocks, whose output was flattened and used as input to the fully-connected dense layers. The MLP classifier was redesigned into two parallel fully-connected towers, whose output is given to a single unit to generate the classification label.

Model parameters for each tower were retained from the previous parameter selection.

*3) Model parameter selection:* Model parameters were selected based on "per-parameter" performance evaluation: for each parameter, an evaluation was conducted on the MOVE condition (all subjects), and the parameter with the highest sum of accuracy and F1-Score was selected.

To achieve a balance between classifier bias and variance, model parameter selection evaluates performance as complexity increases, identifying the point at which the model begins overfitting (drop in performance) or reaches a state of diminishing returns (classification performance increases more slowly than complexity).

### *C. Time Advantage Estimation*

The task of action anticipation seeks to provide a robotic control and decision system with a time advantage relative to the onset of human motion. However, in literature, examples of a unifying time advantage metric are scarce, and generally context-specific. The Total Time Advantage (TTA) a classification system has, relative to movement onset, can be thought of as the difference between the computational time the system requires to output a decision (henceforth Computational Time Delay (CTD)) and the time advantage of the earliest sample the classifier consistently labels as preceding action, which we will refer to as Decision Time Advantage (DTA)-

Obtaining a measure of CTD is straightforward: the computation time spent applying pre-processing and making a prediction for each epoch must be tallied, and a distribution may then be estimated.

Defining the DTA requires handling the fuzziness of the concept of "reliable action anticipation", striking a balance

between optimism and pessimism, and indicating how reliably a certain result may be achieved.

We may thus combine the DTA with a measure of classification consistency. As such, we will define our DTA based on a consistency parameter $k$, such that our metric $\text{DTA}_k$ becomes: the earliest point in time at which a sample is classified by the system as anticipating action at least $k\%$ of the time, and after which the positive prediction rate remains above $k$ until movement onset. This metric is only defined for $k$ between the False Positive Rate and the True Positive Rate.

To estimate this value, a sliding window was moved from resting epochs to anticipation epochs in the validation set in 0.1 second time increments; all pre-processing steps were applied to samples generated this way, which were then labelled by the trained classifiers for each subject and condition; this process was repeated over thirty runs. The positive prediction rate at each of these time increments was computed, and PCHIP interpolation was performed to produce a spline which could be used to obtain an estimate of $\text{DTA}_k$ for each classifier at several $k$ values.

### D. Data Analysis and Validation

Apart from classification, EEG data was also analysed, based on literature on BCI and neurophysiology, to quantify and evaluate features that could impact the task of action anticipation. The focus of this analysis resided on three features in particular: the presence of an $\alpha$ rhythm peak in each subject's EEG recording, which was used as a proxy for BCI-illiteracy; ERSP during action anticipation, in the $\alpha$ and $\beta$ bands, to evaluate the magnitude and consistency of these phenomena that form the basis of MI-BCI, and could potentially inform action anticipation classifiers.

*1) Presence of alpha peak:* In BCI applications, a significant portion of subjects (10 to 30%) demonstrate notably poorer performance relative to the rest of the population. Research into these individuals, usually referred to as "BCI-illiterate", suggest there are specific EEG signal patterns that can be used to identify them with a reasonable degree of confidence. During Motor Imagery tasks, Ahn et al. [18] link this condition with high theta and low alpha powers across different mental states.

To evaluate the presence of an $\alpha$ rhythm peak, the PSD of central channels C3 and C4 was computed using Welch's method [19] (Hamming window, 64 sample size, 31 sample stride), during resting and anticipation periods.

*2) Event-Related Spectral Perturbation:* According to literature, an ERSP should be noticeable around the central electrodes during action anticipation, becoming more pronounced upon onset of movement/motor imagery tasks. This perturbation may be a desynchronization in the $\alpha$ frequency around the sensorimotor area (ERD) and/or a synchronization in the $\beta$ frequency (ERS). These phenomena should be visible as a reduction or increase in band power relative to baseline, for ERD and ERS, respectively, around the central electrodes.

To evaluate the presence of ERSP during action anticipation and task execution, the Morlet Wavelet transformation of signals from central channels C3, C4 was computed for the $\alpha$ (f = $[7, 13]$ hertz, $0.1$ hertz resolution) and $\beta$ (f = $[15, 30]$ hertz, $0.1$ hertz resolution) bands around action anticipation epochs, integrating over the frequencies, and expressed as a percentage of average baseline band power.

### E. Experiments

To obtain multimodal data to develop and support this thesis, a human-robot interaction experiment was conducted. During this experiment, subjects interacted with a robotic arm in a pick-and-place task, while their EEG signals and eye movements were recorded. Timed visual cues were shown to the participants during the experiment, allowing for a period of action anticipation to be isolated, analyzed, and eventually classified. This section describes the experiment that was carried out (subsection III-E1), the materials used (subsection III-E2), as well as a dataset obtained from a similar experiment, used to validate the analysis and classification pipeline (subsection III-E4).

*1) Experiment design:* Five subjects were recruited to take part in the experiment, each providing their informed consent. The experiment consisted of two conditions: Motor Imagery ("NOMOVE"), and Motor Execution ("MOVE"), both performed during the same session. During each condition, subjects performed a total of 16 trials, with 8 trials for each arm, in a randomized order. The order in which subjects performed each condition was also random: the initial condition was randomly selected, its trials were performed in succession, and then followed by the other condition's trials. For this experiment, participants were wearing an EEG cap, as well as an eye tracker.

Participants were sat in front of a robotic arm, behind which was a computer screen displaying "Graz" paradigm cues [20], and were asked to remain motionless, with each hand placed forward, upon a table in front of them. An object was placed centrally, between the robot and the subject, with markers indicating its 3 possible positions: 'L'- left, 'C'- central, 'R'- right. Before each experiment, a trial run was conducted during which study participants could prepare for the experimental tasks, ensuring they understood the experimental paradigm and could interact safely with the robot.

Experiments began with a 15 second baseline/setup wait. During each trial, lasting 18 seconds, a cross would appear on screen to signal the start for 4 seconds; a red arrow would then point to either side of the screen for 1 second; subjects were instructed to, upon disappearance of the arrow, place the object in front of them in the marker corresponding to the side the arrow had pointed, using the arm on that side. during Motor Execution condition, or imagine the aforedescribed movement, during Motor Imagery. The robot would then pick up the object on that side, and place it on its central position once again, coinciding with the end of the trial; during Motor Imagery condition, the object remained in its central position for the duration of the trials. Trials were followed by an inter-trial period lasting 4 seconds. Each condition lasted for around 8 minutes.

*2) Materials:* The following materials were used during the experiments:

- Brain Products GmbH actiCAP 32 electrode cap [21], with gel-based active electrode system;
- Brain Products GmbH LiveAmp 32 wireless signal amplifier [22], with a 500 hertz sampling frequency, 32 electrode channels + 3 accelerometer channels (built into the amplifier);
- Kinova Gen3 ultra lightweight robot arm;
- PupilLabs Pupil Core eye tracker [23], with 120 hertz binocular eye camera, wide-angle lens world camera.

The robotic arm was controlled using a Python script via the Kinova Kortex-API. Graz paradigm visual cues and timing markers were generated using neuXus [24]. Amplifier readings were relayed to a LabStreamingLayer [25] stream using the LiveAmp LSL Connector app, and eye tracker data was sent to an LSL stream using the Pupil Capture LSL plugin. Data streams for the amplifier, eye tracker, and neuXus markers were recorded from LabStreamingLayer streams using LabRecorder, and stored as an XDF file, synchronized over the local LSL clock.

*3) Subject description:* A total of 5 subjects were recruited for this experiment, referred to by numbers 1 through 5 to preserve anonymity. Of these, 4 subjects were male and 1 was female; all subjects were right-handed, as per the Edinburgh Handedness Inventory [26]; none of the subjects had been previously diagnosed with a chronic neurological condition; subjects 1 and 3 had previous experience with MI-BCI.

*4) Validation dataset experiment description:* The "MI-EEG Dataset during Arm Control" was obtained by Farabbi et al. [27] to support a thesis on robot control using a BCI system [28]. As explored in section II, action anticipation phenomena occur even during MI, making a dataset such as this a valid proxy for studying action anticipation using EEG. For this reason, this dataset was used to validate prototypes of the data analysis and classification pipeline.

## IV. Results

This chapter presents the results of data analysis and action anticipation classification, performed according to section III. The chapter begins with a description of data analysis results in subsection IV-A; this is then followed by the model parameter selection results, in subsection IV-B; after this, the main action anticipation results are presented, in subsection IV-C.

### A. Data analysis results

*1) Presence of $\alpha$ rhythm peak:* The PSD averaged over the central electrodes C3, C4, during anticipation and rest epochs, is presented in figure 1. This analysis shows all subjects, apart from subject 1, exhibit a clear $\alpha$ frequency peak, suggesting this subject may be BCI-illiterate. Another notable finding is a subtle, but clear, reduction in $\alpha$ band power during anticipation epochs, when compared to resting state, for subjects 3, 4, 5, suggesting the occurrence of ERD, as expected from literature.

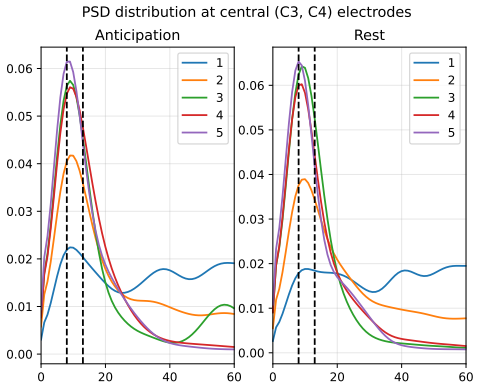


Fig. 1. Power Spectral Density at central electrodes during anticipation, resting epochs, per subject, expressed as a percentage of total power as a function of frequency. $\alpha$ frequency range between dashed lines.

*2) Event Related Spectral Perturbation:* The Event Related Spectral Perturbation during action anticipation and task execution, for the $\alpha$ and $\beta$ bands, is shown in figure 2 and figure 3 (subject average, MOVE condition), respectively for each band. Literature suggests there should be a decrease in $\alpha$ band power relative to baseline (ERD), and an increase in $\beta$ band power (ERS), during action anticipation and task execution. These perturbations should also be greater in the electrode contralateral to the task performed or anticipated.

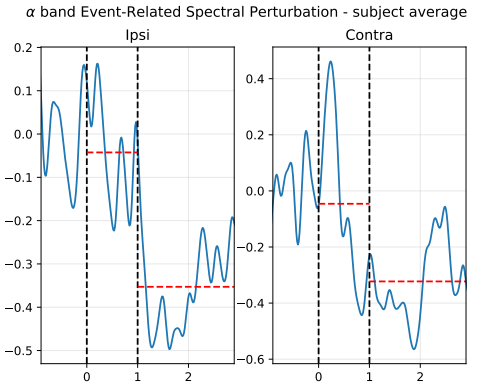


Fig. 2. Alpha ($\alpha$) band power relative to baseline at ipsi/contralateral electrodes before, during, and after action anticipation (between dashed vertical lines). In red dashed lines (horizontal), average band power during action anticipation and task execution epochs.

Indeed, when averaged across all subjects, $\alpha$ ERD can be noticed. However, this phenomenon is not very pronounced during the action anticipation epoch (between vertical dashed lines), only becoming more evident during task execution, as would be expected from literature. Additionally, while figure 3 may suggest the presence of $\beta$ ERS, an inspection of per-subject ERSP revealed the increase in $\beta$ band power is a result of artifacts in the high $\beta$ band in some subjects, rather than a broad, consistent ERS, as well as showing considerable inter-subject variability.

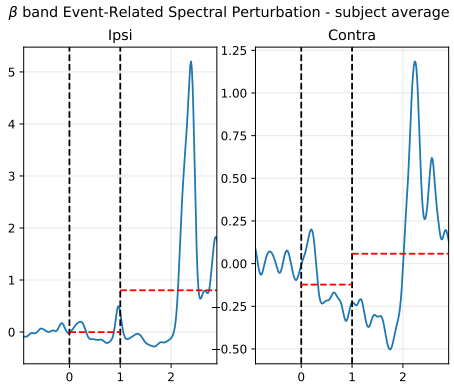


Fig. 3. Beta ($\beta$) band power relative to baseline at ipsi/contralateral electrodes before, during, and after action anticipation (between dashed vertical lines). In red dashed lines (horizontal), average band power during action anticipation and task execution epochs.

### B. Model parameter selection

*1) Short-Time Fourier Transform feature classifiers:* Training of the MLP classifier involves the selection of 3 model parameters: the number of units in each fully connected layer $U$, the dropout rate $D_r$, and the number of training epochs. Final MLP parameters selected for STFT feature classification were $U$ **= 128,** $D_r$ **= 0, 50 training epochs**.

The architecture of the CNN classifier demands the selection of a larger number of parameters: the number of filters in each convolutional layer $F$, the size of the convolutional kernel $k$, the number of units in fully-connected layers $U$, the dropout rate between layers $D_r$, and the amount of training epochs. The final parameters chosen for the CNN classifier based on STFT features were $F$ **= 32,** $k$ **= 5,** $U$ **= 32,** $D_r$ **= 0, and 75 training epochs**.

The CNN_shallow classifier, with no fully-connected layers and a shallow configuration, requires only the evaluation of 3 parameters: the number of filters in the convolutional layer $F$, the kernel size $k$, and the number of training epochs. Parameters determined to yield the best performance for the CNN_shallow classifier were $F$ **= 64,** $k$ **= 2, and 150 training epochs**. Remarkably, unlike the deeper CNN, this classifier sees better performance for lower values of kernel size.

*2) End-to-End classifiers:* Model parameters chosen for the MLP classifier during end-to-end classification were $U = $ **128 units, no dropout (**$D_r = $ **0), and 50 training epochs**.

The most performance-impacting parameters in the analysis of CNN classifier performance were the number of units in fully-connected layers $U$, and the dropout rate, $D_r$. Final chosen parameters for this classifier during the end-to-end approach were $F$ **= 4,** $k$ **= 9,** $U$ **= 32, no dropout, 600 training epochs**. In the case of the CNN_shallow classifier, the best performance results were achieved using $F$ **= 128,** $k$ **= 2, 600 training epochs**.

The CNN_EEGNet classifier, as proposed by Lawhern et al. [12], saw little effect on action anticipation performance as a result of changes to its parameters $F1$ and $D$. Parameters chosen for this model, as a result, were the values $F$ **= 12,** $D$ **= 2**.

*3) Gaze classifiers:* Using gaze features, the MLP classifier presented the best performance metric using parameters $U$ **= 64, no dropout, and 25 training epochs**. Parameter changes had little effect on classifier performance.

The CNN classifier was implemented with parameter $F$ **= 32,** $k$ **= 10,** $U$ **= 32, 250 training epochs, and** $D_r$ **= 0.3**. Once again, the effect of classifier design parameters on performance was muted.

Finally, parameters chosen for the CNN_shallow classifier using gaze features were $F$ **= 32,** $k$ **= 10, and 500 training epochs**.

### C. Action anticipation results

This section presents the main results of this thesis, divided by classification approach. For each method, classification performance metrics accuracy and F1-score are presented considering all subjects, for MOVE (motor execution, table I) and NOMOVE (MI, table II) conditions. Additionally classifier $\text{DTA}_k$ at 75% and 90%, are displayed in table III.

| | | Accuracy | | F1-Score | |
|---|---|---|---|---|---|
| Approach | Classifier | Mean | $\sigma$ | Mean | $\sigma$ |
| STFT | MLP | 52.57 | 22.58 | 49.12 | 29.83 |
| | CNN | **88.95** | 16.67 | 85.45 | 23.45 |
| | CNN_shallow | 87.90 | 19.71 | **87.71** | 20.06 |
| End-to-End | MLP | 73.14 | 16.07 | 73.36 | 17.52 |
| | CNN | 77.14 | 16.90 | 77.27 | 17.89 |
| | CNN_shallow | **88.00** | 13.81 | **88.01** | 14.02 |
| | CNN_EEGNet | 64.00 | 20.65 | 65.25 | 22.22 |
| Gaze | MLP | 56.71 | 23.34 | 52.70 | 26.48 |
| | CNN | **62.71** | 22.76 | **57.88** | 30.57 |
| | CNN_shallow | 59.57 | 26.50 | 57.40 | 29.85 |
| Hybrid | MLP | 54.86 | 18.81 | 49.04 | 28.14 |
| | CNN | **93.14** | 15.18 | **92.43** | 16.05 |
| | CNN_shallow | 85.71 | 19.17 | 85.81 | 19.07 |

TABLE I
CLASSIFICATION PERFORMANCE FOR MOVE CONDITION

## V. DISCUSSION

### A. Data analysis results considerations

The results of the data analysis process carried out should be taken into consideration when discussing the performance and validity of our action anticipation classification pipeline.

Analysis of subject PSD led to the identification of subject 1 as potentially BCI-illiterate. Under this circumstance, classification results for this subject should be noticeably below those of other subjects. While comparison of subject results seems to confirm our hypothesis, there is a caveat, since this subject appeared to be confused about the experimental protocol, which may suggest they were distracted, potentially resulting in a confounding effect.

During analysis of ERSP phenomena in the $\alpha$, $\beta$ bands during action anticipation and task execution, occurrence of the perturbations described in literature was only clearly

| Approach | Classifier | Accuracy Mean | Accuracy $\sigma$ | F1-Score Mean | F1-Score $\sigma$ |
|---|---|---|---|---|---|
| STFT | MLP | 49.29 | 13.76 | 44.08 | 29.40 |
| | CNN | 91.00 | 13.50 | 89.82 | 16.98 |
| | CNN_shallow | **95.57** | 8.74 | **94.91** | 10.47 |
| End-to-End | MLP | 68.00 | 18.64 | 68.33 | 21.06 |
| | CNN | 76.00 | 16.52 | 74.30 | 19.13 |
| | CNN_shallow | **92.57** | 11.51 | **91.57** | 14.51 |
| | CNN_EEGNet | 56.00 | 20.77 | 56.30 | 22.93 |
| Gaze | MLP | 48.71 | 21.29 | 46.12 | 25.61 |
| | CNN | **53.86** | 22.21 | **50.50** | 25.96 |
| | CNN_shallow | 48.14 | 23.18 | 45.73 | 26.26 |
| Hybrid | MLP | 52.29 | 19.08 | 43.31 | 32.45 |
| | CNN | **94.57** | 10.26 | **93.59** | 12.08 |
| | CNN_shallow | 89.71 | 13.11 | 89.65 | 13.18 |

TABLE II
CLASSIFICATION PERFORMANCE FOR NOMOVE CONDITION

| Approach | Classifier | $DTA_{75}$ (ms) | $DTA_{90}$ (ms) |
|---|---|---|---|
| STFT | CNN | 308.8 | 47.8 |
| | CNN_shallow | **389.0** | **120.0** |
| End-to-End | MLP | 223.2 | 64.2 |
| | CNN | 170.8 | 38.3 |
| | CNN_shallow | 293.6 | 101.4 |
| | CNN_EEGNet | 272.6 | 113.3 |
| Hybrid | CNN | 314.6 | 70.5 |
| | CNN_shallow | 378.9 | 107.6 |

TABLE III
CLASSIFIER DECISION TIME ADVANTAGE

detected during motor action. For the duration of action anticipation epochs, ERSP effects were subtle. Further analysis of individual subject ERSP plots also demonstrated considerable inter-subject variability, and lacked a broad, consistent ERS effect in the $\beta$ band.

These findings suggest that, while ERSP occurs during action anticipation, inter-subject variability, as well as lower amplitude prior to task execution, use EEG signals for action anticipation may require the employment of more sophisticated pre-processing and feature extraction techniques, such as CSP, or the design and training of Machine Learning classifiers with enough complexity to model and extract complex patterns from time-frequency features or raw signal, such as CNNs.

### *B. Comparison of EEG based action anticipation approaches*

The first performance comparison to be made, to answer the thesis research questions, is between the feature extraction approach, using STFT time-frequency decomposition, and the end-to-end approach.

Classification performance for the CNN_shallow classifier is similar using both methodologies; the end-to-end approach yields poorer results when using the deeper CNN classifier, but greatly improves MLP performance, potentially due to the much reduced number of features. Notably, the CNN_EEGNet classifier, which boasts State-of-the-Art performance during MI-BCI trial classification, presents the lowest classification scores. This may be a result of low classifier complexity, as shown by the low number of trainable parameters. EEGNet design by Lawhern et al. [12] was greatly informed by features identified in neurophysiology research to detect the occurrence of Motor Imagery, which may not translate consistently to action anticipation.

Overall, these findings motivate the employment of end-to-end approaches, specifically using shallow CNN networks, for the task of action anticipation, as the lack of a feature extraction step reduces computational burden without a significant impact on performance.

### *C. Gaze-based and Hybrid classification*

To understand the potential benefit of introducing gaze features as a complement to EEG during action anticipation classification, we may compare the results obtained by classifiers using either of these features separately, and in a hybrid approach.

Analysis of gaze-based classification results shows poor performance across all classifiers. As previously described, gaze data was heavily corrupted by noise and artifacts, and as a result, classification is being performed on a feature estimated after heavy signal processing. It is thus excessive to dismiss the potential of gaze features for this task.

Use of these features in complement with EEG resulted in only a small improvement (around 4% increases to accuracy and F1-Score) for the CNN classifier. Shallow CNN classifier performance deteriorated slightly with the introduction of gaze features.

### *D. Action anticipation in Motor Imagery*

To establish whether Motor Imagery may be anticipated in a similar way to action execution, we may compare classification results obtained during the MOVE and NOMOVE conditions. Analysis of results under these two conditions reveals Motor Imagery anticipation is classified more accurately than action anticipation for a majority of classifiers.

Interpretation of this result must be done with care. Intuitively, there would appear to be no reason for Motor Imagery to be easier to anticipate than executed action. There is a chance, under the experimental protocol defined, that subjects may begin their task, be it Motor Imagery or motor action, prior to the correct time (red arrow disappearance). When performing MI, this tendency may be accentuated, which would result in more pronounced ERSP phenomena, leading to (superficially) better classification performance.

Nevertheless, these results demonstrate the possibility of anticipating Motor Imagery task performance using the classifiers developed in this work.

### *E. System Decision Time Advantage*

The Decision Time Advantage was evaluated for all classifiers, with results shown in III, for values $k$ of 75% and 90%. Recall the threshold established in the problem statement, demanding an action anticipation classifier to consistently predict action with at least 100 milliseconds of advantage.

For a $k$ value of 75%, all classifiers shown were capable of anticipating action with an advantage above the threshold level. However, demanding a classifier positive prediction rate of 90%, only the shallow CNN architecture, as well as EEGNet, were capable of reaching the desired level of performance. This value is arbitrary: the adequate value of time advantage must be determined by evaluating how the system will be applied in a physical HRI setting, taking into consideration factors like robot weight, and movement speed.

Evaluation of this novel measure provides HRI system designers with a clearer trade-off between classifier confidence and time advantage provided. The DTA also complements traditional performance metrics, such as accuracy and F1-score, by establishing a measure of classifier performance in a real-time setting for robotic control and decision systems.

## VI. Conclusion

The work conducted for this thesis aimed to shed light on the viability of using an EEG-based action anticipation, particularly in the context of physical HRI.

With this aim, an experiment was conducted during which participants interacted with a robotic arm in a pick-and-place task. EEG signals and gaze information were recorded and analysed to evaluate the potential of several CNN classifiers for action anticipation. Two processing pipelines were evaluated: one performing feature extraction, by computing the STFT time-frequency decomposition of action anticipation epochs, and one classifying only the noise-filtered EEG signal. These classifiers were also adapted to gaze-based classification, and extended to explore the potential of a hybrid EEG and gaze classification. Additionally, a novel metric was proposed and evaluated for each classifier, combining time advantage provided relative to movement onset with classifier confidence.

Through a comparison of classifiers making use of feature extraction and an end-to-end approach, this work demonstrates CNN classifiers are capable of correctly classifying action anticipation epochs with similar performance in both circumstances, suggesting these classifiers may dispense with the need for explicit feature extraction, reducing computational burden. The end-to-end, shallow CNN classifier was able to reach an accuracy of 88.00% ($\sigma$ = 13.81%), and an F1-score of 88.01% ($\sigma$ = 14.02%) when classifying action anticipation epochs.

Employment of hybrid classifiers, combining EEG and gaze information, yielded only slight performance improvements when leveraging context independent gaze features. Using a hybrid classification approach, a deep CNN classifier was able to reach an accuracy of 93.41% ($\sigma$ = 15.18%) and F1-score of 92.43% ($\sigma$ = 16.05%) when classifying Motor Imagery-preceding epochs.

A comparison of classification performance during Motor Imagery and motor execution revealed similar performance under both conditions, attesting to the potential of this methodology for task anticipation in rehabilitation contexts.

Finally, evaluation of a novel time advantage metric revealed shallow, fully-convolutional CNN were capable of consistently anticipating action sooner than other architectures tested, classifying 90% of epochs as anticipating motor action with a time advantage of 120 milliseconds.

This research demonstrates the viability of systems utilizing CNN, particularly shallow, fully-convolutional configurations, for the task of human action anticipation using EEG signals. This work provides a foundation for the use of such an approach in the context of physical HRI, and establishes a new way of measuring action anticipation performance which makes confidence/time advantage trade-offs clearer for HRI system designers.

### A. System limitations

The main limitations of this research relate to the experimental protocol. Firstly, this protocol is cue-based, rather than self-paced: while literature suggests slow-cortical potentials, such as the BP, and oscillatory phenomena, such as ERSP, occur in both circumstances during action anticipation [29], research under a more realistic scenario would be needed to confirm this. Secondly, it is impossible to guarantee subjects won't begin task execution while the anticipation cue remains on screen, despite being instructed otherwise, muddying the distinction between these two states, and leading to optimistic performance estimates (assuming task execution will be easier to detect than anticipation).

Additional limitations include model parameter selection, which could be performed using a grid search process; the limited number of different classification models; exploration of STFT features as the only alternative to end-to-end approaches; and the low number of experiment participants, without the inclusion of any subjects with know neurological conditions, such as stroke victims.

### B. Future work

Further research is needed to shine more light on the topic of action anticipation from EEG signals. Future studies should seek to address the limitations and shortcomings of this work, as well as explore other avenues of research, namely by increasing the number of participants; including subjects suffering from neurological conditions, such as the sequelae of stroke; conducting an experiment using a self-paced protocol, leveraging motion detection sensors, such as accelerometers and cameras, to more clearly define action onset; decoding motor intention, achieving limb-level resolution, rather than a binary decision on action anticipation; exploring different classification models and configurations, potentially testing architectures with attention mechanisms; and investigating other pre-processing and feature extraction methodologies, such as by using Wavelet decomposition.

Studies on this topic may also seek to evaluate the benefits of action anticipation on the safety, reliability, and efficacy of HRI, as well as the way human counterparts perceive and interact with robotic colleagues capable of anticipating their movements.

## ACKNOWLEDGMENT

First and foremost, I want to thank my parents for their unconditional love and support throughout my entire life. I would also like to share my gratitude to my friends who have been with me through the course of my academic journey, and without whom this would not be possible.

I also want to thank my supervisors, Professors Plinio and Athanasios, for their help and guidance in writing this thesis.

## REFERENCES

[1] C. H. Brunia, G. J. van Boxtel, and K. B. Böcker, *The Oxford Handbook of Event-Related Potential Components*. Oxford University Press, 2011, ch. Negative slow waves as indices of anticipation: the Bereitschaftspotential, the contingent negative variation, and the stimulus-preceding negativity.

[2] E. Lew, R. Chavarriaga, S. Silvoni, and J. d. R. Millán, "Detection of self-paced reaching movement intention from eeg signals," *Frontiers in neuroengineering*, vol. 5, p. 13, 2012.

[3] D. Planelles, E. Hortal, Á. Costa, A. Úbeda, E. Iáñez, and J. M. Azorín, "Evaluating classifiers to detect arm movement intention from eeg signals," *Sensors*, vol. 14, no. 10, pp. 18 172–18 186, 2014.

[4] A. Buerkle, W. Eaton, N. Lohse, T. Bamber, and P. Ferreira, "Eeg based arm movement intention recognition towards enhanced safety in symbiotic human-robot collaboration," *Robotics and Computer-Integrated Manufacturing*, vol. 70, p. 102137, 2021.

[5] D. Rozado, A. Duenser, and B. Howell, "Improving the performance of an eeg-based motor imagery brain computer interface using task evoked changes in pupil diameter," *PloS one*, vol. 10, no. 3, p. e0121262, 2015.

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

[7] L. Defebvre, P. Derambure, J. Bourriez, A. Destee, and J. Guieu, "Event-related desynchronization and parkinson disease. importance in the analysis of the phase of preparation for movement," *Neurophysiologie Clinique= Clinical Neurophysiology*, vol. 29, no. 1, pp. 71–89, 1999.

[8] K. A. Ludwig, R. M. Miriani, N. B. Langhals, M. D. Joseph, D. J. Anderson, and D. R. Kipke, "Using a common average reference to improve cortical neuron recordings from microelectrode arrays," *Journal of neurophysiology*, vol. 101, no. 3, pp. 1679–1689, 2009.

[9] J. Onton, M. Westerfield, J. Townsend, and S. Makeig, "Imaging human eeg dynamics using independent component analysis," *Neuroscience & biobehavioral reviews*, vol. 30, no. 6, pp. 808–822, 2006.

[10] I. J. Rampil, "A primer for eeg signal processing in anesthesia," *Anesthesiology*, vol. 89, no. 4, pp. 980–1002, 10 1998. [Online]. Available: https://doi.org/10.1097/00000542-199810000-00023

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[12] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[13] E. N. Cannon and A. L. Woodward, "Action anticipation and interference: A test of prospective gaze," in *CogSci... Annual Conference of the Cognitive Science Society. Cognitive Science Society (US). Conference*, vol. 2008. NIH Public Access, 2008, p. 981.

[14] D. Kahneman, *Attention and effort*. Citeseer, 1973, vol. 1063.

[15] S. Moresi, J. J. Adam, J. Rijcken, P. W. Van Gerven, H. Kuipers, and J. Jolles, "Pupil dilation in response preparation," *International Journal of Psychophysiology*, vol. 67, no. 2, pp. 124–130, 2008.

[16] M. Naber, G. A. Alvarez, and K. Nakayama, "Tracking the allocation of attention using human pupillary oscillations," *Frontiers in psychology*, vol. 4, p. 919, 2013.

[17] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.

[18] M. Ahn, H. Cho, S. Ahn, and S. C. Jun, "High theta and low alpha powers may be indicative of bci-illiteracy in motor imagery," *PloS one*, vol. 8, no. 11, p. e80886, 2013.

[19] P. Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on audio and electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[20] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "Bci competition 2008–graz data set a," *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, vol. 16, pp. 1–6, 2008.

[21] May 2023. [Online]. Available: https://www.brainproducts.com/solutions/acticap/

[22] May 2023. [Online]. Available: https://www.brainproducts.com/solutions/liveamp/

[23] M. Kassner, W. Patera, and A. Bulling, "Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction," in *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication*, 2014, pp. 1151–1160.

[24] A. Vourvopoulos, S. Legeay, and P. Figueiredo, "Neuxus: A biosignal processing and classification pipeline for real-time brain-computer interaction," 2020.

[25] C. Kothe *et al.*, "Lab streaming layer (lsl)," 2014.

[26] R. C. Oldfield, "The assessment and analysis of handedness: the edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, 1971.

[27] A. Farabbi, F. Ghiringhelli, L. Mainardi, J. M. Sanches, P. Moreno, J. Santos-Victor, P. Figueiredo, and A. Vourvopoulos, "Motor-imagery eeg dataset during robot-arm control," Apr. 2020, Approved by the Ethics Committee of CHULN and CAML (Faculty of Medicine, University of Lisbon) with reference number: 245/19. [Online]. Available: https://doi.org/10.5281/zenodo.5882500

[28] A. Farabbi, F. Ghiringelli, L. Mainardi, P. Figueiredo, J. M. Sanches, and A. Vourvopoulos, "Robot-control through a motor imagery brain-computer interface : evaluation of learnability and feedback perspective," Master's thesis, Politecnico di Milano, 2020. [Online]. Available: http://hdl.handle.net/10589/164451

[29] L. A. Wheaton, H. Shibasaki, and M. Hallett, "Temporal activation pattern of parietal and premotor areas related to praxis movements," *Clinical neurophysiology*, vol. 116, no. 5, pp. 1201–1212, 2005.