

Decoding the Brain for Target Detection in Aerial Imagery: A High-Throughput Brain-Computer Interface Approach

Miguel Rocha e Costa



FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Master's degree in Biomedical Engineering

Supervisor: João Paulo Cunha, PhD

July 24, 2024

Decoding the Brain for Target Detection in Aerial Imagery: A High-Throughput Brain-Computer Interface Approach

Miguel Rocha e Costa

Master's degree in Biomedical Engineering

Resumo

Interfaces cérebro-computador (BCI) são um tipo de tecnologias emergentes focadas na interação direta de cérebros humanos e computadores para criar sistemas que substituem e/ou aumentam as capacidades individuais. As BCIs são frequentemente usadas para auxiliar indivíduos que sofrem de deficiências motoras, mas a sua aplicabilidade estende-se para além de áreas médicas. O fator chave é que o mesmo paradigma de evocação cerebral pode ser facilmente extrapolado de finalidades médicas para não-médicas, se uma boa generalização for alcançada através do processamento de dados.

Nesta tese foi desenvolvido um sistema BCI no contexto de busca visual aérea, acelerando a deteção de alvos (neste caso identificação de humanos) num vasto ambiente de campo aberto coberto por imagens de drones. Esta deteção é efetuada a uma grande cadência (5 imagens por segundo) pela análise no domínio do Eletroencefalograma de observadores humanos. Para isso, foram usados Potenciais Relacionados a Eventos (ERPs) associados a mecanismos neuronais de atenção visual, nomeadamente o N2pc, N2pcb e P3, elicitados por um objeto de interesse dentro das imagens. Adicionalmente, foi testado o uso do N2pc e o N2pcb para estimar a localização do alvo humano dentro da imagem nas dimensões horizontal (esquerda vs. direita) e vertical (superior vs. inferior), respetivamente usando a mesma taxa de apresentação de imagens de 5 Hz.

Para testar esta abordagem, dois paradigmas independentes de Apresentação Visual Rápida e em Série (RSVP) foram desenhados e apresentados separadamente a 6 sujeitos depois de obtida uma autorização da comissão de ética institucional para esta experimentação humana. Estes dois paradigmas foram desenhados com características muito diferentes, nomeadamente um conjunto diferente de ambientes/terreno e características de alvos humanos, incluindo a distribuição dos seus centroides. O primeiro (240 imagens com 24 targets) foi usado para treinar/validar os modelos de aprendizagem computacional (ML) e o segundo (160 imagens com 16 targets) usado exclusivamente para testá-los. Além disso, o segundo paradigma foi visto pela primeira vez pelos 6 sujeitos durante o procedimento, permitindo a avaliação da capacidade do cérebro de discriminar alvos num ambiente e alvo humano desconhecidos. Foram testados modelos ML personalizados (específicos para cada sujeito) e modelos generalizados (treinados com os dados de todos os sujeitos para o primeiro paradigma e testados com o segundo).

Usando apenas o N2pc com um modelo LDA personalizado para cada sujeito para discriminar entre imagens alvo e não-alvo, obteve-se uma AUC média de 0.77 no conjunto de imagens de treino e 0.75 de AUC no conjunto de dados de teste. Em comparação, foram obtidos 0.77 e 0.71 de AUC usando o componente P3 para as imagens de treino e teste, respetivamente, com uma arquitetura SVM personalizada para cada sujeito. Ao combinar os dois eventos com uma SVM também personalizada para cada sujeito, uma AUC média de 0.84 foi conseguida nas imagens de treino enquanto uma AUC de média de 0.74 foi obtida nas imagens de teste. Já os modelos generalizados que usaram o N2pc e o P3, tanto individualmente como em conjunto, apresentaram resultados ligeiramente inferiores aos obtidos para os modelos personalizados para cada sujeito. No entanto, combinando as diferentes variações de modelos aqui averiguadas para preencher as

lacunas de desempenho em cada sujeito, uma especificidade e sensibilidade média de 0.76 e 0.79 foi alcançada ao tentar discriminar estímulos alvo e não-alvo o que demonstrou uma boa performance, um pouco acima da literatura atual. Como curiosidade, o BCI proposto obteve resultados melhores do que um algoritmo de visão por computador clássico para este tipo de tarefas testado nas mesmas imagens.

Além disso, foi tentada uma distinção específica para cada sujeito entre o Campo Visual Direito e o Campo Visual Esquerdo usando N2pc, atingindo uma média de AUC de 0.75 em ambos os paradigmas com um modelo SVM. Ambas as tarefas de discriminação atingiram resultados em par com a literatura. Explorando ainda mais a ideia de discriminação do campo visual, uma Rede Neural artificial foi treinada para estimar a posição do centroide do alvo ao longo da dimensão horizontal da imagem, atingindo uma correlação média de 0.52 entre o ponto previsto pela rede e o valor real nas imagens apresentadas, um valor superior aos observados na literatura.

Estes resultados apontam para que este tipo de sistemas BCI de alta cadência (híbridos, recorrendo a observadores humanos e técnicas de ML sobre os potenciais cerebrais medidos) generalizam bem quando usamos os ERPs N2pc e P3. Assim, pensamos ter efetuado uma prova de conceito de que este tipo de sistemas poderão ser usados em cenários de busca e salvamento reais, adicionando a deteção do alvo à sua localização na imagem e com o potencial de ter melhor performances e capacidades de generalização que outras abordagens computacionais e recorrendo a uma menor necessidade de recursos.

Palavras-chave:

Interfaces cérebro-computador, Potenciais Relacionados a Eventos, N2pc, N2pcb, P3, Apresentação Visual Rápida e Serial.

Abstract

Brain-computer interfaces (BCIs) are a type of emerging technology focused on the direct interaction of human brains and computers to create systems that replace and/or augment individual capabilities. BCIs are often used to assist individuals who suffer from motor impairments, but their applicability extends beyond medical areas. The key factor is that the same brain evocation paradigm can be easily extrapolated from medical to non-medical purposes, if good generalization is achieved through data processing.

In this thesis, a BCI system was developed in the context of aerial visual search, accelerating target detection (in this case identification of humans) in a vast open field environment covered by drone images. This detection is carried out at a high frame rate (5 images per second) by analyzing the time domain of the Electroencephalogram of human observers. For this purpose, Event-Related Potentials (ERPs) associated with visual attention neural mechanisms, namely N2pc, N2pcb and P3, elicited by an object of interest within the images were used. Additionally, the use of N2pc and N2pcb was tested to estimate the location of the human target within the image in the horizontal (left vs. right) and vertical (top vs. bottom) dimensions, respectively using the same 5 Hz image presentation rate.

To test this approach, two independent Rapid Serial Visual Presentation (RSVP) paradigms were designed and presented separately to 6 subjects after obtaining authorization from the institutional ethics committee for this human experimentation. These two paradigms were designed with very different characteristics, namely a different set of environments/terrain and human target characteristics, including its clothes and the distribution of their centroids within the images. The first paradigm (240 images with 24 targets) was used to train/validate the machine learning (ML) models and the second (160 images with 16 targets) was used exclusively to test them. In addition, the second paradigm was seen for the first time by the 6 subjects during the procedure, allowing the evaluation of the brain's ability to discriminate targets in an unknown and unseen environment and human target characteristics. Customized ML models (specific to each subject) and generalized models (trained with the data of all subjects for the first paradigm and tested with the second) were tested.

Using only N2pc with a personalized LDA model for each subject to discriminate between target and non-target images, an average AUC of 0.77 was obtained on the training image set and 0.75 AUC on the test dataset. In comparison, 0.77 and 0.71 average AUC were obtained using the P3 component for the training and test images, respectively, with a personalized SVM architecture for each subject. By combining both with a subject-specific SVM, an average AUC of 0.84 was obtained for the training image set while an average AUC of 0.75 was achieved for the test dataset. The generalized models that used N2pc and P3 both individually or combined presented slightly lower results than those obtained for the personalized models for each subject. However, by combining the different variations of models investigated here to fill the performance gaps in each subject, an average specificity and sensitivity of 0.76 and 0.79 was achieved when trying to discriminate target and non-target stimuli, which demonstrated good performance, slightly above the

current literature. As a curiosity, the proposed BCI obtained better results than a classic computer vision algorithm for this type of task tested on the same images.

In addition, a subject-specific distinction was attempted using N2pc between the Right Visual Field and the Left Visual Field, achieving an average AUC of 0.75 in both paradigms with an SVM model. Both discrimination tasks achieved results in line with the literature. Further exploring the idea of all-visual field discrimination, an artificial Neural Network (ANN) was trained to estimate the position of the target centroid along the horizontal dimension of the image (not only right vs. left but any position), achieving an average correlation of 0.52 between the point predicted by the ANN and the real value in the presented images, a performance higher than those observed in the literature.

These results indicate that this type of high-frame rate BCI system (hybrid, using human observers and ML techniques on the measured brain potentials) seems to generalize well when using N2pc and P3 ERPs. Thus, we believe we have carried out a proof of concept that this type of system can be used in real search and rescue scenarios, adding fast target detection to its location in the image and with the potential to have better performance and generalization capabilities than other computational approaches, demanding much less resources.

Keywords:

Brain-Computer Interface, Event-Related Potentials, N2pc, N2pcb, P3, Rapid-Serial Visual Presentation.

Agradecimentos

Depois de 3 anos de trabalho árduo, dedicação e muita aprendizagem, está finalmente concluído. Não foi fácil e não foi conseguido sozinho, mas orgulho-me de todo o trabalho árduo, conhecimento adquirido e amigos que fiz ao longo destes anos. Por isso, agradeço a todos aqueles que me acompanharam nesta jornada.

Em primeiro lugar, gostaria de agradecer aos meus colegas de casa no Porto por todo o apoio que me deram durante estes 3 anos, inclusive alguns dos quais me ajudaram diretamente neste projeto durante o período de aquisições. Por isto, deixo um agradecimento ao Vasco, Margarida, Francisco, Ricardo e Bia.

Um obrigado também aos meus colegas do MEB e MBio que me acompanharam diretamente em todo o percurso percorrido na FEUP, com quem partilhei noites sem sono e horas de trabalho.

Não podia deixar de agradecer também a todos os membros do grupo do BRAIN por me terem acolhido e com quem pude partilhar grandes momentos, não só nos laboratórios mas também em contextos descontraídos. Um obrigado especial ao Vítor e ao Francisco, que me apoiaram mais diretamente na aprendizagem necessária para concluir este trabalho. Um obrigado também ao Duarte por toda a disponibilidade e ajuda sempre que foi preciso. Não podia também deixar de agradecer imenso ao Professor João Paulo Cunha por todas as oportunidades que me deu ao longo destes três anos, por motivar sempre a discussão e com quem aprendi imenso, dentro e fora da engenharia.

Por fim, um especial obrigado a toda a minha família por todo o apoio incondicional que recebi e por acreditarem sempre em mim quando era mais difícil.

Miguel Rocha e Costa

*“If the human brain were so simple that we could understand it,
we would be so simple that we couldn’t ”*

Emerson Pugh

Contents

1	Introduction	1
1.1	Context	1
1.2	Motivation	1
1.3	Objectives	3
1.4	Achievements	4
1.5	Thesis Structure	4
2	Brain-Computer Interfaces (BCI): A Theoretical background	6
2.1	The BCI evolution: A brief historic perspective	6
2.2	Categorization of BCI systems	8
2.3	The Electroencephalogram	10
2.3.1	A measure of brain activity	11
2.3.2	System of electrode placement	11
2.3.3	The rhythm of the brain	12
2.4	The geography of thought	14
2.4.1	The human cortex	15
2.4.2	The visual cortex	16
2.4.3	The motor and sensorimotor cortex	19
2.5	From the brain to a computer: An overview of the BCI Pipeline	20
2.6	The paradigms of BCI technology	22
2.6.1	Event-Related Potentials	22
2.6.2	Steady State Evoked Potentials	27
2.6.3	Motor Imagery	28
2.7	BCI performance metric	30
2.8	Chapter 2 summary	30
3	EEG-based Rapid-Serial Visual Presentation (RSVP) and target localization pipelines: A literature review	32
3.1	RSVP-based BCIs	32
3.2	Non-RSVP spatial localization-based algorithms	35
3.3	EEG classification algorithms and optimal feature selection	37
3.3.1	Mutual-Based Information (MBI)	38
3.3.2	Support Vector Machines (SVM)	38
3.3.3	Linear Discriminant Analysis (LDA)	38
3.3.4	Random Forest	39
3.3.5	XGBoost	39
3.3.6	AdaBoost	40
3.3.7	Artificial Neural Networks	40

3.4	Model evaluation	42
3.4.1	Cross-Validation (CV)	42
3.4.2	Classifiers Grid Search	43
3.4.3	Model evaluation metrics	44
3.5	Chapter 3 summary	45
4	Materials and Methods	47
4.1	Signal acquisition	47
4.1.1	Designing the RSVP <i>Oddball paradigm</i>	47
4.1.2	EEG Instrumentation	57
4.1.3	EEG set-up and acquisition parameters	64
4.1.4	Participants	66
4.2	EEG Processing pipeline	67
4.2.1	Estimating signal quality	69
4.2.2	Signal Pre-processing	72
4.2.3	Event-Related Potential analysis	81
4.2.4	BCI actuation structure: Feature vector composition and class discrimination	87
4.2.5	Horizontal centroid estimation: Using Artificial Neural Networks to describe Biological Neural Network activity	91
4.3	Establishing a target identification performance ground truth with a pre-trained Deep Learning model	94
4.4	Chapter 4 summary	94
5	Results and Discussion	97
5.1	The LDR PCB	97
5.1.1	The hardware design	97
5.1.2	Segmentation of the LDR channel and stimuli labeling	105
5.2	EEG signal analysis	108
5.2.1	EEG quality estimation underlying the pre-processing choices	108
5.2.2	ERP Grand Averages	120
5.3	BCI performance	135
5.3.1	Target vs Non-Target classification	135
5.3.2	Sensitivity and Specific: Optimal model combination for target and non-target discrimination	154
5.3.3	Right Visual Field vs Left Visual Field classification (N2pc)	155
5.3.4	Top Visual Field vs Bottom Visual Field classification (N2pcb)	162
5.3.5	Four quadrant classification (N2pc)	163
5.3.6	Horizontal centroid estimation (N2pc)	164
5.4	Chapter 5 summary	167
6	Conclusion	169
7	Future work	171
References		174
Appendix		186

A	186	
A.1	Written consent	187
A.2	Signal quality estimation analysis	190
A.2.1	Frequency spectra	190
A.2.2	Electrode bridging	191
A.2.3	SSVEP SNR estimation	192
A.3	ERP Grand-Averages (Testing paradigm)	193
A.3.1	P3	193
A.3.2	N2pc	195
A.3.3	N2pcb	199
A.3.4	Four quadrant N2pc	202
A.4	BCI performance	203
A.4.1	Horizontal centroid estimation (N2pc) - Train/Validation set	203
A.4.2	Horizontal centroid estimation (N2pc) - Testing set	203

List of Figures

2.1	The number of papers published over the years that were identified from the <i>PubMed</i> database using the keyword "Brain-Computer Interface" [28].	8
2.2	Categorization of the main types of BCI systems by dependability, recording method, and mode of operation. The acronym ECoG refers to Electrocorticography, fMRI to functional Magnetic Resonance Imaging, EEG to Electroencephalography, MEG to Magnetoencephalography, PET to Positron Emission Tomography, and fNIRS to Functional Near-infrared Spectroscopy. Adapted from [30]	9
2.3	Comparison in the temporal resolution, spatial resolution, and complexity planes of all invasive (rectangular) and non-invasive (oval) brain activity recording technologies used in BCI systems [34]. The acronym ECoG refers to Electrocorticography, fMRI to functional Magnetic Resonance Imaging, EEG to Electroencephalography, MEG to Magnetoencephalography, PET to Positron Emission Tomography, and fNIRS to Functional Near-infrared Spectroscopy.	10
2.4	Illustration of the 10-20 EEG system derived from the circumferential measurements of the skull [41].	12
2.5	Illustration of the 10-10 EEG system. Adapted from [41].	13
2.6	Illustration of the different frequencies transmitting information in the brain [24]. It is important to note that the frequency ranges defined here may vary slightly depending on the source, which can generate some confusion when addressing them [44].	14
2.7	The $\frac{1}{f}$ shape characteristic of the EEG recorded on my scalp. Notice the presence of a harmonic oscillation related to line noise at 25 Hz, a common occurrence when capturing the EEG at the BRAIN laboratory at INESC TEC due to the 50 Hz of the power line.	15
2.8	The Lateral (A) and Midsagittal (B) views of the human brain with the four main lobes highlighted in different colors and the cerebellum illustrated in gray below the occipital and temporal lobes. The midsagittal view must be interpreted as the brain cut in half dividing the left and right hemispheres, revealing the structure that connects them both, the <i>corpus callosum</i> . Adapted from [47].	16
2.9	Illustration of the <i>retino-geniculo-cortical</i> pathways [49]. Image (a) encompasses the <i>magnocellular</i> (red) and <i>parvocellular</i> (green) pathways and their contribution to the dorsal and ventral streams of visual information. It also shows the subcortical communications between LGN, V5 and V1 areas. Adapted from [48]. Image (b) does not make the distinction between both cell pathway types to the ventral and dorsal streams but includes their respective summarized function. Adapted from [49]	17

- 2.10 Illustration of the main areas of the visual cortex, namely areas V1, V2, V3, V3A, V4, and V5, including subcortical pathways and structures (LNG). Adapted from [49]. 18
- 2.11 Illustration of the somatosensory homunculus in the *postcentral gyrus* (blue) of the left hemisphere and motor homunculus in the *precentral gyrus* (red) of the right hemisphere [47]. 19
- 2.12 Generalized pipeline for decoding EEG information into executable commands through a BCI system [30]. 21
- 2.13 The target (red) and non-target (blue) visual ERP response during an *oddball face recognition paradigm* recorded with 16 electrodes on my scalp. 100 target responses were averaged to obtain the target-related ERP while 1000 were used for the non-target response. The recorded EEG signals were referenced to both earlobes and posteriorly band-pass filtered between 0.1 and 30Hz. The visual stimuli were composed of several faces of recognizable celebrities with the face of Daniel Radcliffe being attended to, representing the target response occurring with 10% of probability. The image depicts the ERPs mentioned at channels Cz (above) and PO8 (below) revealing the opposing components associated with human face recognition VPP and N170, respectively, around the same period (green patch). Furthermore, the components N2 (blue patch) and P3 (light orange patch) can be seen around 200 ms and 300 ms respectively, with a clear difference in amplitudes between the target and non-target responses. Although not being highlighted, the accentuated negativity following the P300 around 450 ms represents the N400, also showing significant differences in amplitude between both responses. An important consideration to have when visualizing ERPs and comparing them to the literature is that researchers sometimes invert the y-axis making the negative peaks appear pointed upwards and the positive peaks pointed downwards. Despite being relatively common, it is not a standard and as a result, will not be done throughout the rest of the document. 25
- 2.14 EEG signal (Oz-Cz) obtained during a visual light stimulation at 15 Hz. The left image exhibits the time domain SSVEP resulting from the average of the time-locked signals over 10 trials, showcasing a clear transient VEP at the start of the stimulation ($t = 0.0\text{s}$) with the subsequent oscillations related to the SSVEP (starting around 0.3s). The right images constitute the frequency spectrum of the signal, demonstrating the pronounced power of the stimulation frequency (15 Hz) and its higher harmonics [73]. 28

2.15 Illustration of the ERD and ERS that accompanies left wrist movement generation on my motor cortex captured during a motor imagery paradigm. The instant $t = 0\text{s}$ depicts the instruction to imagine the movement. The first half of the image depicts the ERD and subsequent ERS of the <i>mu</i> rhythm in the time domain at channel C4, contralateral to the moved limb, and the changes in Power Spectral Density (PSD) across the 16 electrodes recorded in a topographic map of the scalp (1200 ms - 2400 ms). The second half of the image depicts the same trial recorded but with information related to the upper <i>Beta</i> frequency. Electrode C4 is highlighted with a white dot in both topographical maps. Note that in both maps, an increase in frequency power (ERS) focused on electrode C4 can be seen around the time the movement is executed and a depression (ERD) around the same area that precedes the execution for both frequencies. There is also an apparent delay between the ERD and ERS of the <i>Beta</i> rhythm in relation to the <i>mu</i> wave, corroborating with the literature. Before plotting, the signals, in addition to being filtered in the corresponding frequency ranges, were processed using the Common Spatial Pattern (CSP) algorithm to improve the source detection of activity in each electrode, thus revealing the activity of the left wrist generation over the right motor cortex in channel C4. Furthermore, each instant depicted in the topographic map corresponds to the average PSD activity 50 ms before and after that instant.	29
3.1 Diagram representing an ANN architecture [87].	42
3.2 An example of the inner workings of Cross-Validation with 5-folds [88].	43
4.1 The used DJI Mini 4 Pro drone and its main specifications.	48
4.2 Depiction of the 6 types of environments described that are contained in the Training image dataset. In each of these, a human target can be seen wearing a yellow t-shirt.	50
4.3 Depiction of the type of images contained in the Testing image dataset. On the left, an image containing a human target wearing a red sweater, and on the right, an image without a human target.	50
4.4 Dimensions of the aerial images taken for the Testing dataset by a drone at FEUP and the subsequent cropping (3 x 2) into smaller images, indicated by the blue dashed lines. The target stimuli are identified by the presence of a human lying down contained in the sub-image, as pinpointed by the red circle, whereas in the non-target class, by its absence.	52
4.5 An example of the target (left) and non-target (right) stimuli displayed in the designed <i>oddball</i> experiment. Note that the red circle surrounding the object of interest in the target image is only used to highlight its location and is not included during the stimuli presentation.	53

4.6 Flowchart of the algorithm applied to extract the centroid of my body in the Testing target images along the horizontal and vertical axis. To achieve this, an HSV color mask was constructed based on the average and standard deviation values of my red sweater's color, segmented manually from a random target image (A). The mask was then applied to every image iteratively after reducing its size to half and converting it to the HSV color format. The 2D geometric moments were then calculated from the foreground pixel values that form the isolated target (me) and its centroid (\bar{x}, \bar{y}) was extracted (B). The computed centroid can be seen correctly placed through this methodology by a blue dot over my body. Note that the HSV color range does not have a standard format. <i>OpenCV</i> displays the Hue values between 0 and 179 while online color pickers typically use the 0-360 range. Similarly, the Saturation and Value channels in <i>OpenCV</i> (0-255) also differ depending on the online sources that present these values from 0 to 100%. This is the reason why the HEX color codes are displayed here next to the HSV values (A).	54
4.7 Diagram containing the routines that form the RSVP <i>oddball</i> paradigm, including both the Training and Testing portions that are done concurrently. The routines that have a "SPACEBAR" key placed next to them indicate that the participant has to press it to continue the session, having total power on the resting period's duration.	58
4.8 Representation of the Micromed SD LTM PLUS 64 amplifier and the <i>BRAIN QUICK® Software</i> [98]. Note that this figure has two of these amplifiers in parallel amounting to a total of 128 electrodes recorded, showcasing their system's workaround for higher-density acquisitions. Yet, only one unit of these was used for this project.	60
4.9 Schematic of a voltage divider circuit comprising an LDR, drawn in the <i>EasyEDA</i> online tool for schematic drawing and circuit simulation.	61
4.10 Depiction of the EEG synchronization concept through an LDR implemented in a voltage divider circuit. During the target class presentation, a white square is placed on the top left corner of the monitor, inducing the highest voltage output on the circuit (orange line). Similarly, darker greyscale squares are used for the non-target class, generating lower voltages (green). The use of two intermittent greyscale values in the non-target class is due to images being presented with no intervals, thus enabling separation between two consecutive non-targets. The EEG acquisition and voltage values read by the circuit are combined in the amplifier, allowing to track the instant each stimulus appears and its duration.	62
4.11 Connection of the DC dongle to the EEG amplifier.	62
4.12 Representation of the <i>Voltera V-one</i> PCB printing machine [101].	64
4.13 The Besdata 64 channel cap (a) and the EEG montage (b) depicting the scalp distribution of electrodes.	65
4.14 The signal acquisition setup using two computers, an LDR circuit, a monitor, and the <i>Micromed</i> EEG amplifier.	66

4.15 Overview of the EEG data pipeline. Signal quality estimation groups are displayed in orange color and pre-processing techniques are in light blue. A dashed arrow connecting the frequency tagging and ICA blocks indicates that data did not flow directly between these two processes but the information extracted in the frequency tagging analysis guided the ICA supervised decisions in component removal, hence the "compare" keyword. The ERP analysis is illustrated in purple, the feature vector composition with subsequent model training in red, and the best model's testing in green. Each one of these groups of data conditioning is detailed in the following sessions. Note that on the classification block, the Neural Networks were only trained and tested with all channel features, whereas the classification grid was approached with both an optimal channel selection and all channel information.	68
4.16 The cluster of occipital channels (circled in orange) used to estimate the SNR of the SSVEP caused by the constant image presentation rate (5Hz).	72
4.17 The frequency (blue) and phase (green) response of the low-pass filter applied at 28Hz on the EEG signals.	73
4.18 Comparison between the different filtering approaches to the LDR channel. . . .	77
4.19 The defined thresholds for the target (orange), non-target (green), and SSVEP time windows, intersecting the filtered LDR channel. Notice the equidistant characteristic of each threshold relative to Target and Non-Target stimuli instances. . . .	78
4.20 Representation of the unmixing ability of ICA to decompose signals from different sources linearly summed and captured by two microphones. The linear weights for each microphone (a_1, b_1 and a_2, b_2), derived solely by the sensor data, reflect the proximity of each speaker to the respective microphones. By unmixing the signals captured with these weights, the original sources can be obtained [112]. .	79
4.21 Minimization of the Gaussianity in two mixed signals through ICA to recover their statistically independent sources [112]. Image (A) depicts the data distribution of both signals projected on each axe before ICA and image (B) represents the transformed data with ICA.	80
4.22 The effect of baseline normalization/correction (blue) on the same non-normalized single trial at channel CPz (red). In the baseline period (-200 to 0 ms), the blue signal has a mean of 0, indicative of this normalization.	82
4.23 Illustration of ERP extraction method. Individual trials (left) are averaged together for each electrode individually, resulting in a waveform comprising only the target stimuli processing while eliminating all spontaneous and variable neural activity associated with individual responses (right).	83
4.24 The regions of interest targeted for the N2pc/N2pcb (green) and P3 (red) ERP components. The color scheme defined here will be maintained for all figures depicting these ERPs. In other words, N2pc and P3 plots will always be represented by the green and red color schemes throughout the document, respectively. As for the N2pcb, it will always be plotted referencing the orange color.	84
4.25 Summary of the conceptual computations required to obtain the N2pc/N2pcb ERP group through electrode pair subtractions. The contralateral and ipsilateral components of the N2pc change depending on RVF or LVF appearance, as indicated by the color code of each electrode hemispheric region. Left hemisphere channels are depicted in light blue and right hemisphere ones in light green. On the N2pcb, bilateral activity refers to the sum of electrode hemisphere pairs. Both channel and temporal parameters for all ERP analyzed are displayed.	85

4.26	The visual angle conceptualization [114].	86
4.27	Distribution of target centroids within the images presented during the Training (A) and Testing (B) sessions. The lateral (green) and vertical (red) 1.2° visual angle (VA) thresholds are also exhibited. Notice that the vertical centroid position starts (0 px) from the top as was registered during the centroid tracking algorithm used when preparing stimuli for the <i>Psychopy</i> sessions.	87
4.28	The Grid Search CV that encapsulates all hyper-parameter and architecture model combinations.	90
4.29	Flowchart of the feature vector composition and subsequent classification and centroid estimation. Once again the green color depicts N2pc, orange the N2pcb, and red the P3-based data flow. The purple color represents the data flow common to every model analysis, excluding the channel data pairs at the start of the diagram that are only common to the N2pc/N2pcb group.	92
5.1	The <i>PGM1201-MP</i> LDR component selected for the circuit developed [117].	98
5.2	Noise measurements on the breadboard circuit prototype compared between using an isolated battery (A) and a USB (B) connected to the computer for an upper Voltage range (≈ 2.4 V) in the first image row and a lower Voltage range in the images below (≈ 500 mV). The images were generated directly on the oscilloscope and saved on a pen drive through a dedicated USB port.	99
5.3	The LDR circuit diagram comprising all components. Extracted from <i>Altium Designer</i> . The J1 component represents the USB Type C component, VR1 the Voltage Regulator, PhR1 the LDR, and PR1 the $20\text{ k}\Omega$ potentiometer.	100
5.4	The PCB 2D layout design of the LDR circuit and its dimensions on the left. On the right, is an analogous 3D version. Extracted from <i>Altium Designer</i>	100
5.5	The board resulted from printing on the <i>Voltera v-One</i> machine (A) and an analogous version ordered from <i>Eurocircuits</i> (B). The difference in quality is very evident.	101
5.6	Four different viewpoints of the case designed to encapsulate the LDR PCB alongside its main dimensions. Extracted from <i>Fusion360</i>	102
5.7	Four different viewpoints of the printed case with the assembled PCB.	103
5.8	Four different viewpoints of the printed clamp designed to fix the LDR circuit to the monitor alongside its main dimensions. Extracted from <i>Fusion360</i>	104
5.9	The LDR circuit fully mounted and fixed to the paradigm's presentation monitor at C1 lab.	105
5.10	Raw LDR channel acquired during the Training stimuli presentation for subject 6. The various thresholds for segmentation are once again displayed.	106
5.11	Performance of the segmentation algorithm for the target stimuli.	107
5.12	Performance of the segmentation algorithm for the non-target stimuli.	107
5.13	The frequency spectrum obtained during the experiment for subject 1 (A) and subject 4 (B). Each color represents the oscillatory data from one electrode.	108
5.14	The frequency spectrum obtained during the Training experiment for subject 2 (A) and subject 3 (B). Each color represents the oscillatory data from one electrode.	109
5.15	Impedance check for subject 1 taken directly from the Micromed BRAIN QUICK software. The ground reference placed on the earlobe measured $2.7\text{ K}\Omega$	110
5.16	Interpolated channel O1 for subject 2. The magnitude is displayed also in Volts (V)	110
5.17	Electrode bridging analysis for subject 1.	111
5.18	Electrode bridging analysis for subject 2.	112
5.19	Electrode bridging analysis for subject 5.	113
5.20	The PSD (blue) and SNR (red) spectra for subject 3.	113

5.21 The PSD (blue) and SNR (red) spectra for subject 4.	114
5.22 From top to bottom, the scalp SNR distribution for subjects 1 to 6 was measured at the stimulation frequency (left column), and 1st (middle column) and 2nd (right column) harmonics. No colorbar is specified because the range is different for every topographic map, ranging from 1 to two times the maximum SNR in each estimation. Note that the anterior regions of the scalp are not covered with electrodes and the activity represented in these regions are artifacts of the <i>MNE</i> package trying to fill the entire scalp when computing the topographies.	116
5.23 The ICA solution for subject 1.	118
5.24 Analysis of ICA components 000, 002, and 004 for subject 1, corresponding to images (A), (B), and (C), respectively.	118
5.25 The Grand-Average P3 (left) and its standard deviation (right) along the 5 analyzed channels for the Training experiment across 6 participants.	122
5.26 The grand averaged Target (A) and Non-Target (B) topographies.	123
5.27 The Grand-Average N2pc (left) and its standard deviation (right) along the 7 channel pairs for the Training experiment.	125
5.28 The grand averaged RVF N2pc on the left (A) compared to the LVF N2pc on the right (B) across all channel pairs for the Training experiment.	126
5.29 The grand averaged RVF (A) and LVF (B) topographies.	129
5.30 The Grand-Average N2pcb along the 7 channel pairs for the Training experiment.	131
5.31 The grand averaged TVF (A) and BVF (B) topographies.	132
5.32 The Grand-Average N2pc along the 7 channel pairs divided into four quadrants (RVF-TVF vs LVF-TVF vs RVF-BVF vs LVF-BVF) for the Training experiment.	134
5.33 The Train/Validation results of the subject-specific P3 model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the SVM.	136
5.34 The training of the best subject-specific P3 model with all the Training session data (A) and model evaluation with all the Testing data session (B)	137
5.35 P3 performance in discriminating Target from Non-target stimuli over the 8 trial blocks.	138
5.36 The Train/Validation results of the subject-generalized P3 model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the SVM architecture and the model retrained with the Training data session achieved 67.4 % CV accuracy with 1.66 % std.	139
5.37 The Train/Validation results of the subject-specific N2pc model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found for the 7 best channels was the LDA.	141
5.38 The training of the best subject-specific N2pc model with all the Training session data (A) and model evaluation with all the Testing data session (B). The 7 most discriminative channels were used.	142
5.39 Subjects performance for the best 7 N2pc channels in discriminating Target from Non-target stimuli over the 8 trial blocks.	143
5.40 The Train/Validation results of the subject-specific N2pc model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the LDA.	144

5.41	The training of the best subject-specific N2pc model with all the Training session data (A) and model evaluation with all the Testing data session (B).	145
5.42	Subjects performance for all 14 N2pc channels in discriminating Target from Non-target stimuli over the 8 trial blocks.	146
5.43	The Train/Validation results of the subject-generalized N2pc model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the LDA architecture and the model retrained with all Training data session achieved 65.83 % CV accuracy with 2.24 % std.	147
5.44	The Train/Validation results of the subject-specific N2pc-P3 combined model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found for the 9 best channels was the SVM.	149
5.45	The training of the best subject-specific N2pc-P3 model with all the Training session data (A) and model evaluation with all the Testing data session (B). The best 9 channels were used for each participant.	150
5.46	Subjects performance for the best 9 N2pc-P3 channels in discriminating Target from Non-target stimuli over the 8 trial blocks.	150
5.47	The Train/Validation results of the subject-specific N2pc-P3 combined model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found for all 19 channels was the SVM.	151
5.48	The training of the best subject-specific N2pc-P3 model with all the Training session data (A) and model evaluation with all the Testing data session (B). . . .	152
5.49	Subjects performance for the N2pc-P3 model combining all 19 channels in discriminating Target from Non-target stimuli over the 8 trial blocks.	152
5.50	The Train/Validation results of the subject-generalized N2pc-P3 model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the SVM architecture and the model retrained with all Training data sessions achieved 88.41 % CV accuracy with 2.28 % std.	153
5.51	The Train/Validation results of the subject-specific N2pc model to discriminate RVF vs LVF target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found for the 7 best channels was the SVM.	156
5.52	The training of the best subject-specific N2pc model with all the Training session data (A) and model evaluation with all the Testing data session (B) for the RVF vs LVF discrimination with the 7 best channels.	157
5.53	Subjects performance for the N2pc model combining the best 7 channels in discriminating RVF from LVF stimuli over the 8 trial blocks.	158
5.54	The Train/Validation results of the subject-specific N2pc model to discriminate RVF vs LVF target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found for the 9 best channels was the SVM.	159
5.55	The training of the best subject-specific N2pc model with all the Training session data (A) and model evaluation with all the Testing data session (B) for the RVF vs LVF discrimination.	160

5.56 Subjects performance for the N2pc model combining all 14 channels in discriminating RVF from LVF stimuli over the 8 trial blocks.	161
5.57 The Train/Validation results of the subject-specific N2pcb model to discriminate TVF vs BVF target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the XGBoost.	162
5.58 The Train/Validation results of the subject-specific N2pc model to discriminate the 4 quadrants (RVF-TVF, RVF-BVF, LVF-TVF, LVF-BVF) of target appearance across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was LDA.	163
5.59 Subject 1 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model's train and validation loss (right).	166
5.60 Subject 2 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model's train and validation loss (right).	166
5.61 Subject 1 (A) and 2 (B) performance in predicting the correct pixel where a lateral target appears within the image for the single-trial testing data.	167
 7.1 Comparison between the VR environment of Camp Roberts already built (A) and the one created during this project (B).	172
 A.1 Page 1 of the written consent signed by every participant	187
A.2 Page 2 of the written consent signed by every participant	188
A.3 Page 3 of the written consent signed by every participant	189
A.4 From left to right, the frequency spectra of subjects 5 and 6.	190
A.5 From top to bottom, electrode bridging analysis for subjects 3, 4, and 6.	191
A.6 The PSD and SNR estimation metrics of subjects 1 (top left corner), 4 (top right corner), 5 (lower left corner), and 6 (lower right corner).	192
A.7 The Grand-Average P3 (left) and its standard deviation (right) along the 5 analyzed channels for the Testing experiment.	193
A.8 Grand average topographies of the Testing paradigm.	194
A.9 The Grand-Average N2pc (left) and its standard deviation (right) along the 7 analyzed channels pairs for the Testing experiment.	195
A.10 The grand averaged RVF N2pc (A) and LVF N2pc for the Testing session.	196
A.11 The grand averaged RVF (A) and LVF (B) topographies for the Testing session.	197
A.12 The Grand-Average N2pcb along the 7 analyzed channels pairs for the Testing experiment.	199
A.13 The grand averaged TVF (A) and BVF (B) topographies for the Testing session.	200
A.14 The Grand-Average N2pc along the 7 channel pairs divided into four quadrants (RVF-TVF vs LVF-TVF vs RVF-BVF vs LVF-BVF) for the Testing experiment.	202
A.15 Subject 3 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model's train and validation loss (right).	203
A.16 Subject 4 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model's train and validation loss (right).	203
A.17 Subject 5 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model's train and validation loss (right).	204
A.18 Subject 6 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model's train and validation loss (right).	204
A.19 Subject 3 (A) and 4 (B) performance in predicting the correct pixel where a lateral target appears within the image for the single-trial testing data.	205

- A.20 Subject 5 (A) and 6 (B) performance in predicting the correct pixel where a lateral target appears within the image for the single-trial testing data. 205

List of Tables

3.1	The confusion matrix concept for a binary problem discriminating between Target and Non-Target classes.	44
4.1	Participant demographics (mean and standard deviation for age).	67
5.1	<i>t-test</i> results of the SNR between the visual ROI and the entire scalp for all subjects (S). These testes were done for the image stimulation frequency (5 Hz) and its main harmonic components at 10 Hz and 15 Hz. The values marked as red represent the statistical significant SNR values between the entire scalp data and the visual ROI defined within the corresponding frequency assessed.	115
5.2	The sensitivity and specificity of the subject-specific P3 target vs non-target model. The chance level is 50 %.	139
5.3	The sensitivity and specificity of the subject-generalized P3 target vs non-target model. The chance level is 50 %.	140
5.4	The 7 most discriminative channels to classify target vs non-target stimuli through the N2pc ERP.	141
5.5	The sensitivity and specificity of the subject-specific N2pc target vs non-target model for the most discriminative 7 channels. The chance level is 50 %.	143
5.6	The sensitivity and specificity of the subject-specific N2pc target vs non-target model for all 14 channels. The chance level is 50 %.	146
5.7	The sensitivity and specificity of the subject-generalized N2pc target vs non-target model for all 14 channels. The chance level is 50 %.	148
5.8	The 9 most discriminative channels in classifying target vs non-target stimuli for the combined N2pc and P3 ERPs.	148
5.9	The sensitivity and specificity of the subject-specific N2pc-P3 target vs non-target model for the best 9 channels. The chance level is 50 %.	150
5.10	The sensitivity and specificity of the subject-specific N2pc-P3 target vs non-target model for all 19 channels. The chance level is 50 %.	152
5.11	The sensitivity and specificity of the subject-generalized N2pc-P3 target vs non-target model for all 19 channels. The chance level is 50 %.	154
5.12	The best model combination fill to improve each individual specificity and sensitivity. SG refers to subject-generalized models while SS represents a subject-specific model. BC acronym represents the best channel model variation, whereas AC depicts the all-channel model variation.	155
5.13	The 7 most discriminative channels to classify RVF and LVF stimuli through the N2pc ERP.	155
5.14	The sensitivity and specificity of the subject-specific N2pc RVF vs LVF model for the 7 best channels. The chance level is 50 %.	158

5.15 The sensitivity and specificity of the subject-specific N2pc RVF vs LVF model for the 14 N2pc channels. The chance level is 50 %	161
5.16 The Pearson Correlation Coefficient (PCC) and corresponding correlation statistical significance (p-value) between the predicted and real x-axis centroids for each subject in the validation set	164
5.17 The Pearson Correlation Coefficient (PCC) and corresponding correlation statistical significance (p-value) between the predicted and real x-axis centroids for each subject in the testing set	165

Abbreviations and Symbols

ANN	Artificial Neural Networks
ALS	Amyotrophic Lateral Sclerosis
API	Application Programming Interface
AR	Augmented Reality
BCI	Brain-Computer Interface
BVF	Bottom Visual Field
CV	Cross-Validation
DOC	Diseases Of Consciousness
ECOG	Electrocorticography
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyogram
EPSP	Excitatory Postsynaptic Potential
ERD	Event-Related Desynchronization
ERP	Event-Related Potentials
ErrP	Error-Related Potential
ERS	Event-Related Synchronization
fMRI	Functional Magnetic Resonance Imaging
fnIRS	Functional Near-Infrared Spectroscopy
FPR	False Positive Rate
GUI	Graphical User Interface
HD-EEG	High density Electroencephalogram
HMD	Head-Mounted Display
IFCN	International Federation of Clinical Neurophysiology
INR	Intracortical Neuron Recording
IPSP	Inhibitory Postsynaptic Potential
ITR	Information Transfer Rate
LDA	Linear Discriminant Analysis
LVF	Left Visual Field
LVFA	Left Visual Field Advantage
MBI	Mutual-Based Information
MEG	Magnetoencephalography
MI	Motor Imagery
MND	Motor Neuron Diseases
mVEP	Motion-related Visual Evoked Potentials
NF	Neurofeedback
PET	Positron Emission Tomography

RSVP	Rapid Serial Visual Presentation
RVF	Right Visual Field
SCP	Slow Cortical Potential
SNR	Signal-to-Noise Ratio
SSEP	Steady State Evoked Potentials
SSVEP	Steady State Visual Evoked Potentials
SVM	Support Vector Machine
VEP	Visual Evoked Potentials
VR	Virtual Reality
TPR	True Positive Rate
TVF	Top Visual Field

Chapter 1

Introduction

1.1 Context

Technology has come a long way these past four decades in improving the quality of life of billions of people. Based on modern tools, industrialists can produce better quality products in every sector, including biomedical engineering where technology gets implemented to improve and solve problems related to biological systems, namely the human body.

Since Licklider introduced the concept of Brain-Computer symbiosis in 1960, mankind has fantasized about conjugating the ability of computers to execute complex repetitive tasks in a very fast, accurate, and efficient manner, with the human's brain flexibility to contextualize a lot of information from our world [1]. Indeed, the human brain and computers complement each other's flaws. While the former has a large computational breadth in the sense that it is able to process information from several parallel sources, but low computational depth, related to not being able to execute an arduous and long task, computers are the exact opposite [2].

Establishing symbiosis between man and computer attempts at trying to reduce communication time linking both systems, to ultimately improve the interaction and possibilities of a new and better version of human beings.

1.2 Motivation

Albeit establishing a symbiotic relationship between computers and human brains may unlock countless possibilities in augmenting healthy individuals' lives, the main driver for integrating both systems is substituting impaired function in humans to restore independence in motor-disabled individuals [3]. According to the World Health Organization (WHO), 121,800,000 people worldwide require a wheelchair to move with an estimated of further 110 million requiring wheelchairs but not having access to them [4]. Further evidence suggests an increase in the prevalence of Motor Neuron Diseases (MND) globally, including hereditary spastic paraplegia, amyotrophic lateral sclerosis (ALS), spinal muscular atrophy, primary lateral sclerosis, and pseudobulbar palsy progressive muscular atrophy, with 268,673 subjects suffering from them currently [5]. Considering

that verbal communication is also a form of motor ability, these diseases affect much more than movement.

Brain-Computer Interfaces (BCI) based on Electroencephalographic (EEG) activity propose mitigating non-invasively these limitations in human conditions by decoding information directly from the brain and translating it into commands. Despite still facing some challenges, a wide variety of studies have emerged proving the undeniable uses of these technologies in several levels of disability, namely spelling applications [6], bionic limb control [7], environmental navigation, particularly in wheelchair control [8], and even patients in coma or vegetative states, also denominated Diseases Of Consciousness (DOC) [9]. With the latter condition being misdiagnosed in approximately 41% of the cases where some form of consciousness and awareness remain [10], BCIs can improve diagnosis before addressing the communication and motor ability issue.

These interfaces not only can replace hindered motor function with bioengineered devices but also can help restore it. Recent research indicates using BCIs induces neurological recovery and improves functional (the brain's ability to move functions from a damaged area of the brain to undamaged areas) and structural (the brain's ability to change its physical structure) neuroplasticity at a subclinical level and thus, substantially aid the neurorehabilitation of stroke survivors improving motor function, mainly in the upper body [11].

Many BCI solutions rely on the subject's gaze to operate, both directly by incorporating ocular movement or shift of attention with the neural activity and indirectly when direct focus on a certain target is required to evoke targeted neural activity with discriminatory power [12]. However, some individuals affected with the motor disabilities described might also lose the ability to reorient their eyes, motivating the necessity of considering novel gaze-independent solutions [13].

With improvements in the reliability and usability of EEG-based BCIs, a shift in demands and requirements occurred with the devisement of promising new applications of the technology to healthy individuals instead of only disabled individuals [14]. However, because of the similarities in the algorithms engineered regardless of the application, efforts to expand the entertainment applications of BCI for healthy users will inevitably contribute to boosting the field directed for health-related purposes, just as these novel approaches benefited from substantial research on BCIs for disabled users [14].

An example of these non-medical related uses for BCI technologies is the use of drones in combination with the brain [15, 16]. A symbiotic relationship with an aerial vehicle entails situations that can significantly aid rescue missions and tremendously improve visual search efficiency when coupled with Artificial Intelligence (AI) systems, a field denoted as hybrid intelligence [17]. In fact, drones can not only be employed in the surveillance of a vast amount of land very quickly to search for targets but also aid in determining the safest possible path to said target [18].

By displaying aerial images to individuals, the brain mechanisms underlying an object of interest detection and their spatial appearance on the visual field can be borrowed through Electroencephalography (EEG) and transferred to a computer where learning algorithms can learn patterns on specific classes of neural data and cognitive processes to improve visual search paradigms [19]. Notably, these neural processes can be detected and extracted early in the visual processing stream

at the limits of consciousness, when visual information hasn't been yet fully contextualized and integrated. This means that the BCI application can pick up EEG data related to object recognition and its spatial location before the individual is fully aware of the type of object he saw [20]. This type of paradigm has been successfully used in detecting objects of interest to significantly improve high-dimension aerial imagery review speeds [17] and even in unsupervised labeling of images for AI systems [21]. In the aerial imagery department, a horizontal spatial localization of the targets within the images has also been tested [19].

1.3 Objectives

This project is enrolled in a line of research that aims to establish symbiosis between humans and machines to automatize robotic vehicles' behavior with physiological data. Subsequently, the goal of this work is to expand this symbiotic relationship between the human brain through EEG acquisition and a vehicle such as a drone to optimize search and rescue missions in large and possibly inaccessible terrains. To achieve a symbiotic relationship between an Unnamed Aerial Vehicle (UAV) and a BCI, a hybrid intelligence method that is capable of identifying targets of interest and their location within aerial images and the corresponding environment during rescue missions depending solely on EEG data was developed.

Firstly, a visual paradigm was devised to validate the target identification focused on the search and rescue application. Secondly, if identified within the aerial imagery, the algorithm will also provide a gaze-independent spatial localization of these targets. It has been shown that the EEG contains information not only correlated to the detection of an object of interest within visual stimuli but also to the position of a target along the horizontal [19] and vertical axis [22]. For this, this project was divided into two experiments, a first **Training** paradigm where the algorithm was validated and a second **Testing** paradigm where the algorithm will be tested.

In the **Training** experiment, high-altitude aerial images containing images within a wide variety of different environments was divided into several sub-pictures and presented at a 5Hz rate to the participants. In 10 % of these pictures, a defined human target was present at several putative positions within the image, and each time, the participant has to mentally count its appearance. The EEG data evoked during these events was used to develop and train a machine learning algorithm that can detect the target's presence and provide a spatial estimation of its localization within the images. Once validated with the single-trial **Training** data session, the BCI actuation models for target identification and localization was evaluated in the single-trial **Testing** experiment. However, the **Testing** protocol was presented with a completely different and independent set of images, containing not only a different environment but also different human target characteristics, in an attempt to determine generalization to different visual properties, *i.e.*, if the brain is able to reproduce the results on an unseen environment and human target at fast paces. Achieving this would mean that the brain generalized not only the target vs non-target discrimination but also its localization component, guaranteeing that the brain-in-the-loop approach for search and rescue can be used in novel scenarios with different visual properties during the real-time

application. Moreover, this type of generalization achieved by the brain hints at the methodology proposed here working on other applications, including in the mitigation of the gaze-dependent uses of BCIs in the biomedical field.

1.4 Achievements

This thesis focuses on the BCI performance regarding high-throughput aerial imagery for search and rescue missions. Not only a target identification within the images is asserted by also its spatial localization within the environment for 6 subjects. The methods used here yielded target identification results on par with the literature on similar designed high-throughput studies, even though this methodology has never been applied to search and rescue scenarios before. Similarly, the blind distinction between right and left visual fields also produced comparable to the literature. In contrast, the horizontal target centroid results achieved here was greater than on the literature. However, it is important to note that these results were obtained by testing the models with much more data than studies found on the literature, meaning the assessment is more robust.

This thesis also fills a gap in the literature regarding the generalization of the brain to novel aerial imagery with different visual properties in images presented at high speeds. These results indicate the brain can indeed maintain the discriminatory performance in all classifications tasks tested to these novel images comprising unseen environments and human targets surrounded by distractors. This indicates that good performance can also be achieved during posterior real-time scenarios. When comparing to a computer vision model designed to identify humans, the BCI also performed better.

1.5 Thesis Structure

This thesis, apart from the introductory section, will include six more chapters.

Chapter 2 provides all background concepts necessary to start a project of this nature, comprising a description of the historic achievements and evolution regarding BCI technology, its main types, a review of the EEG biosignal including a brief and very simplified report of the main anatomic regions of the brain with the corresponding function, and a terminology distinction of Extend Reality spectrum. Moreover, this chapter will describe the typical BCI framework, including the signal pipeline from the brain into executable commands and the main paradigms used to make prominent the brain patterns of interest.

On the other hand, Chapter 3 will elaborate on the main parameters, both in terms of hardware and algorithms available in the literature associated with the usage of BCI paradigms combined with target identification and localization. A description of the main feature reduction methods, classification metrics, and learning algorithm architectures will also be given.

Chapter 4 delineates the proposed hardware and pipeline solutions to tackle the BCI application proposed here.

Chapter 5 describes all the results achieved throughout this project, including a detailed discussion of their meaning in the context of the BCI application researched.

Chapter 6 will present the main conclusions obtained with the realization of this dissertation.

Finally, Chapter 7 proposes ways to investigate further the concepts and methodology used to improve the BCI function presented.

Chapter 2

Brain-Computer Interfaces (BCI): A Theoretical background

This Chapter focuses on describing the theoretical volley of concepts adequate to the realization of this project. The relevant technical terms related to BCI technologies will be introduced here. As a result, not only a brief historical overview of the BCI technology will be given, but also a discussion on the main types of BCI, the EEG as an acquisition tool, the main areas of the brain including their functions, and the terminology on the categories of virtual environments.

Once the main concepts are thoroughly discussed, the building blocks that establish the bridge between the brain and executable commands in a computer will be explained, with the several stages that the signal flows through to convey each BCI command detailed below. Furthermore, the main paradigms used to optimize the extraction of certain brain patterns will be reviewed to analyze the main options available to consider when developing a BCI system. Finally, the method used to evaluate BCI performance is also mentioned.

2.1 The BCI evolution: A brief historic perspective

Since the discovery of the Electroencephalogram in 1929 by Hans Berger [23], interest in unraveling the meaning of the ongoing brain activity that intrinsically occurs during various processes bloomed. A few decades later, ambitions of using this information to bypass some physical limitations of the human body evolved into what is designated today as Brain-computer interfaces (BCI). BCIs are devices meant to translate user intent mediated by brain activity into a set of actions. Given that mental activity has correspondent bioelectrical brain activity, the ability to record these signals, process them, and discriminate intended function, makes BCIs a valuable technology for various applications. The term BCI was first coined in 1973 as a concept and described "as a means of control over external processes such as computers or prosthetic devices" by the Belgian researcher Jacques J. Vidal working at the University of California [24]. Although dormant for some years, the field of BCIs received one of its seminal papers in 1988, known as the "P300-speller" [25]. It proposed a way for a user to choose a letter on a 6-by-6 letter matrix using

only the brain waves. Utilizing a specific brain pattern related to stimuli processing designated Event-Related Potentials (ERP) (further detailed in section 2.6.1), the researchers, after presenting the set of characters on a computer screen, asked the participant to choose one and each time it was intensified on the screen, to mentally count it, thus triggering a measurable P300 component [25]. To improve efficiency, the intensification was not done on each individual character but rather on each column and then row randomly on the matrix. Thus, after a set of repetitions, the desired character was defined by the intersection of the column and row that separately generated a P300 component while being intensified/flashed. Albeit being achieved by healthy subjects, this innovative approach opened the field of BCI to one of its core applications, the potential to enable communication of motor-impaired but cognitively intact users with the outside world.

Afterward, researchers all around the globe began to focus on the sensorimotor activity from the cortex [24]. In 1991 experiments such as controlling a mouse cursor up and down via neuro-feedback emerged, where participants, after some training, were able to perceive the relationship between their mental state and the cursor movement in real-time [26]. In other words, subjects learned to voluntarily modulate the *mu* rhythm amplitudes of the *central sulcus* where sensorimotor information is generated by observing the resulting bi-directional cursor movement on-screen. By comparing the *mu* peaks to 5 voltage ranges, the computer translated it into the upward (highest amplitude) or downward (lowest amplitude) cursor movements at different speeds [26]. During this period, the first Motor Imagery (MI) paradigms where imagined limb movements were translated into commands also started to emerge [24].

Throughout the 1990s, most of the research that came to define the field of BCIs as a whole was done, setting the necessary prerequisites for the extensive and multidisciplinary field it is today. In this decade, BCI research evolved into a field of its own resulting in the first International BCI meeting held in 1999 with about 50 participants and 22 different research groups [24]. The rapid increase of computational power and technology development in concomitance with a better understanding of the brain that emerged from the last three decades of research sparked the development of new experimental paradigms, the exploration of other brain activity recording techniques, and the development of new algorithms capable of more accurately representing, cleaning, extracting and discriminating the targeted brain activity. Figure 2.1 showcases the evolution of interest around BCIs by tracking the increased number of papers published discussing the technology throughout the years.

Focusing on recent history, the technological improvements have catapulted the term BCI to the mainstream, leading to the creation of several companies that try to penetrate these systems in the market at affordable prices, namely *OpenBCI*, *Neuroosity*, *InteraXon Inc*, *G.TEC* and others trying to solve more serious issues such as Elon Musk's *Neuralink*. Because of this, BCI technology has been recently crowned the technology of the year 2023 according to *Nature Electronics* [27].

Despite being initially conveyed to replace or restore function in motor-impaired subjects, interest grew for BCI systems to be explored in more complex environments, expanding the scope of potential applications of the technology to other domains, namely entertainment contexts [29].

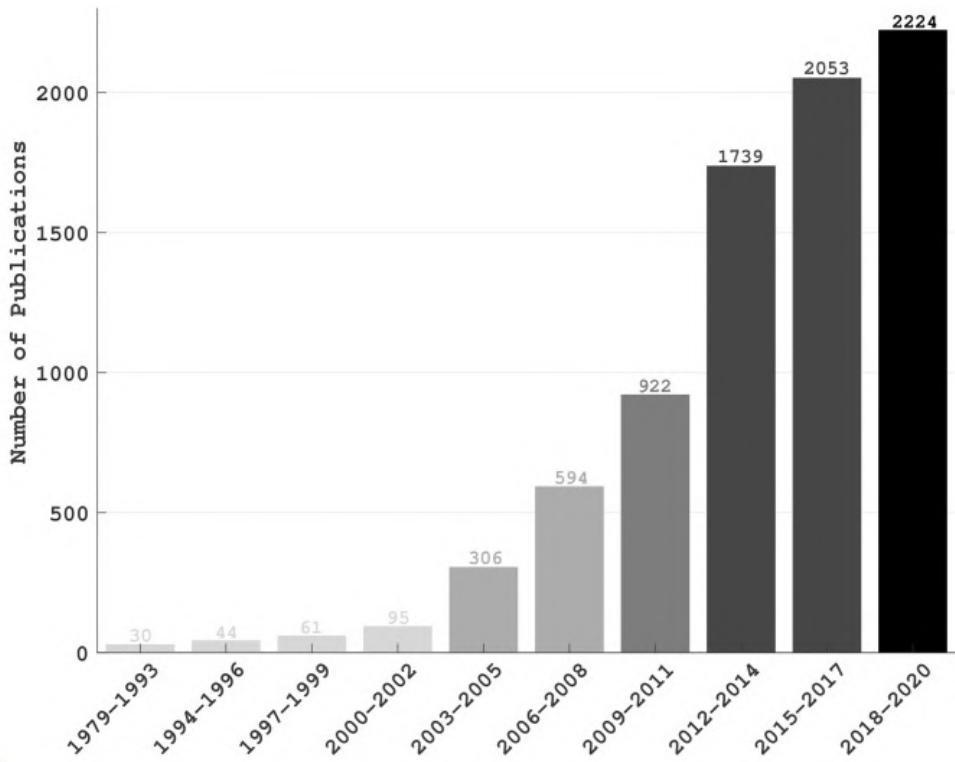


Figure 2.1: The number of papers published over the years that were identified from the *PubMed* database using the keyword "Brain-Computer Interface" [28].

This led to the fusion of BCI technology with Extended Reality (XR), unlocking brain research to many interesting possibilities.

2.2 Categorization of BCI systems

The exponential growth of the field over the last few decades discussed in *Section 2.1*, was logically accompanied by the diversification of the types of systems developed. BCI technology can be categorized into four main groups according to their dependability, recording technology, mode of operation [30], and type of user interaction [14]. Figure 2.2 depicts the branching of the different types of BCI systems. Firstly, regarding the dependability factor, dependent BCIs require the integration of residual motor capabilities with the brain signal such as gaze control-based systems [30]. In contrast, independent BCIs systems, only require direct brain-related information and do not use any motor control from the user [31]. When compared to the former, the latter type is better suited for severely impaired patients or stroke patients who are unable to provide the additional non-brain data reliably [30]. However, it is important to consider that using both systems concomitantly is possible and might better suit the application insofar as integrating both brain and non-brain physiology has been shown to improve performance [32].

The mode of operation category of BCIs regards the temporal dependency of the interaction between the system and the user. Simply put, if the user controls the action of the system at its

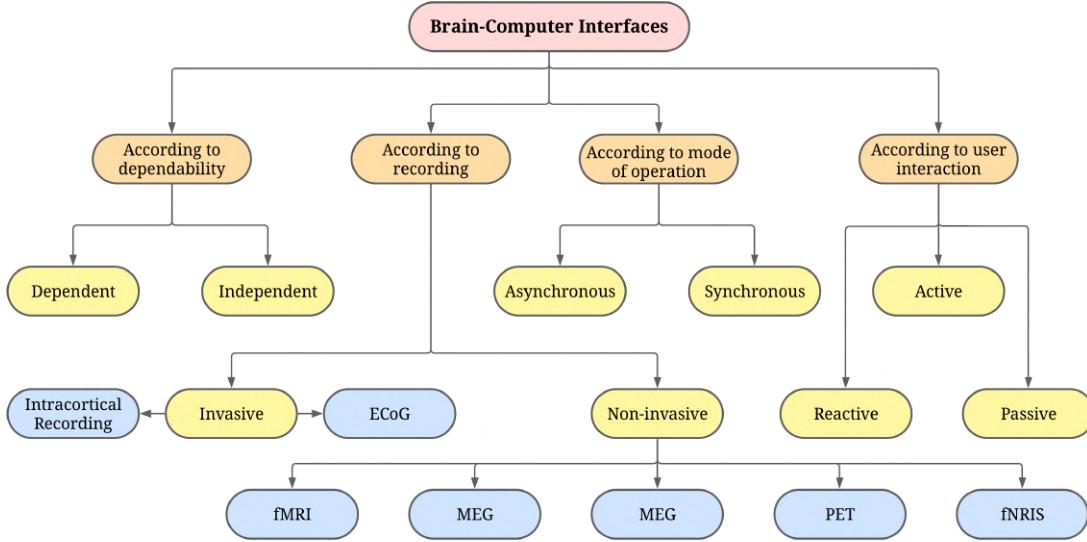


Figure 2.2: Categorization of the main types of BCI systems by dependability, recording method, and mode of operation. The acronym ECoG refers to Electrocorticography, fMRI to functional Magnetic Resonance Imaging, EEG to Electroencephalography, MEG to Magnetoencephalography, PET to Positron Emission Tomography, and fNIRS to Functional Near-infrared Spectroscopy. Adapted from [30]

own desired time, the interaction is time-independent and the BCI is classified as asynchronous, whereas if the interaction is cue-based and information is presented to elicit a certain response, the system is time-dependent or synchronous [30].

BCI technology can be further categorized into active, reactive, and passive depending on the type of interaction with the user. While in active BCIs, the user consciously controls the application through purposeful modulation of neural activity independent of external events, the passive type is able to decode the user state from arbitrary and involuntary brain data without intentional control by the user [14]. Lastly, reactive BCIs work by acquiring brain data related to the presentation of stimuli enabling indirect control by the user. The original P300 speller falls under the latter as it requires the presentation of stimuli for its elicitation. The active and reactive types are generally targeted toward individuals with restricted motor abilities whereas passive BCIs are aimed more at able-bodies users [33].

Finally, concerning the different techniques for recording brain activity, BCI systems can be further categorized depending on their invasiveness to the user. The invasive methods, composing the Intracortical Neuron Recording (INR) and Electrocorticography (ECoG), require surgery for the placement of the electrodes inside the skull, making direct contact with brain tissue [30]. Conversely, non-invasive methods do not require any type of surgery insofar as the brain waves recorded are captured from either electrodes placed on the skin or from machinery surrounding the cranium [34, 30]. The non-invasive methods used in BCIs include the *direct measures* that directly reflect neuron's activity, *i.e.*, Electroencephalogram (EEG) and the Magnetoencephalography (MEG), and the *indirect measures* that reflect brain hemodynamics (blood flow) and metabolism,

specifically the Functional Magnetic Resonance Imaging (fMRI), Functional Near-Infrared Spectroscopy (fNIRS), and the Positron Emission Tomography (PET) [34].

The mentioned techniques related to the acquisition of data within the brain all have certain characteristics that yield advantages and disadvantages depending on the type of application being considered. In Figure 2.3, a comparison of the most important metrics that can be taken into account when addressing the adequacy of the acquisition technology to a certain purpose is represented, ranking them in terms of their cost and both spatial and temporal resolution. Despite the lack of an optimal tool for all situations, the EEG unquestionably is the most attractive solution for the majority of BCI applications. Albeit providing low spatial resolution, the EEG is also an inexpensive, noninvasive, and portable technology with a very high temporal resolution when compared to other recording techniques. A study reviewing BCI-published papers using invasive methods (ECoG), fMRI, NIRS, MEG, and EEG has found that the usage of the latter increased significantly from 2007 to 2011, with 45% and 68% of BCI research conducted in those years, respectively, being EEG-based [35].

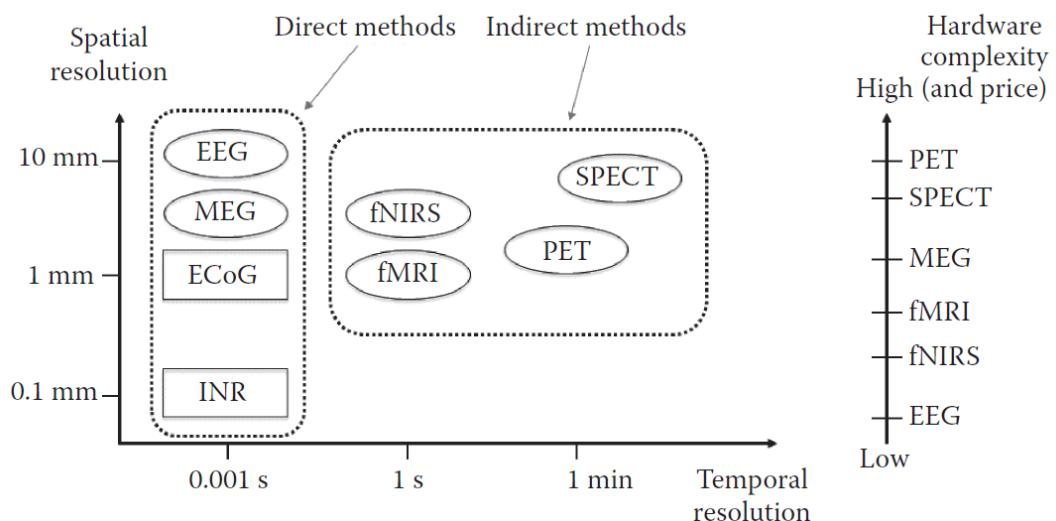


Figure 2.3: Comparison in the temporal resolution, spatial resolution, and complexity planes of all invasive (rectangular) and non-invasive (oval) brain activity recording technologies used in BCI systems [34]. The acronym ECoG refers to Electrocorticography, fMRI to functional Magnetic Resonance Imaging, EEG to Electroencephalography, MEG to Magnetoencephalography, PET to Positron Emission Tomography, and fNIRS to Functional Near-infrared Spectroscopy.

2.3 The Electroencephalogram

When developing systems based on the acquisition of biosignals, some basic knowledge of the tool used is required, namely how the electroencephalogram captures brain activity, the system of electrodes including their terminology, and the main properties of the biosignal. Such topics are introduced in the present section.

2.3.1 A measure of brain activity

The EEG is a brain activity recording technique that measures electrical potentials generated by neuronal activity during various mental and cognitive processes using a system of electrodes on the scalp. The electrical activity collected by the electrodes thus comes from the neuronal activation that can arise in two distinct ways:

- Through the rapid depolarization of membranes caused by the flux of K^+ and Na^+ ions, which generates action potentials in axons and dendrites [36]. These potentials generate brief electrical fields.
- Through synaptic activation mediated by several neurotransmitters, which generates a subtle electric field also called postsynaptic potential [36]. This potential can be either inhibitory (IPSP) or excitatory (EPSP). The electric field generated in this way is slower.

However, not all brain activity can be captured by an EEG. In general, only the synchronized activity of pyramidal neurons, oriented parallel to each other, allows the generation of electrical fields intense enough to be detected by electrodes on the surface of the head [37]. Neurons in deeper parts of the brain, such as the cerebellum, are not arranged parallel to each other, giving rise to electrical fields that cancel each other [38]. The same effect occurs in sulci.

As already briefly mentioned, the EEG allows the identification of cognitive processes at the speed of cognition, with temporal precision in the order of milliseconds, being perfect for studying brain function, both normal and pathological [39, 40]. In contrast, its main disadvantage lies in providing imprecise spatial information when compared to imaging techniques, mainly from deeper parts of the brain below the cortex. In other words, the relatively low spatial resolution provided has sometimes to be complemented with imaging techniques such as MRI, in situations where it is necessary to accurately locate the source of electrical activity [37]. However, with current technological and electrode system advancements, it has become possible to overcome this disadvantage.

2.3.2 System of electrode placement

Standardizing the placement and nomenclature of the electrodes was an essential step in the development of the EEG. Thus, the 10-20 system defines the position of the electrodes based on the circumferential measurements of the skull (cephalometry) [41]. It represents portions of 10% and 20% of the scalp measured longitudinally, from the Nasion to the Inion, through the vertex (Figure 2.4. A), transversally from the left to the right preauricular point, also through the vertex (2.4. B), and from the anterior-posterior direction through the regions above the ears on each side (2.4. C). Regarding the nomenclature of the channels, the letters O, P, T, and F depict the occipital, parietal, temporal, and frontal lobes, respectively. The letter C indicates the central region, and the electrodes on the midline are noted with the letter z (*e.g.*, Cz) for zero. The numbers are incrementally counted from the midline with even numbers for the right and odd

numbers for the left hemisphere. Intermediate sites are named with two (or more) letters where FCz stands for the frontocentral region, for example. This system, comprised of 25 electrodes, is the basic configuration admitted and recommended by the International Federation of Clinical Neurophysiology (IFCN) for clinical use in both adults and children [41, 42].

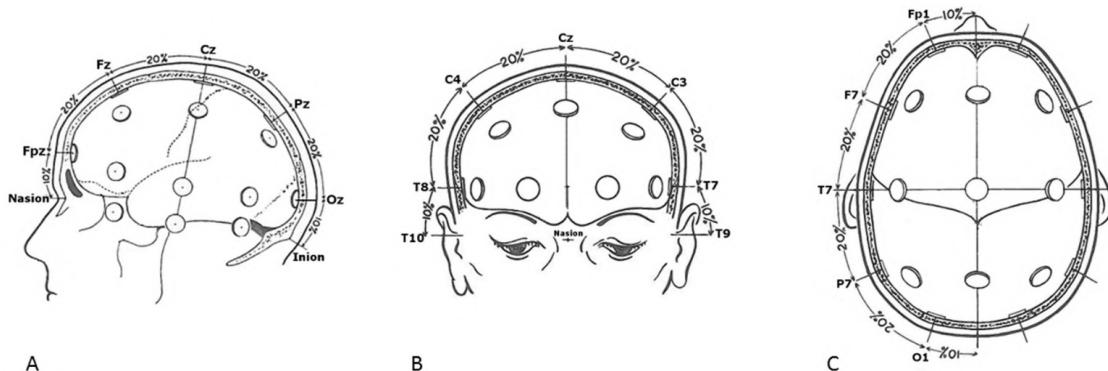


Figure 2.4: Illustration of the 10-20 EEG system derived from the circumferential measurements of the skull [41].

With the advancements in the engineering of amplifiers, higher-density arrays of electrodes (HD-EEG) have emerged, such as 10-10 and 10-5 with 64-256 electrodes, based on the same principles of cephalometry. These systems are mainly used in the context of pre-surgical evaluation in order to accurately identify and locate electrical sources, requiring, at a minimum, 64-76 electrodes to avoid localization errors [41]. A topological representation of the 10-10 system can be seen in Figure 2.5

2.3.3 The rhythm of the brain

The brain is one of the most complex systems in the universe. Inevitably, the biosignal captured by the EEG is highly non-linear and non-stationary with quasi-stationary segments (250ms) [43]. An EEG signal generally presents amplitudes and frequencies ranging from $0.5 \mu\text{V}$ to $200 \mu\text{V}$ and from 0.5 Hz to 60Hz, respectively [24], with frequencies above 50Hz being rarely relevant [37]. Through methods of decomposing signals in their frequencies, namely the Fourier transform, it is possible to identify five main groups of waves, each relevant to a specific mental state in which low frequencies are associated with reduced responsiveness to stimuli [37]. These are:

- **Delta δ (0.5-4 Hz):** The *Delta* waves represent the lowest frequency and highest amplitude component of neural oscillations in the brain. These typically occur frontally during deep sleep in normal-functioning adult brains and posteriorly in awake infants [44, 45]. They are pathological if detected in awake adults, being associated with deep midline lesions, metabolic encephalopathy hydrocephalus, or subcortical lesions [44].
- **Theta θ (4-8 Hz):** *Theta* oscillations are typically observed in children [44], particularly in the parietal and temporal regions. These waves are independent of the task being executed

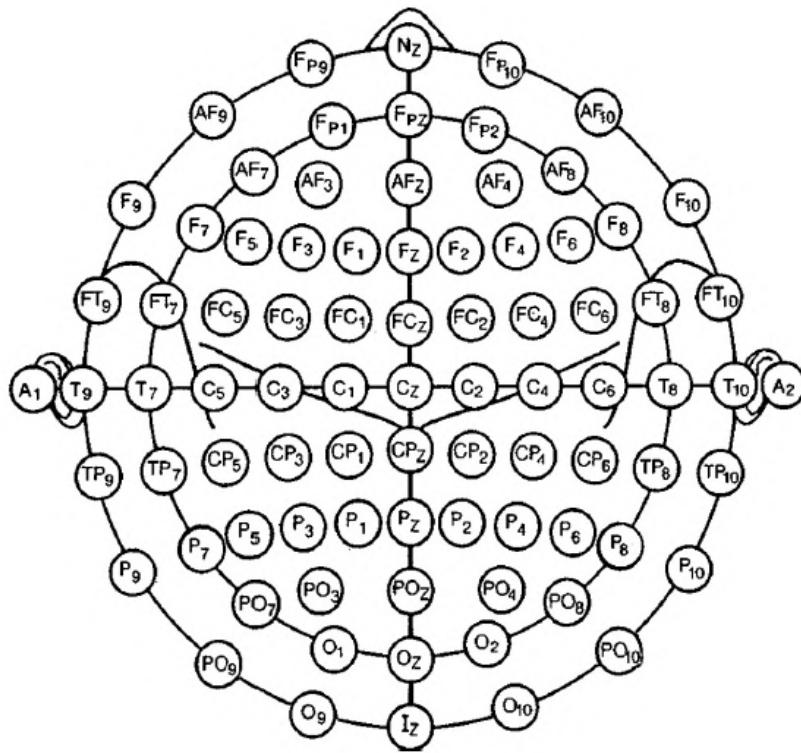


Figure 2.5: Illustration of the 10-10 EEG system. Adapted from [41].

and are often associated with a state of drowsiness, arousal [44], and during emotional stress, mainly related to disappointment and frustration in adults [45]. These waves are also present during the Rapid Eye Movement (REM) sleep, when dreaming and memory consolidation takes place. In adults, an overabundance of *Theta* waves can represent atypical activity in the brain [44].

- ***Alpha* α (8-12 Hz):** The *Alpha* waves are linked to wakefulness, creativity, effortless alertness, and relaxed states in adults, mainly with the eyes closed [44, 45]. However, they are nonexistent during deep sleep. They are commonly observed in the occipital region despite also being detected in the frontal and parietal regions. However, unlike the *Theta* waves, *Alpha* frequencies cannot occur in the cortex without cortical connections to the Thalamus [45].
- ***Beta* β (15-25 Hz):** The *Beta* waves can be recorded from the parietal and frontal lobes and are related to active states and high focus, namely during problem-solving, active attention, and thinking [44, 45].
- ***Gamma* (γ) (25-60 Hz):** The *Gamma* neural oscillations compose the highest frequency and lowest amplitude components, being correlated with integrating information from parallel sensory pathways (*e.g.*, vision and sound) [44]. Furthermore, these waves are also thought to combine populations of neurons.

In addition to the aforementioned frequencies in brain activity, two other neural oscillations must be discussed as they are very important in the context of BCI applications. These may overlap the frequency ranges specified above but are given a different name if generated over a particular area of the brain. The sensorimotor rhythms (SMR) range from 12 to 15 Hz and as the name suggests, originate in the somatosensory and motor cortices during motor execution or imagined movement [24]. Contrastingly, if the oscillations originate from these regions and comprise frequencies between 7 and 13 Hz they are designated the *mu* or rolandic rhythms. Aside from the latter frequency band that overlaps the *Alpha* waves upon elicitation, the mentioned neural oscillations and their corresponding characteristics can be seen in Figure 2.6

Neural oscillations pose a crucial mechanism for enabling coordinated and synchronized activity across neural regions during normal brain function, with lower frequencies (*Delta*, *Theta*, and *Alpha*) establishing communication over distant regions in the brain, while faster waves exchange information over local networks [46].

Brainwave	Sample Pattern	Frequency (Hz)	Amplitude (μV)
Delta		0.5–4	100–200
Theta		4–8	5–10
Alpha		8–12	20–80
SMR		12–15	
Beta		15–25	1–5
Gamma		25–60	0.5–2

Figure 2.6: Illustration of the different frequencies transmitting information in the brain [24]. It is important to note that the frequency ranges defined here may vary slightly depending on the source, which can generate some confusion when addressing them [44].

The human brain produces a specific frequency spectrum pattern which can be used to initially identify if the EEG was correctly recorded. It is characteristic for the EEG to have a $\frac{1}{f}$ shape, meaning it produces more signal power at lower frequencies which gradually decreases with an increase in frequency. This shape is demonstrated in figure 2.7.

2.4 The geography of thought

The human brain is the control center of our body and thoughts, being responsible for regulating, integrating, processing, and coordinating all sensory pathways to construct consciousness and the world around us. The way it achieves this is far from being completely understood. However,

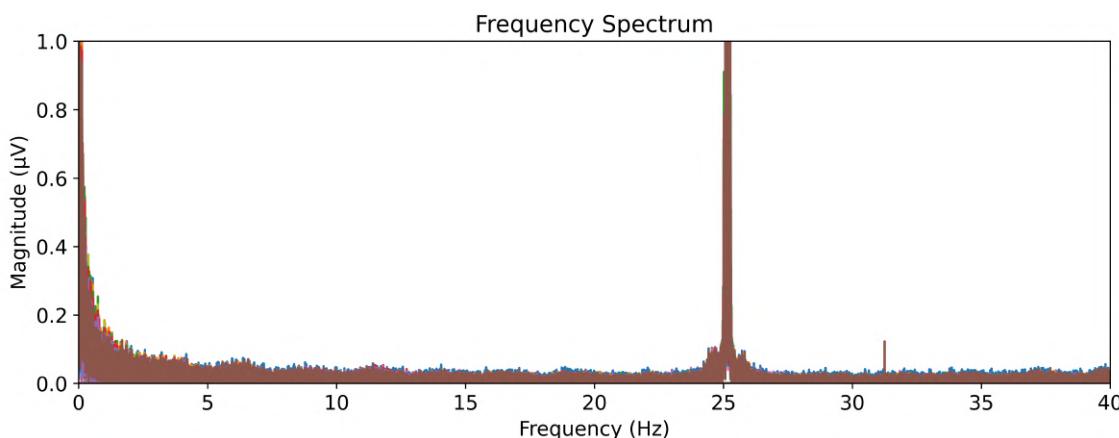


Figure 2.7: The $\frac{1}{f}$ shape characteristic of the EEG recorded on my scalp. Notice the presence of a harmonic oscillation related to line noise at 25 Hz, a common occurrence when capturing the EEG at the BRAIN laboratory at INESC TEC due to the 50 Hz of the power line.

with the combined efforts of neuroscientists and neuroanatomists, certain cognitive and sensory functions have been mapped to specific regions in the brain.

Since the signal captured from the EEG has spacial specificity, a review of the main anatomical components and corresponding functions is also required to fully understand the origin and language of the biosignal. Since neuroanatomy is an extremely complex and extensive subject, a description of only the main brain areas targeted for BCI research is depicted in this section.

2.4.1 The human cortex

Within the context of the EEG, the most significant region is the cerebral cortex, the outer layer of the brain. This region is not a flat surface but is instead characterized by numerous indentations known as *sulci*, whereas the raised areas that are separated by the *sulci* are referred to as *gyri* [47]. This cortical folding allows for a large surface area relative to the volume inside the skull. The cerebral hemispheres, which are the two main parts of the brain, are divided into four main regions: the frontal, temporal, parietal, and occipital lobes, as shown in Figure 2.8.

The **frontal lobe** is the region of the brain responsible for the definition of being human, being involved in personality, rationality, and the ability to generate complex thoughts [47]. It governs behaviors of restraint such as delaying gratification, self-governance, inhibiting inappropriate social responses, judgment, perseverance, and concentration [47]. Moreover, in addition to being the source of initiative thoughts, namely motivation, spontaneity, and creativity, the frontal lobe also is in charge of the ability to successfully complete sequencing tasks and various other cognitive functions.

The **temporal lobe** is involved in the processing of auditory information and in the understanding of the semantics in language and vision [47]. Furthermore, it plays a role in memory storage since the temporal lobes include the *hippocampus*, the region of the brain that consolidates information as memories.

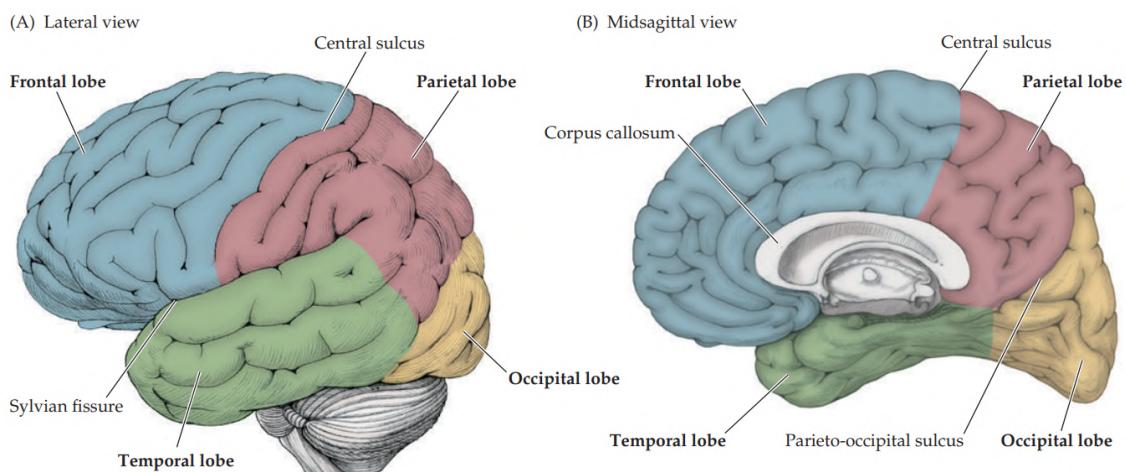


Figure 2.8: The Lateral (A) and Midsagittal (B) views of the human brain with the four main lobes highlighted in different colors and the cerebellum illustrated in gray below the occipital and temporal lobes. The midsagittal view must be interpreted as the brain cut in half dividing the left and right hemispheres, revealing the structure that connects them both, the *corpus callosum*. Adapted from [47].

The **parietal lobe** is the one responsible for the processing of somatosensory information, namely the sensation of touch, temperature, pain, and management of the gustation, olfaction, auditory, and visual modalities [47]. Furthermore, it is also involved in the perception of the body in space which is termed proprioception, allowing navigation within a complex environment.

Finally, the posterior portion of the brain is designated the **occipital lobe**. The main function of this region is the process of visual information as it contains most anatomical areas of the visual cortex [47].

2.4.2 The visual cortex

As already described, most visual processing brain areas are included in the *occipital lobe* but these also communicate with the *temporal* and *parietal* lobes to extract meaning and integrate visual information with other sensory inputs through different pathways.

The visual system begins by the capturing of images in the human body's primary visual receptor, the *retina*. It then travels subcortically through the two main parallel pathways referred to as *retino-geniculo-cortical* pathways [48]. These two pathways can be seen in Figure 2.9 (a) and involve information captured on the retina, projecting through the *thalamic Lateral Geniculate Nucleus* (LGN), the *primary visual cortex* (V1) and finally through the dorsal and ventral visual streams. These two pathways compose two different types of ganglion cells, the *magnocellular* (M), represented by the color red, and *parvocellular* (P), represented by the green color, conveying two types of visual information at different speeds [48]. The M channels are larger and conduct information faster, displaying a strong preference for high temporal frequencies and low spatial frequencies, meaning it is specialized for rapidly presented, coarsely textured visual information, respectively. They are not sensitive to color but exhibit high sensitivity to contrasts. Its response is

characterized by rapid transient activity, facilitated by fast axonal conduction speeds [48]. Moreover, these also project to the brainstem driving visual attention and eye movements. Conversely, the P channels represent the smaller and slower channels with low contrast sensitivity, being tuned to high spatial and low temporal frequencies, *i.e.*, higher detail and slower visual stimuli [48].

Despite the M and P cell type pathways aligning with the ventral and dorsal stream distinction, evidence demonstrates that the subcortical and cortical pathways don't have a straightforward correspondence, and M and P signals substantially intermingle within V1, thus contributing with different proportions of each to the dorsal and ventral streams [49], as evidenced by Figure 2.9 (a). Therefore, while the dorsal stream consists mainly of M channels, thus providing faster conduction, the slower ventral pathway has a more balanced distribution of both channels [48]. The dorsal stream also differs from the ventral pathway regarding its inputs. Through M channels, in addition to the dorsal signal path enabling input through V1 to the motion processing region V5 directly, also called *middle temporal visual area* (MT), the subcortical LGN can also feed neural activity directly to V5, whereas in the ventral stream, information does not bypass V1 [48].

As a result of the mismatch in conduction rates, both the M and P pathways project parallel and separate information that arrives at different times in V1, enabling the tight synchronization of separately timed visual processes with complementary information. This results in the faster M inputs being quickly channeled through the dorsal stream for the analysis of motion and the guidance of arm and eye movements towards targets [48]. Later, some M fibers feed back through the dorsal stream and combine with delayed P channel inputs in V1. Together, they later project through the ventral stream to the inferior temporal cortex, which plays a crucial role in object and face recognition [48]. This cascade of information processing through both signal paths demonstrates the reason behind the dorsal and ventral streams being commonly described as the "where/how" and "what" pathways, respectively, as illustrated in Figure 2.9 (b).

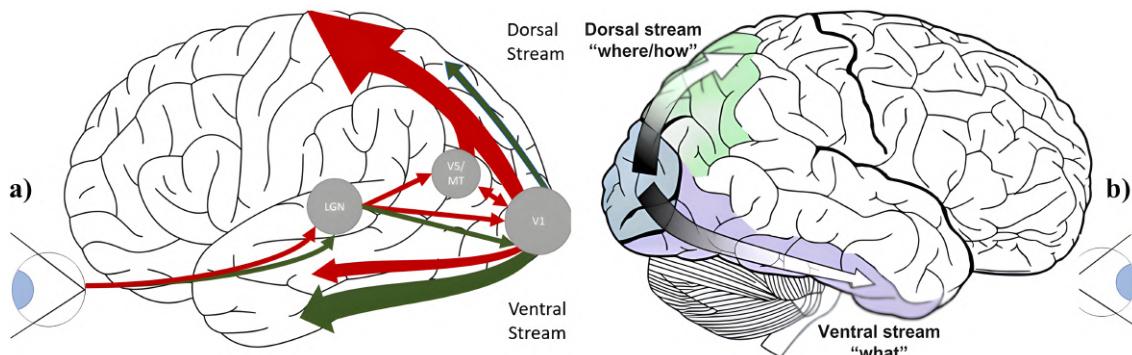


Figure 2.9: Illustration of the *retino-geniculo-cortical* pathways [49]. Image (a) encompasses the *magnocellular* (red) and *parvocellular* (green) pathways and their contribution to the dorsal and ventral streams of visual information. It also shows the subcortical communications between LGN, V5 and V1 areas. Adapted from [48]. Image (b) does not make the distinction between both cell pathway types to the ventral and dorsal streams but includes their respective summarized function. Adapted from [49]

Expanding on the visual processing regions in terms of their function and anatomy, Area V1,

known as the *striate cortex*, is organized in a retinotopic manner [49]. In other words, there is a mapping between the spatial distribution of neuronal regions and specific areas of the retina and thus the visual field. Functionally, V1 is responsible for detecting fundamental visual attributes like colors, bars, edges, and their orientations. Moving downstream beyond V1, we enter the region of *extrastriate* visual areas [49]. The first of these areas, V2, receives the bulk of its input from V1 and primarily projects to visual regions V3 and V4. V2 retains its retinotopic organization but encompasses larger receptive fields compared to V1. It aligns V1 features to execute moderately complex visual tasks, including the detection of spatial frequencies, textures, and the separation of foreground from background [49]. Visual area V3, which maintains its retinotopic organization, builds upon the groundwork laid by V2. V3 excels at generating more intricate, invariant representations. With its large receptive fields, V3 demonstrates an enhanced capability to detect complex visual features related to orientation, motion, depth, and color in visual stimuli [49]. In the context of stereoptic processing (depth), it's important to note that binocular disparities, which are differences in the images received by each eye, are present in all stages of the visual processing hierarchy, starting from V1 [49]. Visual areas V2 and V3 exhibit increasing capabilities for processing depth and their proficiency is influenced by the size of their receptive fields. However, it's the visual area V3A that is thought to hold the specialized role in processing stereoptic depth and, even though the evidence remains inconclusive, this area may also be responsible for calculating relative disparities between pairs of visual stimuli [49]. Figure 2.10 depicts the neuroanatomy of the visual cortex, including the areas discussed.

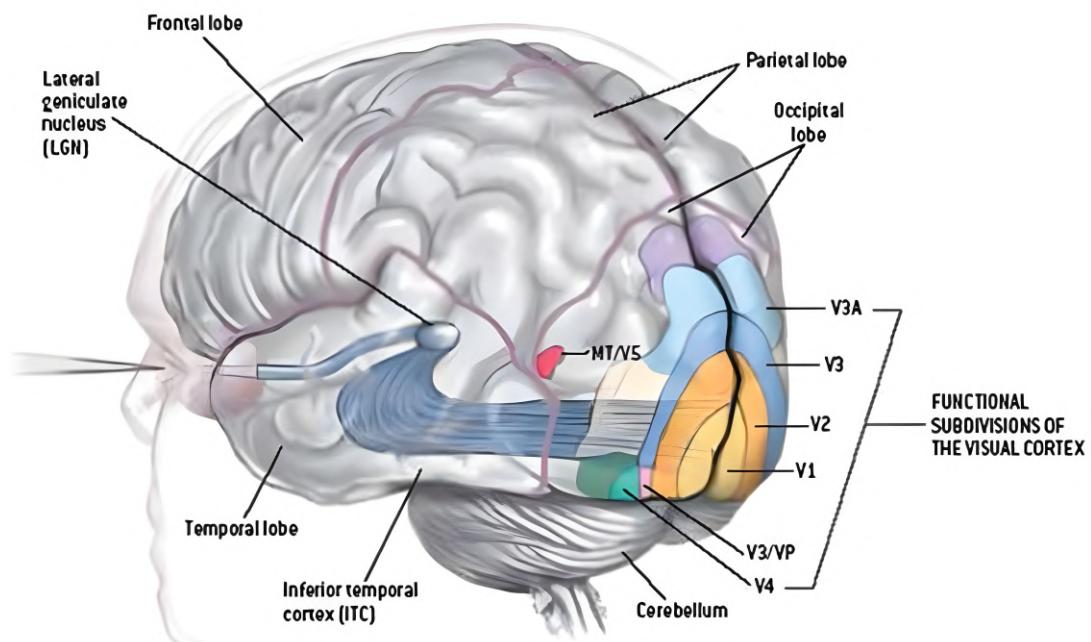


Figure 2.10: Illustration of the main areas of the visual cortex, namely areas V1, V2, V3, V3A, V4, and V5, including subcortical pathways and structures (LNG). Adapted from [49].

2.4.3 The motor and sensorimotor cortex

A wide variety of BCIs target specific locations related to movement planning, execution, and their corresponding sensory processing and perception. These are the motor and somatosensory areas of the brain, located in the *precentral gyrus* (frontal lobe) and *postcentral gyrus* (parietal lobe), respectively. The mentioning of these regions is often accompanied by the homunculus representation (Figure 2.11). The homunculus is commonly depicted as a distorted image of a human, with the size of each body part proportionate to the amount of brain real estate dedicated to controlling that part.

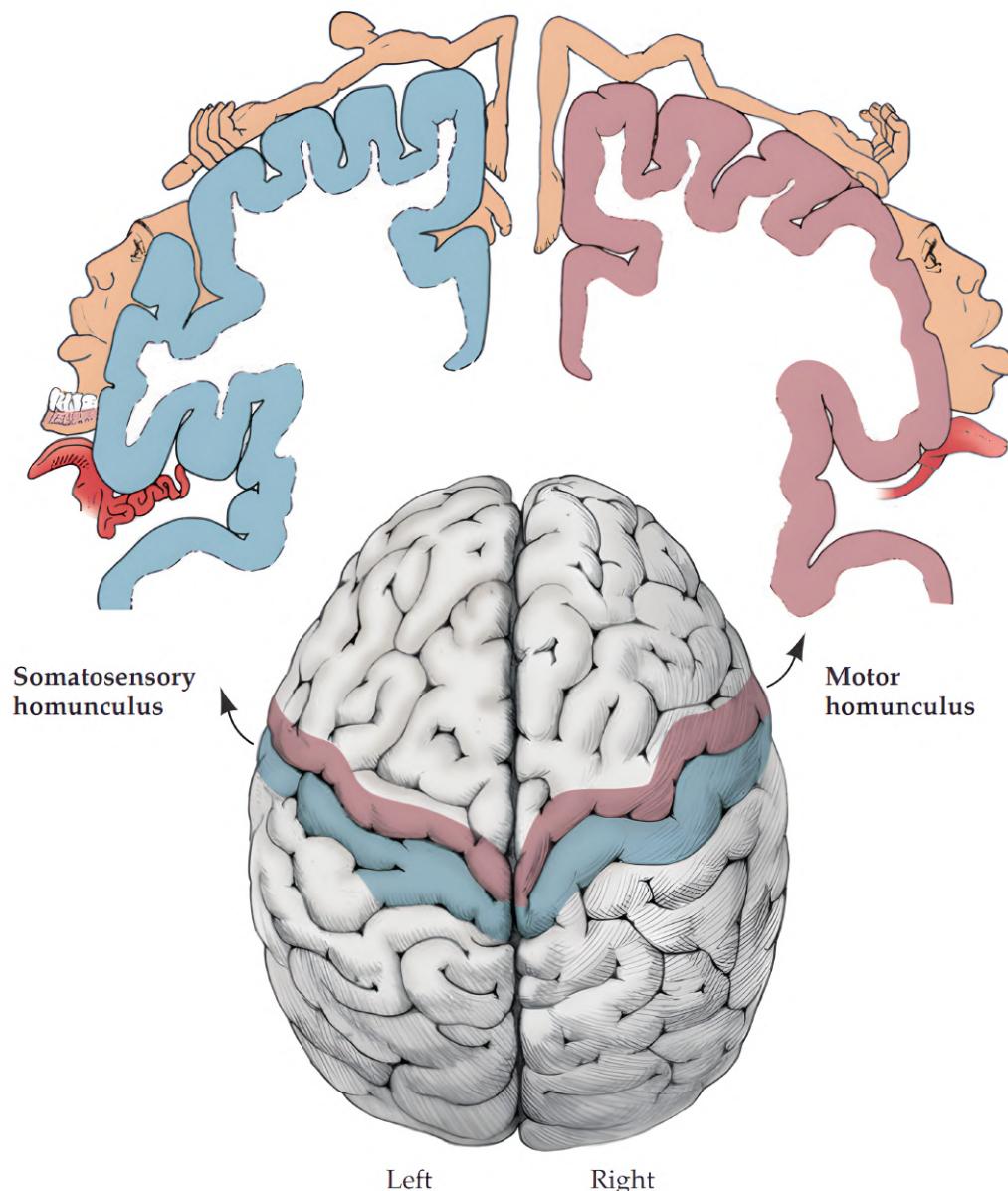


Figure 2.11: Illustration of the somatosensory homunculus in the *postcentral gyrus* (blue) of the left hemisphere and motor homunculus in the *precentral gyrus* (red) of the right hemisphere [47].

The hands, face, and mouth are typically the most enlarged in the motor homunculus, reflecting the importance of fine motor control in these areas. The legs, feet, and torso are typically smaller, reflecting that less brain area is dedicated to controlling those parts. Additionally, one must not forget one of the most intriguing properties of the brain when considering the motor cortex, *i.e.*, the concept of contralaterality. This is related to the fact that the left cortical hemisphere is responsible for controlling the right half of the body, whereas the right hemisphere controls the left side. Such notions are extremely important when trying to discriminate types of intended movements from the EEG signals.

2.5 From the brain to a computer: An overview of the BCI Pipeline

The pipeline of a BCI application describes the path the signal passes through from the moment it is acquired from the brain and digitized, to the execution of a certain task after being translated into commands. Although there are vast amounts of different applications, BCI pipelines normally follow the same set of generalized steps [30]. Firstly, the physiological data is acquired with the use of electrodes placed on top of the skull, spatially distributed to optimize the capturing of the brain function desired. The EEG signal is characterized by having a low Signal-to-Noise ratio (SNR) that can either arise from non-physiological sources, namely power line noise or the incorrect grounding of electrodes, or from the users themselves, mainly due to ocular and head movements [50]. This is particularly relevant for long-term use EEG systems such as BCIs where these movements are expected to occur more frequently. Consequently, the correct interpretation of the neural activity of interest requires first the isolation of the signal from the noise, the so-called artifacts. This stage is named pre-processing.

With the data cleaned and ideally representing solely brain data, the next step aims to transform the signals into objective metrics and extract features related to a specific brain state. This process represents the heart of the pipeline and is highly dependent on the type of application and brain mechanism targeted. Following this, the extracted features can then be fed into a classifier to discriminate several intents into actions executed by the BCI system, *e.g.*, moving a wheelchair or bionic limb. The majority of classification algorithms seen in the literature are generally based on variations of the Support Vector Machines (SVM) or Linear Discriminant Analysis (LDA) as they offer a great trade-off between accuracy and processing speed, ideal for BCI-based applications [30]. The pipeline of processes can be visualized in Figure 2.12.

When building machine learning models that learn to distinguish different classes of brain data, these undergo design considerations that are specific to the BCI field. While the representation in Figure 2.12 portrays the signal path from brain to executable commands, it does not refer to an important and increasingly appealing practice within BCI research, *i.e.*, whether brain metrics resulting from the end of the BCI processing chain are fed back into the brain, ultimately influencing the next obtained results positively by producing more accurate classification and faster model convergence. In other words, whether the system is open or closed-loop. This process was already mentioned in Section 2.1 and is referred to as Neurofeedback (NF). NF is a method

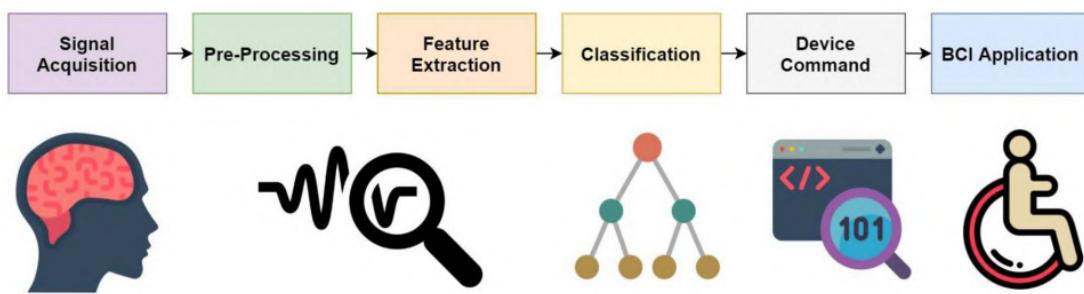


Figure 2.12: Generalized pipeline for decoding EEG information into executable commands through a BCI system [30].

that allows individuals to consciously regulate their own brain activity in real-time by providing them with information about the translated signals and task performance [51]. This process helps individuals become more aware of the variations in their brain activity and through training, adapt and learn to influence neural oscillations in the desired direction. Notably, when comparing users' performance during the same paradigm with and without NF, the condition where NF is present has been shown to significantly improve BCI performance in a wide variety of applications such as in modulating SMR oscillations [52] or regulating levels of focus [17]. NF is typically conducted with visual, auditory, or a combination of both sensory mechanisms [51].

In addition to NF, The learning algorithms introduced above all share a characteristic that is unfortunately common and tricky to overcome in BCI research, *i.e.*, they are subject-specific. This means that a model is trained and tested for each participant by using subject-specific neural data. This can be explained by differences that are intrinsic to the inability to match inter-recording impedances perfectly and contrasting head-cap fittings, but mainly by the term coined as BCI Illiteracy.

The phenomenon of BCI illiteracy is not yet fully understood but is rooted in the fact that some subjects cannot achieve adequate control over BCI systems, even with proper NF [53]. It is thought to arise from subject-specific neuroanatomical and neurophysiological factors with participants revealing widely different optimal oscillatory activity within the SMR rhythm range, for instance [54]. Combining these factors with limitations in EEG acquisition also poses a possible explanation, with activity arising simultaneously in two opposed *gyri* walls being invisible in the EEG due to electric field cancellation [54]. Motivational, psychological, and cognitive factors also play a significant role in BCI performance. High motivation and low anxiety levels predict better BCI control [55], with higher degrees of attentional control also being associated with improved performance [56].

Now that these principles are established, it is possible to extrapolate that designing a pipeline capable of achieving satisfactory results is a non-trivial task, particularly when considering that a significant portion of BCI applications require real-time processing and actuation. However, before thinking about processing the physiological data, evoking and isolating the target neural information on the mesh of activity within the human brain required to achieve the desired result

is equally important. As a result, the next section will dissect the main experimental conditions necessary to elicit distinct slices of neural activity utilized within the BCI scope.

2.6 The paradigms of BCI technology

There are several paradigms that can be used to elicit a specific brain pattern in time, frequency, and space (*i.e.*, the scalp), that can be interpreted and translated into commands. The most widely used are the Event-Related Potentials (ERP), Steady State Evoked Potentials (SSEP), and Motor Imagery (MI) [34]. Despite being less used, the Error-Related Potential (ErrP) and Slow Cortical Potentials (SCP) are also worth mentioning.

2.6.1 Event-Related Potentials

Due to the high temporal precision required, EEG activity time-locked to stimuli designated Event-Related Potentials (ERPs), sometimes interchangeably termed as Evoked Potentials (EP), are commonly used to capture neural activity related to sensory processes. ERPs are nothing but the averaged neural activity upon the repeated presentation of the same stimulus [57]. Because of this, ERPs have lower amplitudes ($\approx 10 \mu\text{V}$) when compared with the ongoing brain waves captured with the EEG ($\approx 50\text{-}100 \mu\text{V}$) [33].

While studying the brain's response to stimuli, participants might elicit spontaneous and involuntary neural activity during any moment of the recording. Simply put, the neural response to stimuli is highly sensitive to the subject's attention, presence of motor acts, and inner thoughts, among others, introducing random segments of activity in the signal that might even overshadow the targeted response. Conversely, the neural activity evoked by that particular stimulus will always be present after the instant of its presentation. Consequently, by averaging activity across trials of EEG activity recorded in response to the same stimulus, the spontaneous and variable activity will be filtered out, whereas the signal phase-locked to stimuli onset will become evident [58].

There are several types of ERPs that arise from a wide variety of sensory inputs, namely visual, auditory, somatosensory (tactile), olfactory [58], gustative [59] or even from internal senses such as gastric interoception [60], with the former two modalities being the most commonly studied. In addition to their modality, ERPs can be classified depending on the pathway it is activating. Early components peaking within 100 ms after stimuli presentation are considered to be in the period of stimulus transduction in the sensory organs and conduction of neural signals along the sensory pathways, thus being called sensory or exogenous ERPs [61]. After the 100 ms mark, these trigger and manifest the cognitive or endogenous ERPs. Exogenous ERPs mainly depend on the physical parameters of the stimulus while the endogenous category represents the manner in which the subject evaluates the stimulus as they examine information processing [61]. Independently of the category, the individual ERPs nomenclature is defined according to the period they occur in relation to the respective stimulus and their positive or negative concavity [58].

Despite being termed here as a BCI paradigm, in actuality, ERPs are in a different class of experimental design than the other two paradigms discussed in this chapter due to their loose conditional parameters. Rather than being an experimental paradigm by definition, ERPs are an umbrella of paradigms where each one is designed to elicit specific neural processing of different types of stimuli.

In a *face recognition paradigm*, images comprising human faces are shown, thus firing the mechanisms related to human face processing and eliciting the N170 and Vertex Positive Potential (VPP) components at the occipitotemporal (P7/PO8 - PO8/PO7) and vertex (Cz) channels, respectively, peaking around 170 ms after stimuli onset [62].

Additionally, the *conditioning-testing paradigm* involves the presentation of a pair of auditory stimuli, S1 (condition) and S2 (testing), which is used to study sensory gating, *i.e.*, the ability of the brain to suppress activity when processing consecutive redundant/equal stimuli presented with short intervals (*e.g.*, 500 ms) [63]. This generates a component designated as P50 with each stimulus and is measured at electrode Cz around 50 ms after its onset, with the amplitude ratio between these two components defining the amount of suppression of the redundant stimulus, which can be relevant to the study of certain diseases like schizophrenia [63].

However vast, in the context of BCI applications, the number of different options usually tapers down to the *oddball paradigm* in which the participant is presented with a series of stimuli and must respond to rare or infrequent "target" stimuli while ignoring the more frequent "standard" ones [58]. Targets usually comprise around 5-10% of the total stimuli to generate the desired "surprise" result [33]. The target neural response elicits a whole body of interesting ERP components, one of which is the already mentioned P3/P300 peak that kick-started the BCI field, holding its name due to appearing positively 300ms after stimuli onset [30]. This potential is considered to reflect the information-processing cascade associated with attentional and memory mechanisms and can be seen widely distributed over the scalp electrodes but achieves its maximum for the visual modality in the centro-parietal channels (Pz) [30]. The P3 can be further divided into the P3a and P3b subcomponents with different mechanism correlates and scalp distributions [64]. While the P3a indexes the orienting of attention to significant or unexpected events in the environment, being more focused frontally, the parietally distributed P3b reflects the up-dating of working memory which is recruited while consciously attending to the target stimuli [64]. Putting it differently regarding the P3b, the subject is required to hold a representation of the target in memory to properly recognize and consciously attend to it upon its emergence. Unless targeted separately with an experimental design, both subcomponents contribute to the P3 peak and even overlap, being hard to distinguish. However, the P3b is the one that contributes the most to the target processing.

The P3 displays an interesting option for BCI systems as it consistently provides high discriminability, it is easy to compute and implement, requiring only a few electrodes, and because it is generated automatically in response to a stimulus, training is not necessary, being only an option in cases of complex paradigms or to improve classification accuracy [34]. Furthermore, albeit being a commonly brought out disadvantage when discussing P300 experiments, the habituation issue, *i.e.*, the reduction in efficacy due to the brain gaining "resistance" to the stimulus, studies

have shown that both the latency and amplitude remain stable over a continuous usage period of 40 weeks in users severely disabled by ALS, not affecting BCI performance [65]. Additionally, in spite of being shown to be generated regardless of whether the user is focusing directly on the target or not [13], the P3 discriminative power significantly decreases the further away from the target stimulus the user's attention is, inferring to its gaze-dependent modulation [66].

Preceding the P3 negatively at around 200 ms, the N2/N200 potential is another relevant component that co-occurs and typically co-varies with the positive peak during the *oddball* experimental design. This means that, similarly to the P3 component, an increase in amplitude (more negative in the N2 case) is expected in response to target stimuli in comparison to its standard counterpart. However, this is only true for one of its three subcomponents, the N2c, which is related to the orienting of visual attention during stimulus target classification, having the biggest contribution over the centroparietal scalp [64]. This is the component that accompanies the P3b. Contrastingly, the N2b subcomponent is more focused on the frontocentral region and encompasses response inhibition, response conflict, and error monitoring on the non-target stimuli [64] or even when detecting changes in stimuli properties that are task-relevant [61]. In the *oddball paradigm*, the inhibitory response represents the conscious decision to not attend to non-target stimuli, thus triggering increased negativity to non-target images than to targets [64]. This is the subcomponent that is seen in combination with the P3a down the processing cascade. The visual N2a, more recently relabeled to visual Mismatch Negativity (MMN), is also observed frontocentrally but reflects the identification of novel or mismatching elements compared to the perceptual expectations during attended stimuli [64]. This potential is elicited even when the participant is not paying attention to the stimuli, unlike with the N2b and N2c.

Furthermore, other potentials related to the N2 peak have been researched, providing interesting additional properties when designing BCI systems. The N2pc, where the "p" and "c" are an acronym for posterior and contralateral, respectively, is a lateralized component related to visual selective attention, appearing in the cortex contralaterally to where the object of interest appears in the visual field [19]. In other words, if the object appears on the left visual field relative to fixation, the component will have a stronger contribution on the right side of the hemisphere and vice-versa, enabling the exploitation of additional information in addition to the target vs non-target image classification, namely in visual search [64, 19]. Spatially, the N2pc reaches its maximum amplitude in the parieto-occipital electrode sites P7/P8 and PO7/PO8 and it does not require attention focused on the stimulus, providing a gaze-independent BCI alternative [13]. Notably, spatial attention underlying the N2pc generation can be allocated separately and independently to more than one target when presented sequentially [67]. This was demonstrated by Eimer and Grubert (2014) when a pair of target visual stimuli were shown in quick succession separated by 10-100ms, each generating its own N2pc with latency differences corresponding to the stimulus onset asynchrony (SOA). The difference in latency between the two N2pc components generated matched the two targets' presentation difference, even at 10 ms. Analogously to the horizontal target specificity of the N2pc, a recently labeled component, the N2pcb where the "b" stands for bilateral, has been shown to be related to the visual search of targets on the vertical midline [68].

It is important to note that some of these paradigms can be fused together into one. Such a case is that of an *oddball face recognition paradigm* where in a pool of several face images, one of them is considered to be the facial target stimuli, thus generating the N170/VPP and later in the processing cascade, the N2 and P3 [69]. Figure 2.13 demonstrates typical ERP responses in a *oddball face recognition paradigm*.

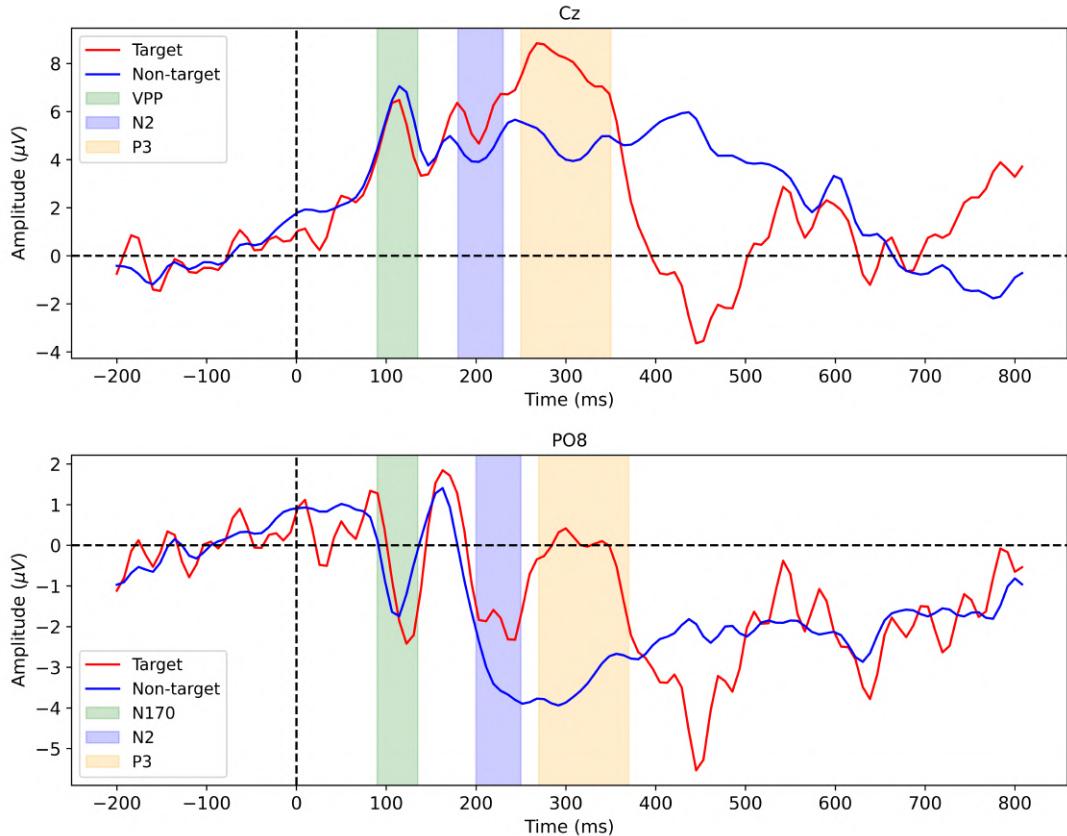


Figure 2.13: The target (red) and non-target (blue) visual ERP response during an *oddball face recognition paradigm* recorded with 16 electrodes on my scalp. 100 target responses were averaged to obtain the target-related ERP while 1000 were used for the non-target response. The recorded EEG signals were referenced to both earlobes and posteriorly band-pass filtered between 0.1 and 30Hz. The visual stimuli were composed of several faces of recognizable celebrities with the face of Daniel Radcliffe being attended to, representing the target response occurring with 10% of probability. The image depicts the ERPs mentioned at channels Cz (above) and PO8 (below) revealing the opposing components associated with human face recognition VPP and N170, respectively, around the same period (green patch). Furthermore, the components N2 (blue patch) and P3 (light orange patch) can be seen around 200 ms and 300 ms respectively, with a clear difference in amplitudes between the target and non-target responses. Although not being highlighted, the accentuated negativity following the P300 around 450 ms represents the N400, also showing significant differences in amplitude between both responses. An important consideration to have when visualizing ERPs and comparing them to the literature is that researchers sometimes invert the y-axis making the negative peaks appear pointed upwards and the positive peaks pointed downwards. Despite being relatively common, it is not a standard and as a result, will not be done throughout the rest of the document.

A category of ERP paradigms that show potential for BCI purposes is dependent on the rate of stimuli presentation designated Rapid Serial Visual Presentation, which provides exciting and novel ways to access neural information.

2.6.1.1 Rapid Serial Visual Presentation

Rapid Serial Visual Presentation (RSVP) 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 under the *oddball paradigm* [33]. 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 brain built-in visual mechanisms to triage images in an accurate manner under a few hundred milliseconds by extracting relevant information from brain circuitry related to the identification and detection of an object of interest [21].

The most impressive feature of these types of BCI is that they can discern these changes in brain waves and perform decisions at the borderline of human consciousness, meaning that they can do it faster than the user is able to consciously perform that decision. Evidence has shown that it is possible to capture useful information with presentation rates up to 10 images per second [33]. Thus, these BCIs offer an important field of research that may be able to help people perform tasks more efficiently such as the analysis of large image datasets in reduced time [33, 21]. In fact, RSVP-driven 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 [33]. However, this design has the disadvantage of requiring sustained attention which can lead to lapses in vigilance due to visual fatigue.

2.6.1.2 Motion-related Visual Evoked Potentials

Albeit RSVP paradigms provide time-locked neural information in a very efficient manner, visual processing activity can also be exploited on dynamic and much richer visual environments, *i.e.*, live scenes (*e.g.*, video-games) with moving stimuli rather than static images [70]. In fact, object motion within a video or live scenery generates time-locked Visual Evoked Potentials (VEP) that can be picked by BCI systems to perform decisions. Among the motion-based potentials, the motion onset Visual Evoked Potentials (mVEP) reliably produce the largest amplitudes and lowest intra and inter-subject variability [71]. However, other types of responses related to moving stimuli exist, namely, the motion offset VEP representing the evoked activity upon the cessation of a moving object, motion-reversal VEP produced by changing motion direction, Steady-State motion Visual Evoked Potentials (SSmVEP) that occur during continuously moving stimuli, among others [71].

Object motion typically elicits three components. These are the P1 (110-130 ms), N2 (200 ms), and P2 (240 ms), with the latter being replaced by the later P3 wave during a motion discrimination task [71]. Depending on several factors underlying stimuli presentation, the response

can have a dominant contribution of the P1 or the N2, corresponding to two variants of the mVEP. These factors include the relative position of the moving object in relation to the center of the eye, its contrast, temporal frequency (speed of motion), and inter-stimulus interval [71]. The P1 prevalent response underlies the presentation of macular stimuli, referring to the macula, the specialized area at the center of the retina, near the back of the eye responsible for high-acuity vision, color perception, and detailed visual tasks. Thus, in macular stimuli, these coincide with the field of vision perceived at the center of the eye, emanating a dominant P1 on the *striate cortex* (channel Oz) [71]. Additionally, high temporal frequency (> 6 Hz), high contrast, and short inter-stimulus intervals (< 1 s) also result in the dominant P1. Contrastingly, the N2 peak, which in fact has been shown to represent the motion onset processing system activity, is induced by extramacular stimuli that emerge in the peripheral vision where movement is better perceived, lower contrast and temporal frequency stimuli motion, long inter-stimulus intervals, and around 200 ms of motion duration [71]. Moreover, originates from from *extrastriate cortex* and has been linked to *magnocellular* pathway activity. Compared to ERPs generated by static stimulation, the N2 provides the most distinct peak in response to dynamic stimuli rather than the P3 [70].

2.6.2 Steady State Evoked Potentials

Similarly to ERPs, SSEPs are brain patterns generated through the presentation of stimuli. However, while ERPs consist of the brain response to discrete stimulation, typically shown at rates around 1 or 2 Hz (on non-RSVP designs) that are time-locked to stimuli onset, SSEPs are elicited through continuous stimulation at higher frequencies [72]. In the case of the visual modality that gives rise to Steady State Visual Evoked Potentials (SSVEP), in which most research regarding this paradigm for BCI applications is held [34], frequencies generally range from 6 Hz to 30 Hz, comprising the flickering of lights or photos [30].

This paradigm's most important property is that the stimuli frequency is mimicked in the brain. In other words, while the user maintains the gaze on the intended stimuli, if they are exhibited with a frequency of 15 Hz, the EEG activity enters into a *steady state* and produces SSVEPs matching the fundamental frequency of the flicking stimulus, including its harmonics [30]. When analysing the frequency spectrum, a clear distinction between the stimulation frequency and the other oscillatory activity can be seen manifested over the primary visual cortex region, as showcased in Figure 2.14 [73]. As a result, by presenting stimuli with distinguishable frequency bands, each one of these frequencies can be associated with a certain control option. Therefore, much of the feature vectors extracted from this method comprise spectral metrics from the target frequency and its harmonics.

The SSVEP experiments share and in fact, improve some of the advantages stated about the P300-BCIs, insofar as not requiring any form of training from the user and being considered the most reliable and fast form of communication in the category of EEG BCI, evaluated by metrics specific related to BCI performance defined in 2.7 [34]. However, because of the nature of the stimulation, SSVEPs require a certain level of focus and can cause fatigue. Furthermore, SSVEP-based BCIs provide better results with overt attention, *i.e.*, physically directing the gaze to

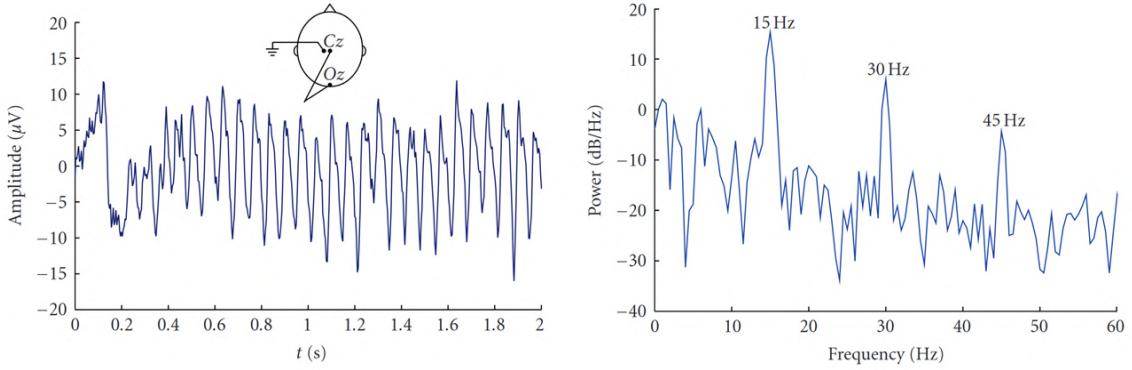


Figure 2.14: EEG signal (Oz-Cz) obtained during a visual light stimulation at 15 Hz. The left image exhibits the time domain SSVEP resulting from the average of the time-locked signals over 10 trials, showcasing a clear transient VEP at the start of the stimulation ($t = 0.0\text{s}$) with the subsequent oscillations related to the SSVEP (starting around 0.3s). The right images constitute the frequency spectrum of the signal, demonstrating the pronounced power of the stimulation frequency (15 Hz) and its higher harmonics [73].

a stimulus, rather than with cover attention, described by the act of mentally focusing on stimuli through vision periphery, thus not enabling reliable gaze-independent information [13].

2.6.3 Motor Imagery

As already described, the act of moving a limb produces oscillations in motor areas of the brain known as sensorimotor rhythms (SMR). Notably, without even tensing the muscles, imagining movements also produces these waveforms, hence the paradigms name Motor Imagery (MI) [30].

This paradigm is generally employed to evoke the activity corresponding to the limb areas of the cortex (Figure 2.11). The areas related to hand and foot motor activity are large enough for EEG-based BCI to discriminate both intentions. In fact, motor activity of the right and left hands can be captured by C4 and C3 electrode positions, respectively [30]. On the contrary, due to the cortical regions being extremely close and deeper in the brain (Figure 2.11), motor activity of the left and right feet is often impossible to distinguish with non-invasive methods, being both captured at the Cz channel location.

During the imagined or execution of motor acts, changes in the *Mu* band and *Beta* frequencies occur. These are designated Event-Related Desynchronization (ERD) and Event-Related Synchronization (ERS) [34]. ERD refers to the drop in frequency band amplitude in the brain's sensorimotor areas while ERS refers to the rise in band amplitude after the movement or imagined movement. The *Mu* band ERD begins just before the movement starts, hits its maximum shortly after, and then returns to baseline level. On the other hand, the beta band experiences a brief ERD at the start of the movement and then switches to ERS, reaching its maximum after the movement is completed, while the *Mu* rhythm is already at normal levels [34]. An illustration of this behavior is shown in Figure 2.15.

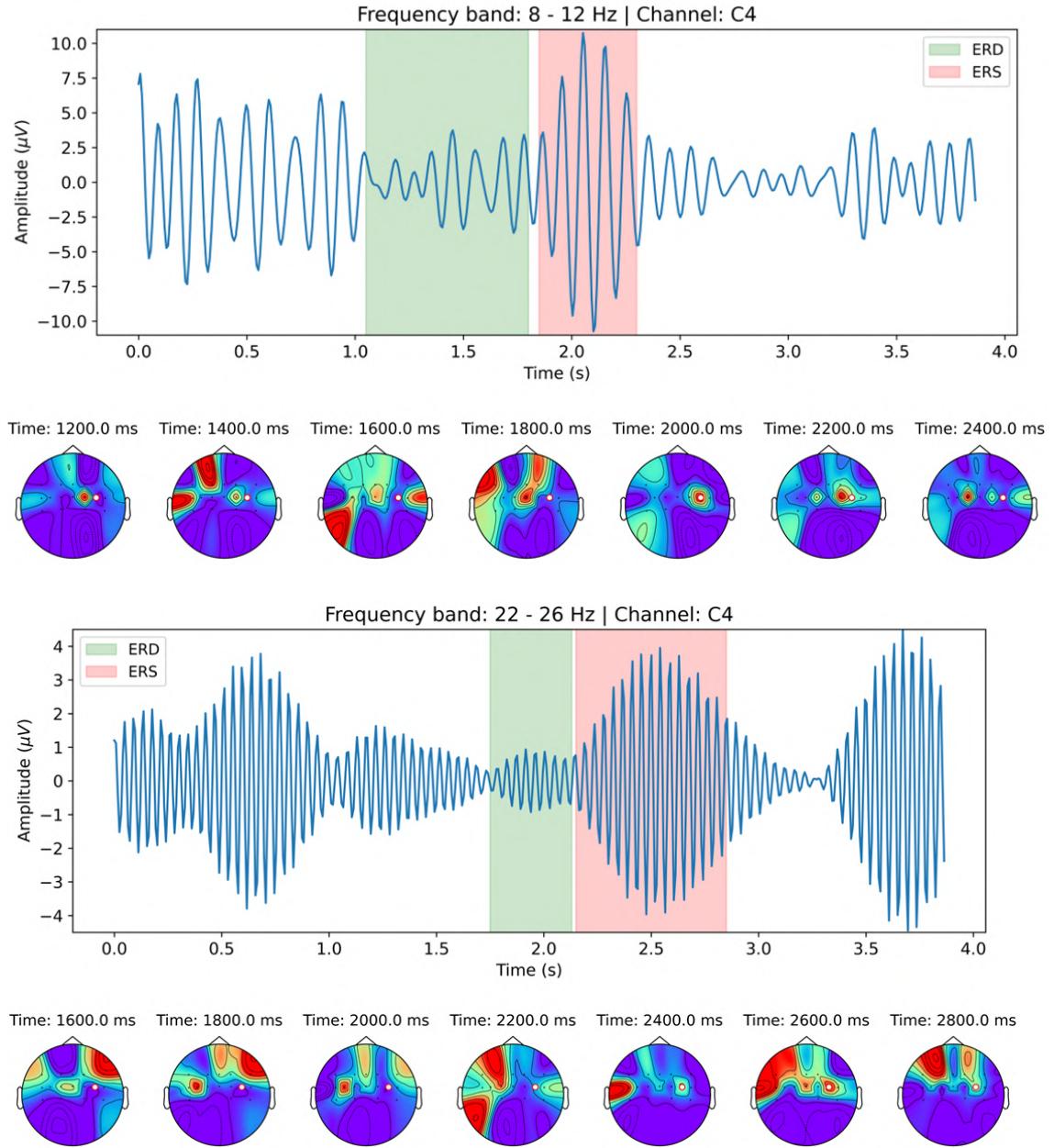


Figure 2.15: Illustration of the ERD and ERS that accompanies left wrist movement generation on my motor cortex captured during a motor imagery paradigm. The instant $t = 0\text{s}$ depicts the instruction to imagine the movement. The first half of the image depicts the ERD and subsequent ERS of the *mu* rhythm in the time domain at channel C4, contralateral to the moved limb, and the changes in Power Spectral Density (PSD) across the 16 electrodes recorded in a topographic map of the scalp (1200 ms - 2400 ms). The second half of the image depicts the same trial recorded but with information related to the upper *Beta* frequency. Electrode C4 is highlighted with a white dot in both topographical maps. Note that in both maps, an increase in frequency power (ERS) focused on electrode C4 can be seen around the time the movement is executed and a depression (ERD) around the same area that precedes the execution for both frequencies. There is also an apparent delay between the ERD and ERS of the *Beta* rhythm in relation to the *mu* wave, corroborating with the literature. Before plotting, the signals, in addition to being filtered in the corresponding frequency ranges, were processed using the Common Spatial Pattern (CSP) algorithm to improve the source detection of activity in each electrode, thus revealing the activity of the left wrist generation over the right motor cortex in channel C4. Furthermore, each instant depicted in the topographic map corresponds to the average PSD activity 50 ms before and after that instant.

MI-based BCIs are great as they provide a lot of control options without the use of external cues but their correct operation is conditioned by prior extensive training as the user needs to optimize the imagination of each limb [34]. Additionally, these design, despite enabling gaze-independent control, suffers from substantial differences between users and a high proportion of BCI illiteracy related to the inability of some users to generate discriminatory activity, impossibil- itating the use of an MI-based BCI [74, 13].

2.7 BCI performance metric

Evaluating the performance of these systems is essential to compare results and improve methodologies. A variety of methods to evaluate BCI performance have emerged throughout the years and despite the lack of a standard method, the Information Transfer Rate (ITR) is the most widely used metric to compare communication performance in BCI systems in the literature. The equation described by Wolpaw in 1998 [75] goes as follows:

$$B = \log_2(N) + P \times \log_2(P) + (1 - P) \times \log_2\left(\frac{1 - P}{N - 1}\right) \quad (2.1)$$

where B represents the Bit rate in bits/symbol, N is the number of classes, and P the classifier accuracy. The symbol can represent a specific event, state, or decision that is encoded as a binary value, such as a stimulus, response, or action. However, in the literature, instead of bits/symbol, ITR is represented in bits/min (Bt):

$$B_t = B \times (60/T) \quad (2.2)$$

with T (seconds/symbol) representing the time needed to output each command/symbol.

2.8 Chapter 2 summary

This chapter aimed to introduce the Brain-Computer Interface as an emerging field and discuss its main properties, types, and applications in today's world, starting with its evolution throughout the years. Because BCIs require capturing neural activity, the main recording techniques used to decode brain patterns were presented and compared, motivating the versatility and scalability of the EEG to construct these applications. Because of this, the EEG as a brain recording technique was thoroughly presented and explained according to the system of electrode placement and its biosignal properties.

With the brain being the signal generator targeted, a review of its main anatomical regions and sources of activity exploited in the context of BCIs was also provided, namely the visual and motor cortices, from which the main BCI paradigms are carefully designed to isolate these neural sources. These include Motor Imagery, Event-Related Potentials, and Steady State Visual Evoked Potentials, which were all considered in the early stages of the project proposal. The typical signal

conditioning pipeline required to design an operational BCI system was also reviewed according to the main considerations involved in its design and the main performance metric (ITR).

From this analysis, specific Event-Related Potentials were identified to provide a way to produce a BCI that fits the project requirements for search and rescue scenarios, namely the N2pc, N2pcb, and P3 potentials.

Chapter 3

EEG-based Rapid-Serial Visual Presentation (RSVP) and target localization pipelines: A literature review

The present Chapter aims to evaluate the current status of BCI technology with a focus on target identification RSVP paradigms with or without spatial localization. The type of hardware, relevant electrodes, discriminatory brain activity, and methodologies used will be discussed.

The search was conducted on SCOPUS database with the combinations of the keywords and their abbreviations: *Brain-Computer Interfaces*, *Rapid Serial Visual Presentation*, and *Electroencephalogram*. In total, the search obtained 223 results. With respect to target identification paradigms, the search was more difficult since keywords like *Target Identification* or *Target Localization* did not yield many results that were worth picking. From this, a selection was made based on the relevance to the current study and the most relevant papers ($n = 13$) that met this criteria are described below according to their pipeline and main conclusions.

3.1 RSVP-based BCIs

In terms of their applications, RSVP-based BCIs are mainly categorized into spellers, medical image analysis, face recognition, data categorization, and surveillance, with the latter being the most researched [33] and where this project falls under.

In 2017, the feasibility of the N2pc potential to identify not only the lateral location of a target of interest within aerial images but also its centroid estimation was explored by Matran-Fernandez and Riccardo Poli in 11 healthy participants [76]. The P3 in conjunction with the N2pc potential was also explored to discriminate between target and non-target stimuli through an image presentation rate of 5 Hz in a pool of 2400 images, 10 % of which were from the target class. By combining different sets of electrodes on the parieto-occipital region of the scalp, the

authors were able to achieve an average AUC of 0.873 for target discrimination with 28 channels. Furthermore, the N2pc achieved a 0.76 median AUC in classifying left vs right visual angle target appearance. Furthermore, the N2pc elicitation was also found to be significantly different between right-handedness (RH) and left-handedness (LH) subjects. When attempting to create a regressor based on Neural Networks with a hidden layer composed of 8 neurons, the actual x-coordinate and the predicted point achieved a correlation of 0.42 for the LH group and 0.39 for the RH participants. The EEG signals were acquired with a 64-electrode BioSemi ActiveTwo system at a 2048 sampling rate, referenced to the mean value of both earlobes. The signals were band-pass filtered between 0.1 Hz and 28 Hz and the channel data comprising the feature vectors were downsampled to 64Hz before being classified using the SVM architecture. Later, the same authors approached the same paradigm and participants by combining the BCI output from two individuals with identical performance to the same image [19]. This mitigated the difficult discrimination of the N2pc due to its small size and improved the classification performance by 5% on average. Additionally, in the later study, the same Neural Network architecture used in the former study achieved an average correlation coefficient between the predicted and real x-axis centroid of 0.19 at 5 Hz across 11 participants.

Paul Sajda and colleagues (2010) have developed two EEG-based BCI architectures synergistically coupled with computer vision to aid image sensing and image retrieval routines [21]. They termed these systems as *cortically-coupled computer vision* due to the integration of both systems. The BCI is based on the P3 potential to discriminate between target images and non-target images and uses spatial filters to map the neural response associated with the images presented to an interest score. Spatial filters are data processing techniques that utilize statistical data to maximize the variance of neural data between two classes by using the spatial distribution of electrodes. They are very common in MI BCIs as these target specific activity under the motor cortex. Recently, these have also been used to help ERP analysis. Here, the authors used the Hierarchical Discriminant Component Analysis (HDMA) that accounts for the spatial distribution of electrodes across several time windows after stimuli onset, also retaining the temporal information of the ERP [21]. A variation of the LDA classifier, the Fisher Linear Discriminant algorithm was used for classification within each time window.

The first application consists of helping users review high-dimensional images (several hundreds of giga-pixels in size) more efficiently. To achieve this, a computer-vision algorithm was used to cut the images into several smaller pictures with potential targets of interest. These smaller pictures were then presented at 10Hz to a user and based on the interest score generated according to the target defined by the researchers, fed the cut pictures prioritized by the computed score into a database for later validation. By comparing the results with highly trained specialists that conduct these types of searches, the researchers found that their system improved by 300% the time required to review and find the targets within the aerial images [21]. To record the EEG, they used the 64-electrode system ActiveTwo by Biosemi with a 2048 Hz sampling rate. To train the model to tune the P3 potential with 20 to 30 blocks of 100 each where two target images are used.

The authors then employed the architecture for image retrieval application, combining EEG-based object of interest (OOI) detection with a graph-based visual pattern mining algorithm [21]. Initially, a small subset of images is randomly sampled from a larger database and presented to the subject for EEG measurement. The EEG interest detector identifies a set of highly ranked images as 'pseudo positive' labels for the pattern discovery module. This module refines and propagates these labels throughout the entire image database, generating a larger set of related images based on the user's interest. By minimizing subject participation, the system effectively infers objects attracting attention, allowing users to access a substantial number of interesting images without lengthy EEG sessions. This approach balances user engagement and efficient image retrieval. In comparison to using EEG alone (33.71%) and for chance (1.76%), the system drastically improves the image retrieval task, achieving 69.1% Average Precision (AP) for one subject searching for Dalmatians [21]. In contrast to the first example, in this particular design, the authors used a 6Hz presentation rate with 1000 images (10 blocks of 100) from 62 different categories.

In [16], an RSVP paradigm was coupled with a UAV for real-time target selection. The visual paradigm consisted of the presentation of aerial high-dimension pictures taken by a drone, cut into several non-overlapping subpictures that were presented at 10Hz to 12 volunteers. A total of 80 trials divided into 4 blocks, with each trial consisting of 150 pictures (15s), was used. A variation of the HDCA algorithm was used, the slide-HDCA, which uses a sliding window method to improve classification. Additionally, the xDawn model was used to remove the 10Hz SSEVP associated with the high presentation rate of visual stimuli. The algorithm achieved an 86.6% target detection rate. The EEG acquisition system used was the 64-channel NeusenW series wireless from NEURACLE with a 1000-Hz sampling rate.

In order to improve gaze-independent BCIs, Yijing Wang and colleagues (2022) devised a new spatial coding paradigm combining both the P3 and N2pc for target recognition and spatial localization [74]. A two-level classification algorithm was implemented in which the HDCA was first used to perform the target vs non-target discrimination, while the Discriminant Canonical Pattern Matching algorithm paid more attention to the spatial distinction. The presented stimuli consisted of 2 side by side images where the target could appear on one of them and participants were asked to concomitantly identify the presence target and its spatial localization (left or right). The stimuli presentation rate was 5Hz and 5 blocks of 8 trials each (100 images with 8 targets) were used on 14 subjects. A single-stimulus array of images was presented with only one image instead of two, comprising the same parameters to compare the performance. With an average accuracy of 84.43% across all subjects with 6 subjects scoring above 85%, the algorithm that incorporated the N2pc spatial localization level on classification improved by 10% the target detection efficiency with an ITR of 60.52 bits/min against 54.96% [74]. The 64 channels Synamps2 system from Neuroscan (1000 Hz) was used to record the EEG, which was downsampled to 250 Hz and filtered between 1 and 30 Hz before the decoding process.

3.2 Non-RSVP spatial localization-based algorithms

Similar to other studies already discussed, in order to improve the P3 gaze-dependent BCI system of target selection to navigate an avatar toward a destination mark, researchers have utilized a paradigm based on covert attention to elicit the N2pc potential on 24 subjects [13]. The avatar was controlled by a four-class paradigm in which each movement direction was associated with a specific color, *i.e.*, blue/left, red/up, yellow/right, and green/down. These were displayed during the whole experiment in the margin of the screen while the user focused only on the avatar. The correct move was hinted at by presenting the target color as a ring instead of a full circle and NF was provided by the movement associated with the decoded color in addition to emotional feedback provided by the avatar depending on if the cued direction was in agreement with the decoding (smiley face), the direction did not agree with the hinted color but the distance toward the destination did not increase (neutral face), and if the distance to the destination increased, a sad face was presented. To discriminate between four classes under the constraint of only two conditions, the participants were presented with two potential targets on each side, necessitating various stimulus combinations to determine the participant’s focus. As a result, each trial consisted of 12 stimuli onsets and the algorithm’s aim is to find the right sequence of left/right responses that correlate with the brain data. The algorithm used was the HDCA to decode the EEG signal, acquired with the BrainAmp DC amplifier with 29 channels, after band-pass filtering between 1 Hz and 12.5 Hz and downsampling to 50 Hz [13]. The algorithm achieved great results with 6 subjects exceeding the 80% classification accuracy.

Similarly, a study investigated the viability of employing a multiple correlated components analysis (MCORCA) method for identifying lateral targets in visual search tasks within a single trial employing the N2pc for a BCI system [77]. The approach involves weighting target signals using spatial filters to enhance the amplitude difference between bilateral hemispheres. By extracting linear combinations of multiple channels across trials with an optimal subset of correlated components, the method aims to prevent the loss of efficient information. The proposed method achieved a classification rate of 82% with a single short-duration trial when assessed using Leave-one-out-cross-validation (LOOCV) [77]. To record the signal, the 64-channel Neuroscan SynAmps2 EEG was used, and before decoding the data, the signal was re-referenced to the infinity reference, was downsampled to 250 Hz, and band-pass filtered between 0.1 and 30 Hz. The visual stimuli were arranged randomly in the four quadrants around a central cross, forming a square layout. Each trial displayed a random number of items (4, 8, or 12), evenly distributed around the fixation point. Visual search displays presented either an O among Cs or all Cs, with the circle as the target and the arcs as distractors, indicating a serial search. All 30 subjects performed 20 blocks of 5 trials and each began with a central cross flashing for 500 ms, following a stimulus presentation that would stop after 1500 ms or if the participant responded by right-clicking a mouse button for the right-placed target or a left press for a left target [77].

Because the N2pc does not support the visual search of targets alongside the vertical midline, Mattia Doror (2020) and colleagues investigated and labeled a potential that is triggered during

the same period as the N2pc but with a vertical instead of horizontal modulation, inferring that they correspond to the same mechanism of visual search. They labeled this potential as N2pcb [68]. This enabled a distinction between targets above and below fixation in the vertical midline. To evoke such EEG potentials, they presented to 12 participants a series of dots equidistant around the center point. They divided the experiment into a singleton search, where participants were fed with right and left target positions hinted by a colored dot among grey dots, and a feature search where all dots were colored and the target color was placed above or below fixation. The participants were instructed to press the keyboard button 1 or 2 to attend to both the presence and absence of a target, respectively. Each subject altered between both designs to amount to 10 blocks of 96 trials where the visual stimuli were present for 100 ms and 1300 ms were given to attend them. In contrast to the N2pc, which is computed by subtracting the contralateral to ipsilateral components relative to where the object appeared on the visual field, the N2pcb relies on collapsing the bilateral components (contralateral and ipsilateral) elicited by targets on the vertical midline and subtract from it the ipsilateral activity related to lateral targets [68]. They showed clear amplitude differences between the bottom and top presented targets in which the bottom ones evoked much greater negativity in comparison. Furthermore, the potential during the singleton search emerged slightly earlier when compared to the feature search. This can be explained by the added visual complexity of the latter design. To record the EEG, the authors used 27 scalp channels with EOG for latter artifact removal after low-pass filtering at 40 Hz and baseline correcting them between 0 and -100 ms relative to stimuli onset.

In [22], the authors investigated new ways to improve the target localization within images attending to the limitations of the N2pc. Backward Decoding Model (BDM) and Forward Encoding Model (FEM) with a Channel Tuning Function (CTF) to raw EEG signals to extract and reconstruct feature-based target selection processes in space and in time with targets presented at 4 different quadrants. The algorithm tuned to the temporal window between 260 ms - 270 ms, matching the N2pc elicitation period, demonstrated the highest discriminatory accuracy. Importantly, this model can spatially discriminate individual target positions in these displays and is fully invertible, enabling the researchers to construct hypothetical topographic activation maps for target positions that were never used. When tested against the real pattern of neural activity obtained from a different group of subjects, the constructed maps from the forward model turned out statistically indistinguishable, thus providing independent validation of their model. The EEG was recorded with BrainAmp DC amplifier (500 Hz) from 27 electrodes, the signal was band-pass filtered between 0.1 Hz and 40 Hz, baseline corrected between -200 ms to -100 ms, and downsampled to 250 Hz. The stimuli presentation array featured two successive 20 ms stimulus displays, separated by a 100 ms interval (80 ms blank screen). Each display contained two items in different colors (red, green, blue, or yellow), one being the target color. Participants attended to the target color item either in the first or second display in different blocks. Stimuli included uppercase letters and digits presented at an eccentricity of 2.4° from fixation. Participants reported the category of the target color item by pressing a corresponding key. The experiment comprised 12 blocks, with 64 trials each, resulting in 768 total experimental trials. The second experiment

featured similar parameters but displayed 8 possible positions in which one was the target but this time it never appeared on the vertical or horizontal midline.

In a study where the objective was to control a drone using the P300 paradigm, the differences between the navigation performance in a *Unity* simulation (VR) and with a real drone from the real-time video (AR) was evaluated through the same HMD device (TC-Vive) [15]. A training session was included before the navigation tests in both environments. The EEG signal was acquired from seven electrodes referenced to the left ear (Fz, Pz, Oz, P3, P4, PO7, and PO8) using a DSI-VR300 headset. The signal processing pipeline included a band-pass filter between 0.1 and 30Hz, following the epoching of the first 1000ms after the target stimulus onset with subsequent averaging and baseline correction before applying LDA to classify the signals based on the components' amplitude and latency. The users controlled the direction of the drone by focusing on one of seven different control options that blink in random order on the screen. By synchronizing the time in which the *oddball* stimulus in each control option appears with the elicitation of the P300, the application is able to predict the direction of interest. This study found that there were no significant differences between the accuracies in each environment, being 89.71% overall. The authors noticed, however, a correlation between the user's preferred environment with an increase in accuracy as participants who enjoyed more the VR experience obtained an accuracy of 93.1% in VR and 89.33% in AR. Conversely, individuals that preferred the AR experience achieved 90.76% in that environment and 85.46% in VR, corroborating with previous studies on the importance of user experience and the overall engagement factor in the performance of BCI systems [15].

3.3 EEG classification algorithms and optimal feature selection

As already briefly mentioned in section 3 and shown in the present section, many BCI pipelines are based on variations of the SVM and LDA learning algorithms. This is because these typically perform well in high-dimensional spaces where the number of features is high relative to the number of samples, a characteristic commonly occurring in EEG data. Additionally, the inner workings of these machine learning algorithms are also robust to overfitting and are computationally efficient in real-time scenarios. However, the evolution of the standard computational resources available opens doors to the usage of more powerful classifiers with minimum loss in decision time that have been applied to EEG data but are not very commonly used in BCI systems. These encompass the Random Forest [78, 79], the XGBoost [80, 81], and the AdaBoost [82, 83]. Furthermore, more powerful learning algorithms, the so called Artificial Neural Networks, have been developed and combined with EEG data for more specific problems [19].

However, not all features extracted may optimally describe the differences between classes. In these cases, a feature selection algorithm is often applied prior to classification, namely Mutual-Based Information. A review of the Mutual-Based Information feature selection algorithm followed by a description of the 5 classifier architectures mentioned is given below.

3.3.1 Mutual-Based Information (MBI)

Mutual-Based Information is not completely categorized as a machine learning technique but composes an important step in their pipelines, working as a feature reduction method, ensuring only the most relevant features for a particular classification problem remain. The MBI algorithm is a powerful feature selection method that leverages mutual information to identify and retain the most relevant variables for predictive modeling. Mutual information quantifies the amount of information obtained about one random variable through another random variable, effectively capturing the dependency between features and the target variable [84]. The MBI algorithm systematically evaluates these dependencies to rank and select features that contribute the most to the predictive accuracy of the model. This process helps in reducing dimensionality, mitigating overfitting, and enhancing model interpretability by focusing on the most informative variables [84].

When applying MBI to categorical data, the algorithm calculates mutual information by examining the frequency distributions and co-occurrences of different category combinations [84]. This approach captures the nonlinear relationships and dependencies between categorical features and the target variable. For numeric data, mutual information is computed by discretizing the continuous variables into bins, transforming them into categorical counterparts [84]. The algorithm then similarly evaluates the dependencies as with categorical data.

3.3.2 Support Vector Machines (SVM)

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 [85].

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 [85]. 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 [85].

3.3.3 Linear Discriminant Analysis (LDA)

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 [86]. 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 [86].

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 eigen decomposition of the matrix product of the inverse of the within-class scatter matrix and the between-class scatter matrix is then performed [86]. 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 [86].

3.3.4 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 [86]. 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 [86]. 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 [86]. 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.

3.3.5 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 [86].

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 [86]. 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 [86].

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 [86].

3.3.6 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. [86]

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 [86]. 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 [86].

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 [86]. 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 [86].

3.3.7 Artificial Neural Networks

The models described so far all accomplish good recognition of the interdependencies of vast amounts of data and their corresponding labels. However, some problems require more complex architectures in order to correlate, learn, and output an unknown function describing the data. In this department, Artificial Neural Networks (ANNs) offer a way to achieve this.

ANNs are computational models designed to mimic the way biological neural networks, such as those found in the human brain, process information, operating through several key components

and processes that enable them to learn from data and make predictions. These include activation functions, weights, biases, forward propagation, loss functions, backpropagation, epochs, batches, and optimization algorithms [87].

The basic skeleton of these architectures is the neurons and their connections, associating a weight between two neurons that determine the strength and importance of the inputs to each neuron. They are crucial for the network's ability to learn and adjust based on the data it processes. Biases are additional parameters added to the input of each neuron before applying the activation function. Biases help the network fit the data better by providing flexibility in the decision boundary [87].

Activation functions are mathematical functions applied to the input of each neuron to determine its output. They introduce the required non-linearity into the network, allowing it to learn complex patterns. Common activation functions include the sigmoid function, which maps inputs to a value between 0 and 1, the ReLU (Rectified Linear Unit) function, which outputs the input directly if it is positive and zero otherwise, and the tanh function, which maps inputs to a value between -1 and 1 [87].

The forward propagation is the process where input data passes through the network layer by layer to generate an output [87]. Each neuron's output is calculated by applying the activation function to the weighted sum of its inputs plus the bias. This process continues until the final layer produces the network's prediction.

The loss function measures the difference between the network's predicted output and the actual target values [87]. It quantifies the error of the network, guiding the learning process. Common loss functions include mean squared error for regression tasks and cross-entropy loss for classification tasks.

Backpropagation is perhaps the most important process within the ANN architecture and is actually what introduces the learning portion of the algorithm depending on the loss accumulated through its layers. It involves computing the gradient of the loss function with respect to each weight and bias in the network [87]. This gradient indicates how much each parameter should be adjusted to reduce the loss. The network then updates the weights and biases in the direction that decreases the loss, typically using an optimization algorithm.

Epochs refer to the number of times the entire dataset is passed through the network during training. Each epoch involves forward propagation, loss calculation, and backpropagation, iteratively improving the network's performance. Batches are subsets of the dataset used during training to update the network's parameters. Training on batches, rather than the entire dataset at once, makes the training process more efficient and helps the network generalize better to unseen data [87].

Optimization algorithms are techniques used to update the network's weights and biases during training [87]. They aim to minimize the loss function effectively. Gradient descent is a widely used optimization algorithm, where the parameters are updated proportionally to the negative gradient of the loss function. Variants like stochastic gradient descent (SGD) use individual batches to

compute the gradients, while more advanced algorithms like Adam combine the advantages of both adaptive learning rates and momentum to improve convergence speed and performance [87].

An example of an ANN architecture is given in Figure 3.1, also illustrating the data flow associated with both the forward and backpropagation within the network.

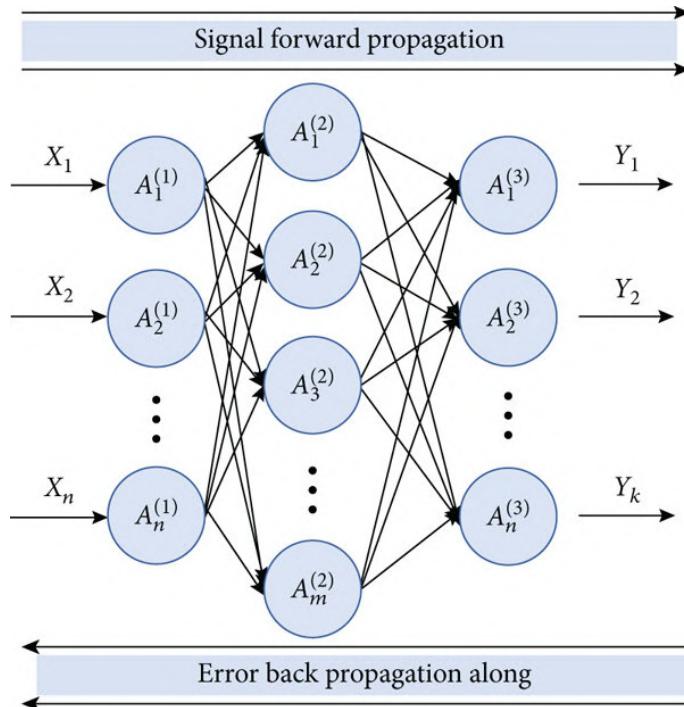


Figure 3.1: Diagram representing an ANN architecture [87].

3.4 Model evaluation

In order to assess what machine learning architectures provide the best performance on unseen data, a Cross-Validation loop is often used to compare several performance metrics associated with classification pipelines. A review on the Cross-Validation loop and how it influences the classification pipelines is given below. Finally, a method to tune the hyper-parameters that best separate the feature vectors of a particular application is also presented.

3.4.1 Cross-Validation (CV)

Cross-validation (CV) is a robust statistical method used to evaluate and improve the performance of machine learning models by partitioning the data into subsets. This technique primarily aims to mitigate overfitting and provide a more accurate assessment of a model's predictive capabilities on unseen data. In its most common form, k-fold CV, the dataset is divided into k equally sized folds [86]. The model is trained on $k-1$ folds and tested on the remaining folds. This process is repeated k times, with each fold being used exactly once as the test set. The results from these

iterations are then averaged to produce a single estimation of the model's performance. This cyclical validation process ensures that every data point has an opportunity to be in both the training and testing sets, providing a comprehensive evaluation of the model's performance across different data subsets [86]. This intuition is contextualized in Figure 3.2.

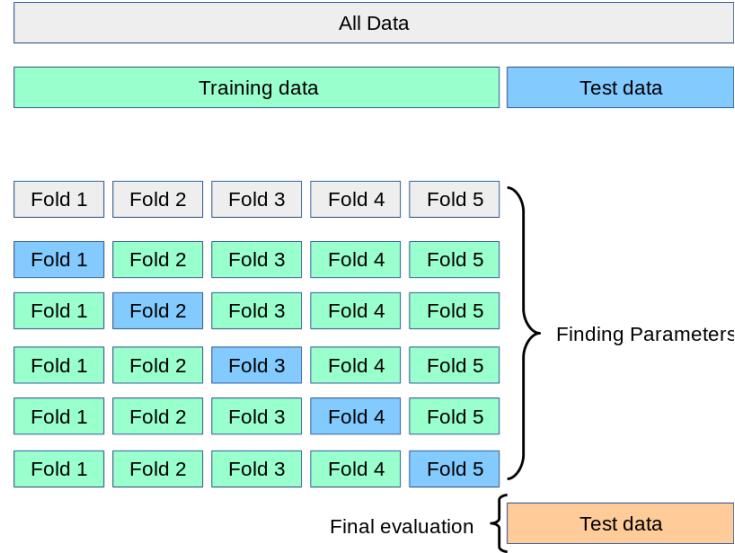


Figure 3.2: An example of the inner workings of Cross-Validation with 5-folds [88].

By systematically rotating through the dataset, CV offers several advantages over traditional single train-test splits. Firstly, it maximizes the utilization of available data, particularly beneficial when the dataset is small. Secondly, it provides a more reliable measure of a model's generalizing ability by capturing variance and mitigating bias that could arise from an arbitrary split of the data. This method is particularly useful for hyper-parameter tuning, model selection, and assessing model stability. Through its iterative and thorough approach, CV fosters the development of more robust, accurate, and generalizable machine learning models, thereby enhancing their real-world applicability and reliability.

3.4.2 Classifiers Grid Search

Grid search is an approach that aims to optimize the hyper-parameters within each model architecture. Hyper-parameters represent properties that are set before the learning process begins and control the behavior of the training algorithm. Unlike model parameters, which are learned during the training phase, hyper-parameters need to be specified before training and can greatly influence the performance and outcome of the model. Finding the optimal combination of hyper-parameters can significantly enhance the model's performance, ensuring that the model generalizes well to unseen data.

Grid Search CV (Cross-Validation) automates this optimization process by exhaustively searching through a predefined grid of hyper-parameter values to determine the best combination based on cross-validation performance metrics. During this process, the algorithm places all possible

combinations of hyper-parameters on a grid and iteratively evaluates each combination's performance.

This exhaustive search ensures that all potential hyper-parameter combinations are considered, providing a comprehensive understanding of how each parameter is tuned specifically for a particular feature vector. While grid search can be computationally intensive, especially with a large number of hyper-parameters and values, it is a robust method for hyper-parameter tuning, particularly when coupled with cross-validation to prevent overfitting and ensure model robustness.

3.4.3 Model evaluation metrics

When designing and comparing different learning algorithms through CV, it is pivotal to understand the different metrics commonly used to evaluate their generalizing ability to unseen data, with each metric providing a different insight into the model's performance. Some metrics are more indicative than others on the model's true performance depending on the type of application desired, class context (*e.g.*, class imbalances), and if discriminatory outputs are given the same weight. For instance, in sensitive applications like medical diagnosis, different types of classification errors can have contrasting consequences, and thus, different weights can be associated with them.

In the pool of evaluation metrics for machine learning purposes, accuracy is one of the most used and intuitive. It measures the capacity of the model to correctly predict the assigned classes by accounting for the proportion of 4 different terms commonly displayed on the confusion matrix [86], as defined in Table 3.1

- **True Positives (TP):** Classes correctly predicted as positive
- **True Negatives (TN):** Classes correctly predicted as negative
- **False Positives (FP):** Classes incorrectly predicted as positive
- **False Negatives (FN):** Classes incorrectly predicted as negative

Table 3.1: The confusion matrix concept for a binary problem discriminating between Target and Non-Target classes.

	Predict Target (Positive class)	Predict Non-Target (Negative Class)
Target (Positive class)	TP	FN
Non-Target (Negative class)	FP	TN

From this, the accuracy metric is obtained:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.1)$$

While accuracy provides a straightforward measure of model performance, it can be misleading, especially in imbalanced datasets where one class dominates. In such cases, a high accuracy might simply reflect the model's ability to predict the majority class correctly, rather than its overall discriminative power. The AUC metric, on the other hand, provides a better estimate in these cases, being derived from the ROC (Receiver Operating Characteristic) curve, which plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings [86]. The TPR, also referred to as sensitivity/recall, and the FPR are obtained as follows [86]:

$$TPR \text{ (Sensitivity/Recall)} = \frac{TP}{TP+FN} \quad || \quad FPR = \frac{FP}{FP+TN} \quad (3.2)$$

The sensitivity (TPR) describes the proportion of correctly classified samples of the positive class (1) while the FPR measures the percentage of negative class samples that are incorrectly classified as positive. The FPR is also obtained through the subtraction of $1 - specificity$, with the specificity describing the proportion of correctly classified samples of the negative class (0). The ROC curve illustrates the trade-off between TPR and FPR across all possible thresholds, effectively summarizing the model's performance irrespective of a specific threshold. The AUC represents the area under this curve and quantifies the overall ability of the model to discriminate between positive and negative classes. It essentially summarizes the model's performance across all possible thresholds, providing a comprehensive evaluation that reflects the model's ability to rank positive instances higher than negative ones [86]. This makes the AUC a more robust and informative metric, especially in the context of imbalanced datasets, as it encapsulates the performance of all possible confusion matrices and offers an aggregated view of the probability that the model will assign a randomly chosen positive instance a higher predicted probability compared to a randomly chosen negative instance.

Despite no guidelines have been established, an AUC above 0.7 is considered an acceptable classification performance, with a higher threshold of 0.8 AUC being optimal for BCI applications [76].

3.5 Chapter 3 summary

This Chapter focused on reviewing the pipeline architecture basis in relevant published research related to high-throughput (RSVP) and target identification with subsequent spatial localization BCIs by decoding the visual processing performed by the brain associated with image presentation. The signal acquisition, processing, and classification pipelines mainly directed at exploiting the N2pc, N2pcb, and P3 potentials are discussed and contextualized with the particular application goals of each study.

Based on the learning algorithm architectures identified within these studies, the most relevant classifiers for EEG data, *i.e.*, SVM, LDA, Random Forest, XGBoost, and AdaBoost were presented and detailed according to their inner workings. Furthermore, Deep Learning algorithms were also identified in some of these studies, and because of this, the Artificial Neural Network

architecture was also explored. Additionally, common methods to evaluate the model's robustness and performance are described, namely Cross-Validation and Classifier Grid search to optimize both model and hyperparameter selection for each discrimination task along with several metrics that enable quantification of the model's ability to discriminate different classes of data. These refer to the accuracy and confusion matrices, from which the AUC, Sensibility, and Specificity metrics are derived. Finally, a commonly utilized algorithm used for optimal feature selection is also presented (Mutual Based Information).

Chapter 4

Materials and Methods

The current chapter is dedicated to thoroughly explain how the study was conducted, including all the steps required to design the paradigm that was ultimately used to acquire the EEG signal, described in section 4.1. Furthermore, the signal conditioning pipeline with subsequent BCI actuation is described in section 4.2. Furthermore, in section 4.3, a way to compare the BCI results here produced is discussed.

4.1 Signal acquisition

The acquisition of biosignals inherits several factors that must be considered and meticulously thought out to guarantee that the physiological phenomena of interest are correctly captured. This section will describe how the stimuli were handled, how the paradigm was engineered to elicit the desired neural activity, and the steps taken to ensure that the EEG data was correctly mapped to each mental state being studied.

4.1.1 Designing the RSVP *Oddball paradigm*

Given the emphasis on target identification, the RSVP *oddball paradigm* has been selected as the most appropriate experimental design for this project. This is because it effectively isolates and salients the target stimulus among a rapidly occurring stream of non-targets, allowing for efficient measurement of the participant's ability for target detection through ERP analysis, as discussed posteriorly.

The paradigm was developed using the *PsychoPy* software [89]. *PsychoPy* is an open-source software package for designing and running experiments in psychology, neuroscience, and related fields. It is written in the *Python* programming language and provides a user-friendly graphical interface for creating, running, and analyzing experiments. PsychoPy supports a wide range of experimental paradigms, including visual and auditory stimuli presentation, response collection, and data analysis. It is transversal to the main operating systems, *i.e.*, Windows, macOS, and Linux, and experiments can be extended using custom *Python* scripts.

The correct execution of any EEG experiment is predicated on the stimuli' quality and appropriate selection. To form the image database, the hallmark that would define the *oddball* stimuli has to be established first. With the BCI application of visual search in rescue missions context in mind, the target response was decided to stem from the presence of a human lying down in the environment. Contrastingly, the absence of this characteristic in the image was delineated as the non-target response. It is important to note, however, that the presence of other human beings in the environment, in concomitance with the defined target or otherwise, is expected and does not dictate the *oddball* if they are not lying down. Such cases form the distractor stimuli. This factor is rooted in the unpredictability of realistic settings, incorporating the necessary challenge to BCIs that attempt to aid real-life situations.

On top of these challenges, for this project to have practical relevance, the identification of an object of interest must be at least environment-independent in the sense that the target identification should be generalized regardless of the background in which it is distributed. Exploring this idea even further, a degree of generalization would exponentially expand the applications of this paradigm. Optimally, this would be achieved by the brain in the loop by leveraging its generalization mechanisms, not requiring compensation from the algorithm counterpart by scaling the models' complexity and decision time. To test this putative cognitive flexibility at high speeds, two slightly different RSVP *Oddball paradigms* were devised. These were labeled as the **Training** and **Testing** protocols, also conveniently aiding the separation of the data used to train and test the models.

Hereby, with the experiment conceptualized, the following subsections will describe the manner this was achieved in practical terms.

4.1.1.1 Capturing stimuli

This subsection will delineate how the raw visual stimuli were captured. To achieve this, the drone model DJI Mini 4 Pro was used, supporting blur-free high fidelity and resolution image capture at high speeds. This particular model and its main specifications are displayed in Figure 4.1.



Figure 4.1: The used DJI Mini 4 Pro drone and its main specifications.

A routine to fly the drone while taking aerial pictures was planned and implemented by converting a chosen path drawn in a map to a `.json` file that can be read by the device. The drone was flying at approximately 25 meters relative to the ground. The altitude is saved on the image properties but can be misrepresented due to the absolute altitude factoring in the height of the geographical location relative to the sea level.

To obtain real-scenario images that would target the brain generalization problem, location scouting was first carried out to obtain images comprising different environmental characteristics. As a result, the **Training** dataset was taken in Madalena Beach at Vila Nova de Gaia, whereas the **Testing** dataset was captured at FEUP. A full description of the image dataset is given below.

4.1.1.2 Image dataset description

After location scouting, the only property of the images left to define would be the human subject that participants would attend to during EEG acquisitions. It was decided that the easiest way was to use the author lying down in the environments captured. However, to test the generalization property of the brain regarding differences in target characteristics, the clothes the author is wearing are significantly different between both image collections. Despite representing the same target, changing its clothes and colors would suffice to test the generalization of different humans since the clothes represent the biggest noticeable region of the target within the images.

Following this specification and touching on the environmental differences captured, the **Training** image dataset contained more complexity and diversity regarding both the settings captured and the objects contained. In total, 65 Ultra HD (4k) 4032 x 2268 px (`.jpg`) were taken containing a wide mix of beach and urban settings. In total, 6 different environment categories can be distinguished, namely a solely sand-based environment, a mix of sand, rocks, and sea, a mix of sand and wooden buildings, a dune setting comprising sand and light vegetation, a full urban setting depicting cars, streets, and sidewalks, and another urban setting covering mostly bicycle paths. All these environments contain images with and without the attended human target, which appears with a yellow shirt. The 6 types of environments described are shown in Figure 4.2.

Similarly, the **Testing** image dataset was captured in the same format and resolution for a total of 61 images. The environments covered were less diverse but completely different from the ones presented in the **Training** dataset, making it possible to test the environment-independent BCI. The images captured a grassy environment with some buildings on the edges and the human target appeared with a red sweater. Two examples of these images are represented in Figure 4.3

Despite the existence of some vegetation on some particular images of the **Training** dataset, the overall color pallet and characteristics of the environments are completely different between the visual stimuli from both datasets. Furthermore, the targets also display completely different characteristics, making it hard to perceive that it is the same person in both datasets, especially when considering the velocity of stimuli presentation.

Despite the intended differences, a careful matching of the cognitive load between both image collections, *i.e.*, the attending task difficulty, had to be ensured. Cognitive load has been known to impact the properties of the ERPs [90] and a good match will establish similarities between

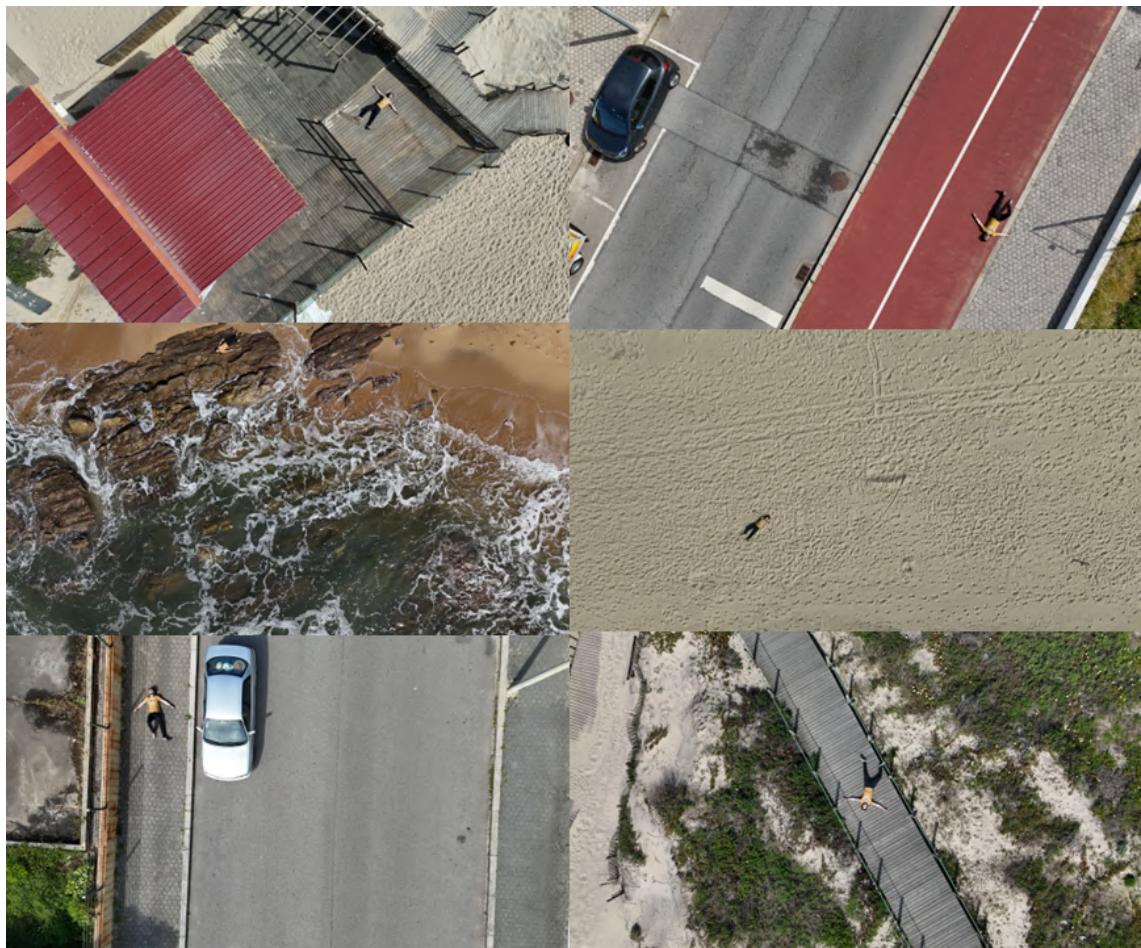


Figure 4.2: Depiction of the 6 types of environments described that are contained in the **Training** image dataset. In each of these, a human target can be seen wearing a yellow t-shirt.

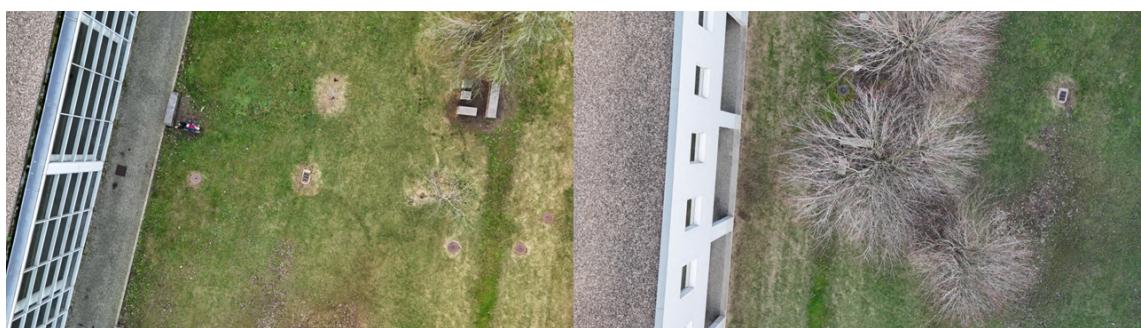


Figure 4.3: Depiction of the type of images contained in the **Testing** image dataset. On the left, an image containing a human target wearing a red sweater, and on the right, an image without a human target.

the EEG used to train and test the data, and consequently, output good results on unseen data. However, matching the cognitive load between these two image collections cannot be done objectively, at least not prior to EEG recordings. As a result, in order to try to approximate these, a few image characteristics were considered and planned before the image capturing. On one hand, the **Training** dataset contains more diverse settings, objects, and color palettes with the t-shirt color being similar to many of the backgrounds, namely the sand and some rocks. However, the target appears slightly bigger on average in the **Training** image pool than in the **Testing** counterpart to balance the added difficulty. In contrast, the **Testing** dataset displays a simpler and more constant environment with a higher contrast target relative to the backgrounds, but also contains more human and object distractors, with the latter being an object with a similar red tone to the sweater. This manipulation thus tried to attenuate the differences in cognitive load associated with the differences required between both stimuli pools to test the BCI independence of background environments and targets.

Once taken, the raw images had to suffer some adjustments to better suit the experiment's design.

4.1.1.3 Preparing stimuli

Before presenting the visual stimuli to the subjects, these were initially processed for a few reasons. The exact same pipeline was followed for both the **Training** and **Testing** image collections but examples will contain mostly images from the **Testing** dataset.

Firstly, the raw images were cropped due to their original dimensions, as referenced in Figure 4.4. Visual stimuli presented in *oddball* designs are typically square-shaped, this being especially relevant for target localization purposes insofar as having the maximal vertical and horizontal field of vision angles containing the image identical. Secondly and most importantly, this approach can be extrapolated to real-time BCI scenarios, as demonstrated by Zhang *et. al.* [16], by distributing spatial complexity through time where instead of presenting a big and complex image, the image is subdivided into smaller and simpler ones that can be displayed in the RSVP format. Thus, the entire environment captured by the original image is subdivided spatially but the same environment is covered in time. Due to the reduction of spatial complexity, the cognitive load is reduced enabling the leveraging of the brain's ability to direct its attention to a simpler space and detect objects that grab its attention in a very fast manner.

Following this rationale, the original Ultra HD (4k) 4032 x 2268 px images captured by the drone were decomposed into 6 sub-images by dividing the original images horizontally by 3 and vertically by 2. This subdivision was chosen to fit as much as possible the square-like shape requirements while zooming the whole image enough without losing quality. Therefore, this process resulted in a total of 378 images with 1344 px in width and 1134 px in height. Posteriorly, some of the generated sub-pictures containing mainly non-relevant visual search information (*e.g.*, building walls or ceilings) were manually deleted, with a total of 313 cropped images remaining in the **Training** image collection and 261 in the **Testing** dataset, respectively. An example of a single

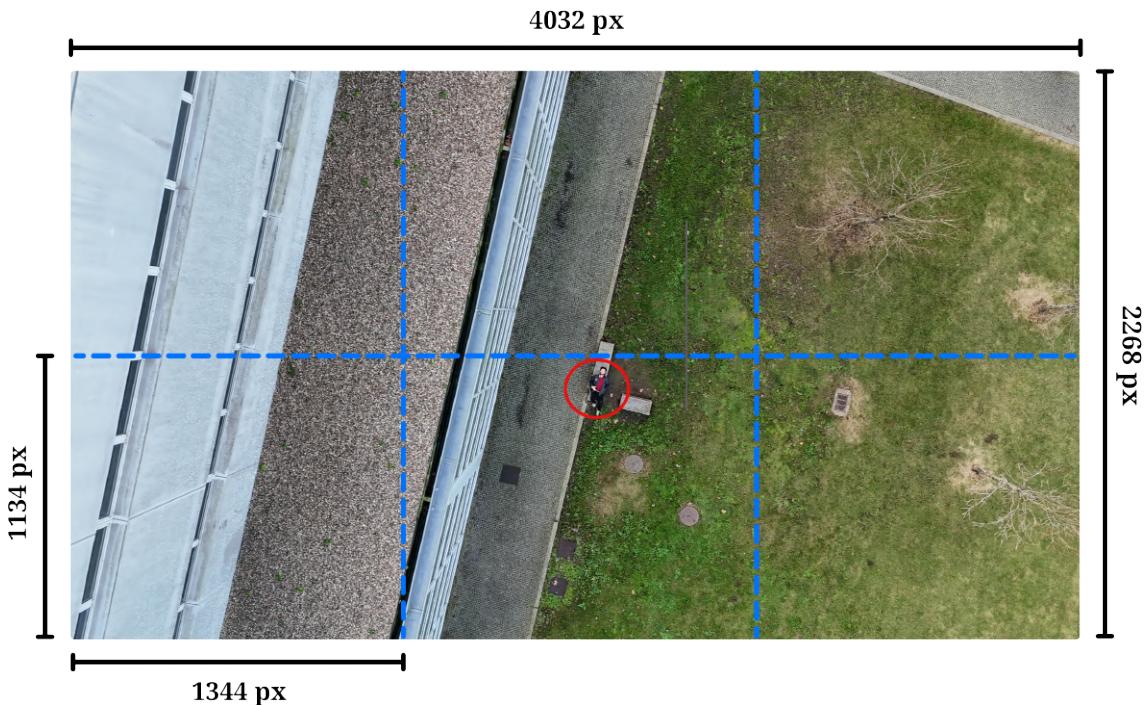


Figure 4.4: Dimensions of the aerial images taken for the **Testing** dataset by a drone at FEUP and the subsequent cropping (3×2) into smaller images, indicated by the blue dashed lines. The target stimuli are identified by the presence of a human lying down contained in the sub-image, as pinpointed by the red circle, whereas in the non-target class, by its absence.

target and non-target image generated from Figure 4.4 and integrated into the **Testing** protocol can be seen in Figure 4.5.

Once appropriately cropped, one important piece of information was left to extract from the sub-images. In order to train a model that could respond to image presentation and estimate where in the image an object of interest is located, data comprising the real centroids (center of mass) along the horizontal and vertical dimensions of these in the target-containing images is required. When capturing the image datasets, a property within the targets was considered before taking the pictures to ease and automate the target centroid labeling.

To automatically find the centroids used for labelling the single-trial target events, a simple color mask using the *OpenCV* package was assembled by leveraging a discriminatory characteristic on the target images, namely the red sweater, as schematized in Figure 4.6 (A) for the **Testing** dataset. Although other red objects exist in the captured images, the red sweater has a particular red tone compared to others. From this, a part containing only my red sweater was manually selected and segmented from a random target image. This image part was afterward converted from the BGR (Blue, Green, Red) format, as it is represented in *OpenCV*, to the HSV (Hue, Saturation, Value) color space. Representing the images in such a format provides a more intuitive way to represent colors, making it easier to isolate a specific color range because the whole color information is portrayed in the first channel, Hue. Contrastingly, the Saturation and Value channels represent the contrast and brightness, respectively. Doing this facilitates the construction of a color



Figure 4.5: An example of the target (left) and non-target (right) stimuli displayed in the designed *oddball* experiment. Note that the red circle surrounding the object of interest in the target image is only used to highlight its location and is not included during the stimuli presentation.

mask that can segment similar colored regions present in other images. To assemble the color mask, a color range was specified based on the average and standard deviation red color values of the segmented part. For the Hue channel, a lower and upper color range was defined using 5 standard deviations from the mean, while 4 standard deviations were used for the Saturation and Value channels. These were the values used for the **Testing** dataset while a standard deviation of 2 was used for both the Hue and Value channels within the **Testing** dataset. These values were established through trial and error for the specific needs of the image datasets formed.

Once established, the HSV color mask was then used to isolate the regions within the yellow range of the t-shirt for the **Training** protocol and the red range of the sweater for the **Testing** one, as depicted in Figure 4.6 (B) for the latter. This process is initialized by looping through the image database comprising target and non-target images where in each iteration the cropped pictures are rescaled to half their size, resulting in a 567 x 672 px image. This was done to equalize the maximum image dimension (672 px) to the image size used in the literature [19]. Following the appropriate rescaling, the image is converted to the HSV color space and the mask is applied. This binarizes the portions of the image contained in the HSV range, setting the pixel values in the said range to 1 (white), whereas color values that do not match this definition are set to 0 (black). From the isolated region, the centroid pixel values along the vertical and horizontal dimensions are extracted by calculating the 2D geometric moments, according to the following equation [91]:

$$M_{i,j} = \sum_{x=1}^W \sum_{y=1}^H x^i y^j I(x,y), \quad (4.1)$$

where $M_{i,j}$ denotes the moment of order for the x-coordinate (i) and y-coordinate (j), x and y the pixel coordinates along the horizontal and vertical directions until the maximum width (W) and height (H) are reached, respectively, and $I(x,y)$ the pixel value at coordinate (x,y) , which in a binarized image can be either 1 or 0. When dividing the average horizontal distance from the

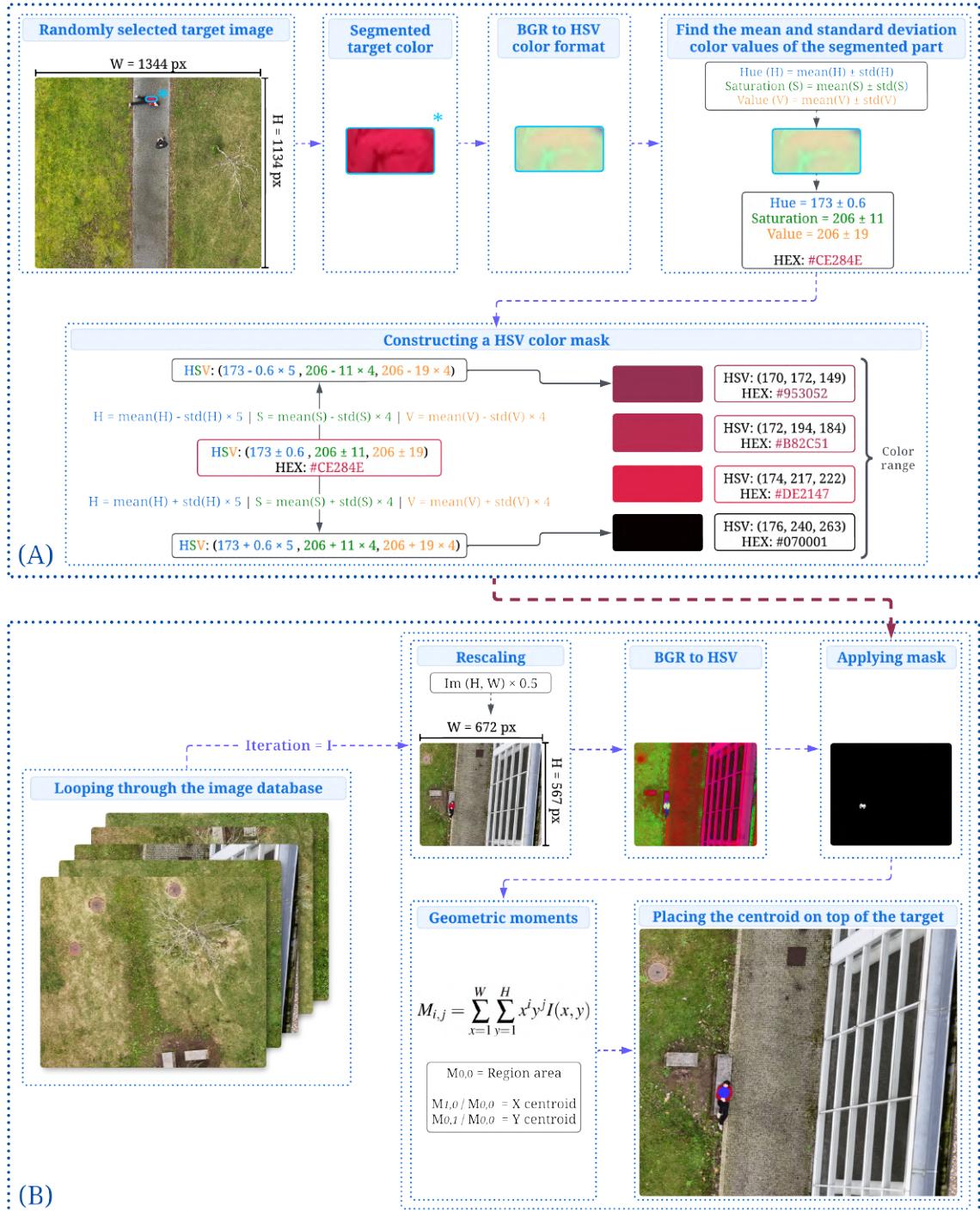


Figure 4.6: Flowchart of the algorithm applied to extract the centroid of my body in the **Testing** target images along the horizontal and vertical axis. To achieve this, an HSV color mask was constructed based on the average and standard deviation values of my red sweater's color, segmented manually from a random target image (A). The mask was then applied to every image iteratively after reducing its size to half and converting it to the HSV color format. The 2D geometric moments were then calculated from the foreground pixel values that form the isolated target (me) and its centroid (\bar{x}, \bar{y}) was extracted (B). The computed centroid can be seen correctly placed through this methodology by a blue dot over my body. Note that the HSV color range does not have a standard format. *OpenCV* displays the Hue values between 0 and 179 while online color pickers typically use the 0-360 range. Similarly, the Saturation and Value channels in *OpenCV* (0-255) also differ depending on the online sources that present these values from 0 to 100%. This is the reason why the HEX color codes are displayed here next to the HSV values (A).

left edge of the image to where foreground pixels are located ($M_{1,0}$) by the binarized region area ($M_{0,0}$), the x-axis centroid can be extracted ($\frac{M_{1,0}}{M_{0,0}} = \bar{x}$). The same can be applied to extract the y-axis centroid through the fraction $\frac{M_{0,1}}{M_{0,0}} = \bar{y}$.

This approach correctly identified all the targets among the non-target images within both datasets. However, one may argue this defeats the purpose of using a brain-in-the-loop in the first place but this is simply not the case. This would be like saying a color mask would be enough to correctly distinguish a particular husky from a specific german shepherd within a dataset only containing images of these two dogs. If the dataset is small, this could work most of the time. However, this approach is not generalizable if another husky and german shepherd are introduced with slightly different color schemes or even for images of the exact same dogs but with different lighting and saturation levels within the same dataset. On the other hand, a computer vision deep learning model can actually learn to pick the differences between both dog races and can generalize this distinction to several different scenarios. The color mask approach used in this project leverages a property that can solely be exploited within the datasets used here for automatic labeling and does not apply to real search and rescue scenarios where a color assumption for the target is impossible.

From this, and because it is standard to use *excel* lists when creating visual stimuli presentation in *PsychoPy*, a spreadsheet (.xlsx) file was created comprising 3 parallel categories, more specifically, a list of the processed image paths, a category depicting if the individual image was classified as a target or otherwise, and if so, where in the image the target is located in the horizontal and vertical directions. Before writing the information to *excel*, this 3-column list was reshuffled row-wise.

4.1.1.4 Experiment parameters

Once the visual stimuli are ready to use, some important parameters related to the RSVP *oddball* paradigm are left to be defined by conjugating both the N2pc/N2pcb and P3 elicitation limits. These include the total number of stimuli per trial block, the number of trial blocks, the proportion of non-target to target images, ITI, and possible restrictions against two consecutive target stimuli.

Regarding the number of stimuli used, also referred to as the number of trials, when the list of images was created, a total of 24 target images were included in the training dataset. Based on the criteria related to *oddball* designs already described, a 10% of the target to non-target stimuli is the common proportion used to produce a reliable P3 peak [33], which is not a relevant parameter for the N2pc/N2pcb[91]. Consequently, to maintain this proportion, 216 non-target images were selected along with the 24 targets, resulting in a 240-trial array for the **Training** dataset. Similarly, the same proportion was maintained in the **Testing** dataset, comprising 16 targets and 146 non-targets for a total of 160 images.

ERPs vary in size, both in amplitude and the time window they cover, resulting in the number of target trials required to obtain reliable potentials being dependent on the number of trials for

each specific component. The P3 is one of the largest ERP components with around $10 \mu\text{V}$ in amplitude and requires fewer trials to achieve a reliable component [92]. A study where this parameter's influence on ERP components was researched states that for the P3, a number of target trials above 14 are enough to obtain a highly discriminant P3 [92], while others indicate 20 as the bare minimum [93]. However, because of its reduced size ($1 \mu\text{V}$), the N2pc/N2pcb requires more target trials and becomes the determinant factor to set this limit. Following the literature guidelines for the N2pc/N2pcb acquisition, it is recommended to produce more than 100 target trials [92]. Nonetheless, to avoid the risk of not producing enough N2pc/N2pcb responses since some trials are expected to be excluded due to random noise during EEG acquisitions, a total of 10 trial blocks were defined for the **Training** image collection, resulting in 240 target and 2160 non-target trials. This matches the equivalent parameter set by the researchers in the EEG dataset used, where a total of 2400 stimuli were presented, of which 240 were the *oddball* class. On the other hand, for the **Testing** images, because fewer data are required, only 8 trial blocks were presented, resulting in 128 total targets and 1152 non-targets. Moreover, subjects were presented with a fixation cross centered on the screen for 3 seconds before each stimulus array burst and were asked to maintain their gaze on its center after its disappearance during image presentation periods.

Despite the numerous trials at first glance, this is where the RSVP design shows its strengths. In this project, a 5Hz presentation rate (200 ms ITI) with no intervals between consecutive images was used, corresponding to a 48-second window for each trial block in the **Training** protocol and 32 seconds for the **Testing** image trial blocks. Considering a hypothetical but realistic resting time of 12 seconds between each trial block, the whole experiment has a duration of just 10 minutes for the **Training** protocol and 6 minutes for the **Testing** counterpart, which provides minimal strain on both vision and mind.

One of the most important considerations that must be taken into account during this experiment' design is related to the limited temporal ability of the brain to deploy visual attention during consecutive presentations of target stimuli, more specifically within less than 500 ms of each other [33]. In other words, with a putative presentation rate of 10Hz, there must be 5 non-target images between two target stimuli so that the subsequent response isn't suppressed. This phenomenon is commonly referred to as *attentional blink* [33].

In addition to the *attentional blink*, other types of so-called cognitive blindness can occur, namely the *change blindness* and *saccadic blindness* [33]. *Change blindness* refers to the latency period that humans take to identify non-trivial variations in the presented images. This means that users will logically take longer to identify targets in cluttered images when targets do not pop easily from the rest of the image, reducing user performance during the experiment. On the other hand *saccadic blindness* can be described as the selective suppression of visual sensitivity and perception during rapid eye movements denominated saccades. These occur approximately three times per second and its function is responsible for increasing the perceptual resolution by capturing different parts of the world on the *fovea* rapidly[94]. Despite the speed at which these are performed and the images are swept through the retina, no motion or blurring is apparent, and a stable world is perceived. This stability results from the mechanism of saccadic suppression and

thus, when a person saccades between two locations, the presented image may not be perceived and *saccadic blindness* occurs, with evidence suggesting loss of visual information for periods of 50 ms [94]. This threshold defines the theoretical absolute minimum duration for a stimulus in the screen (20 Hz) to bypass the saccadic suppression [33].

Given this assumption and because the stimuli order of presentation is randomized before initiating every trial block, an algorithm was added extending the *PsychoPy python* script that encapsulates *attentional blink* periods and prevents two target stimuli from appearing within 500 ms of each other. With an ITI of 200 ms, a target separation of 3 images was coded, representing a period of 600 ms. Additionally, the script averts that a target image occurs within the first second (5 images) of each trial block, avoiding unprepared participants. Although participants choose to start each trial block as they see fit by pressing the "SPACEBAR" key on the computer, the shift of attention between the keyboard and the screen can take a few milliseconds and by imposing this restriction no unnecessary target response is lost.

Extending on this, a few more changes were made to the protocol script. The objects that handle the saved data with digital timestamps and the order of the stimuli were removed to improve efficiency. In addition to not providing useful information, the algorithm that handles target separation requirements happens after the list of the ordered stimuli is already formed, which by default *PsychoPy* automatically saves on an *Excel* sheet. Meaning there is a mismatch between the list and the actual order of the stimuli that appear onscreen. Instead, each new stimuli order defined right before a trial block will be saved in an *Excel* file at the end of each trial block to track the instance each defined target centroid appeared during the experiment.

To conclude the paradigm's design, one routine incorporating a message on the screen indicating participants to perform jaw-clenching movements for 20 seconds was added before the first trial block loop on the **Training** protocol. This refers to another preventive measure that captures common noise sources (EMG) in parallel to brain activity, enabling its later removal from the entire signal. A diagram illustrating the protocol, including both **Training** and **Testing** sessions that each participant underwent sequentially, is represented in Figure 4.7.

4.1.2 EEG Instrumentation

INESC TEC offers many ways to record EEG signals, each with particular applications in mind. These refer to the *cython + daisy* boards produced by OpenBCI [95], the two iterations of the Muse headband [96], Crown by Neurosity [97], and the recently acquired Micromed SD LTM PLUS 64 [98]. Albeit all being acquisition systems, these systems are in completely different categories. The first three are under the umbrella of low-cost systems used for prototyping BCIs, whereas the Micromed SD LTM PLUS 64 falls into the category of high-end, clinical-grade EEG systems designed for more rigorous and comprehensive neurological studies. An important detail is that this model is the same one used in São João Hospital to study and monitor several patient cases, posing a crucial familiarization step for every Biomedical Engineer. Because of all this, the Micromed acquisition system offers other dimensions of functionality and signal quality that the other systems cannot reach, making the choice of which one to use, easy.

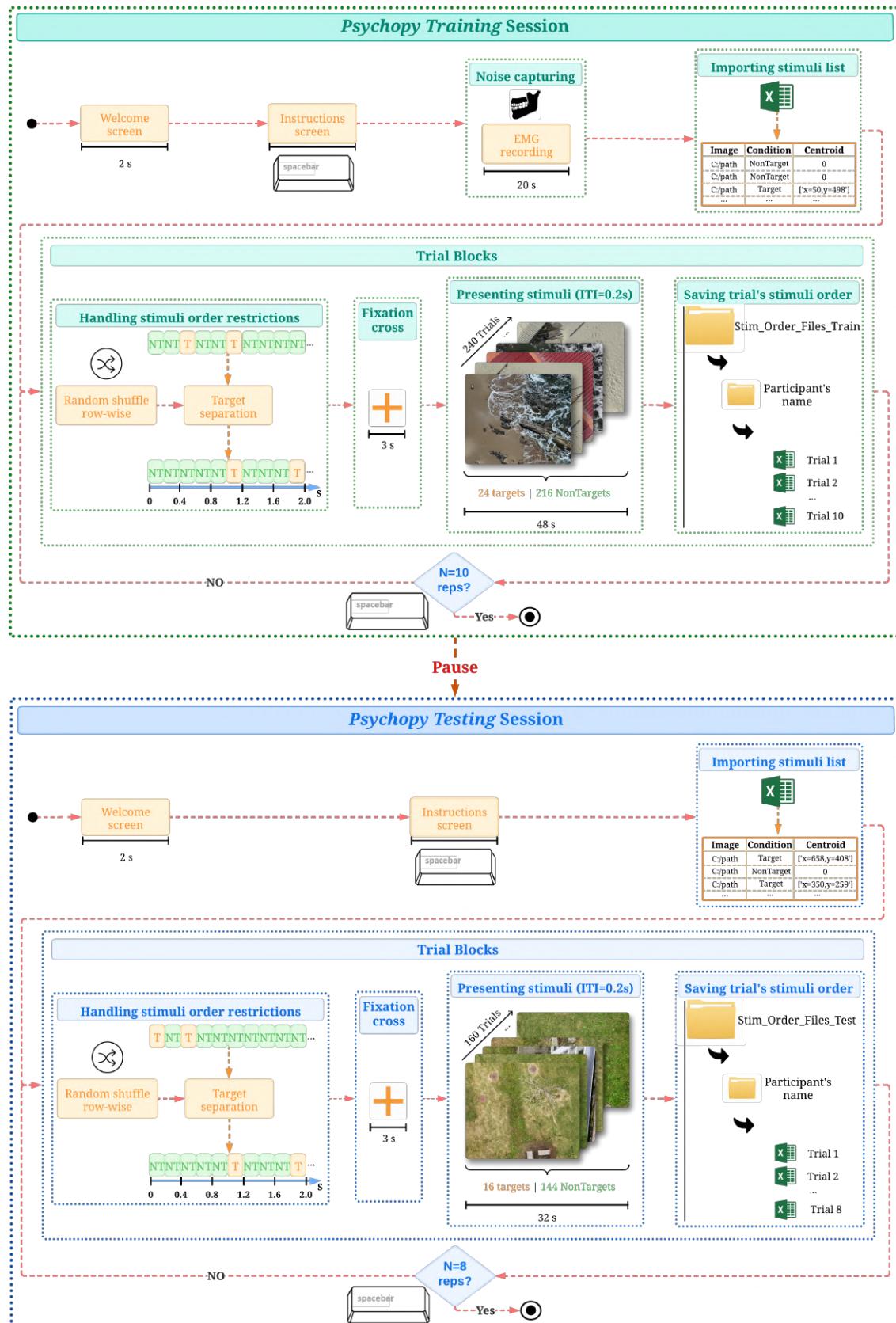


Figure 4.7: Diagram containing the routines that form the *RSPV oddball paradigm*, including both the **Training** and **Testing** portions that are done concurrently. The routines that have a "SPACE-BAR" key placed next to them indicate that the participant has to press it to continue the session, having total power on the resting period's duration.

4.1.2.1 Micromed EEG Amplifier

The Micromed amplifier delivers support to up to 64 simultaneous EEG channels, synchronized Full HD Video with night vision, relevant to monitor both epileptic episodes and sleep conditions, trigger implementation, custom DC modes in all inputs, and up to 16 KHz of sampling rates. Furthermore, this system also provides the *BRAIN QUICK software* that offers capabilities to easily manage several types of EEG montages and acquisition parameters [98].

The SD LTM PLUS 64 also boasts significant versatility in its functionality. The system can be configured to operate in three distinct modes: a wired headbox, a wireless headbox, or an ambulatory recorder. This adaptability allows for flexibility in various recording environments. In a clinical setting, the wired headbox configuration provides a stable and reliable connection for EEG acquisition during patient monitoring. The wireless headbox mode offers increased patient mobility within a limited range, which in concomitance with its lightweight design reduces discomfort during extended monitoring sessions. Most importantly, the ambulatory recorder functionality enables continuous EEG data collection even when the patient is outside the clinical setting.

Furthermore, the SD LTM PLUS 64 is capable of recording not only EEG signals but also other physiological signals relevant to a variety of study types. These additional channels can include respiration and oxygen levels, providing a more comprehensive picture of the patient's physiological state during EEG monitoring. This ability to capture multi-modal data allows for a more thorough analysis of potential factors that may influence certain conditions. A full depiction of this system including a glance at the *software* UI is shown in Figure 4.8

4.1.2.2 Hardware synchronization of the EEG

When studying neurophysiological data derived from the EEG signal, a precise time alignment between the signal and the mental state it is supposed to represent is mandatory to build models that respond to specific neural activity, as in BCIs. This is especially important when considering stimuli-locked and consequently, time-dependent features of the signal, such as ERPs. It has been shown that poor data synchronization results in decreased SNR in ERP responses [99]. However, establishing precise synchronization of the signals using a software approach by aligning the acquired data stream to a common time series with the clock of the computer as a reference is not a trivial task. If not carefully implemented, increased latency and *jitter* translated to the absolute delay that a sample takes to reach the computer, and the variability of these delays, respectively, can be expected [99]. This is the reason many EEG amplifiers provide auxiliary analog channels that accept triggers defined by the researcher to synchronize multimodal data streams.

Amplifiers are able to send packages of data at the same time to computers with no latency and *jitter* provided that the same amplifier is used to simultaneously acquire data from the different sources in parallel [99]. The standard approach in these cases is the use of a Light Dependent Resistor (LDR) implemented in a voltage divider circuit, feeding different DC voltage levels depending on the stimuli type to one of the auxiliary amplifier channels' during the EEG acquisition.



Figure 4.8: Representation of the Micromed SD LTM PLUS 64 amplifier and the *BRAIN QUICK® Software* [98]. Note that this figure has two of these amplifiers in parallel amounting to a total of 128 electrodes recorded, showcasing their system's workaround for higher-density acquisitions. Yet, only one unit of these was used for this project.

The LDR is a component that fits perfectly the purpose of triggering on-screen instances of stimuli presentation or when an action is asked to be performed, such as in RSVP and MI paradigms, respectively. This is due to its photoconductivity properties of varying the resistance with different incident light intensities. In the dark, its resistance is maximum and it decreases as the light intensity increases. By associating the LDR to a resistor with a fixed value in a voltage divider, as configured in image 4.9, the varying resistance will be transformed into a varying voltage that can be read by the amplifier.

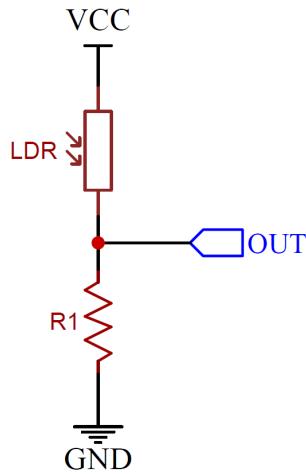


Figure 4.9: Schematic of a voltage divider circuit comprising an LDR, drawn in the *EasyEDA* online tool for schematic drawing and circuit simulation.

In this configuration, the output voltage will approximate zero for the darkest black on the screen due to the majority of the voltage drop occurring in the LDR, while its resistance is maximum. Inversely, the brightest white displayed will significantly decrease the resistance and cause a lower voltage drop, outputting a voltage closer to the upper range depending on the V_{cc} and $R1$ values (Figure 4.9). The inner workings of this process can be visualized in Figure 4.10 and the way the output voltage varies in the configuration used follows the voltage divider equation:

$$V_{OUT} = V_{CC} \times \frac{R_1}{R_1 + LDR} \quad (4.2)$$

Because no such device exists in the *MICROMED* catalog that delivers this functionality, a circuit that interfaces this process between the presented stimuli and the amplifier was developed for the SD LTM PLUS 64 system. The auxiliary channel on the amplifier targeted for this purpose is the DC channel, allowing for two different triggers simultaneously with an adapter given by *MICROMED* during the system's purchase. This connection is highlighted in Figure 4.11. The first problem being tackled during the circuit design was understanding the power delivery restrictions of the EEG amplifier. Each trigger channel comprises a negative and positive DIN 1.5mm Touch-Proof (TP) socket input ranging from -2.4 to 2.4V. These sockets are the same as the ones that feed the EEG signal from the EEG cap to the *MICROMED* system.

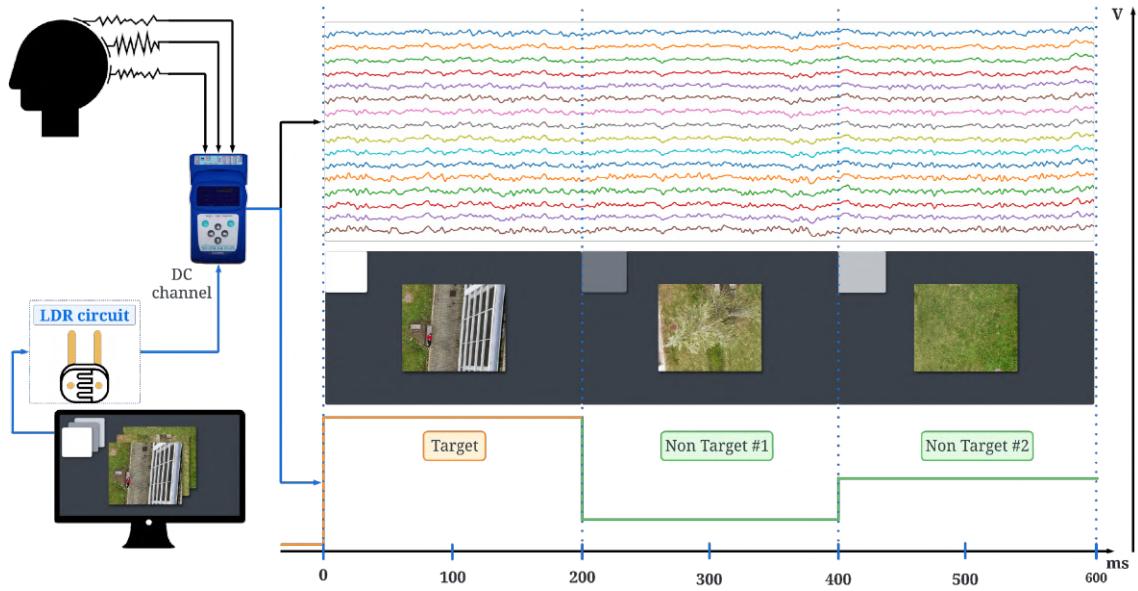


Figure 4.10: Depiction of the EEG synchronization concept through an LDR implemented in a voltage divider circuit. During the target class presentation, a white square is placed on the top left corner of the monitor, inducing the highest voltage output on the circuit (orange line). Similarly, darker greyscale squares are used for the non-target class, generating lower voltages (green). The use of two intermittent greyscale values in the non-target class is due to images being presented with no intervals, thus enabling separation between two consecutive non-targets. The EEG acquisition and voltage values read by the circuit are combined in the amplifier, allowing to track the instant each stimulus appears and its duration.



Figure 4.11: Connection of the DC dongle to the EEG amplifier.

Albeit being a useful range, using negative voltages can significantly increase the complexity of the circuit being designed since Operational Amplifiers (AmpOps) would have to be implemented to create a virtual ground, creating an internal reference in the circuit where the total input voltage would be divided in half, where the lower half would constitute the negative DC signal and the upper half the positive range. Another option would be to use some sort of DC offset to center the input voltage around a positive and negative voltage with the same absolute value. However, this approach would also not be ideal since coupling two voltage dividers without a *buffer*/unity gain amplifier or transistor logic interfacing them would make it difficult to meet both voltage dividers' requirements concurrently. Additionally, in both approaches, the correct power delivery to drive the AmpOps would have to be met. All of this would add to the design's complexity in terms of signal conditioning, the number of components, board size, and, as a result, its cost. For this reason, a simpler and cost-effective but equally useful solution given the number of components required and those already available in the Lab was developed for the range between 0 and 2.4V. This range is perfectly suitable for many types of EEG paradigms, not just the typical binary experiments, such as the *oddball* (target vs non-target stimuli), but also when more types of conditions are required, such as in the MI paradigms where up to 5 different classes of EEG data can be required to segment.

Following the voltage requirements' determination, a few options to deliver the power to drive the circuit were also considered, namely a battery and a USB connection. The doubts about the most convenient approach were whether the noise associated with a USB connection would be significant. By utilizing a battery, the circuit is completely isolated from the main power line and, theoretically, less prone to its noise. Based on this rationale, a study to ascertain whether this noise component would be significantly different between these two options was conducted. For this, a circuit prototype was built on a breadboard comparing both options with a *Keysight InfiniiVision DSOX2004A* oscilloscope.

After defining the input and output voltages, the breadboard prototype was translated into a formal circuit design using *Altium Designer* [100]. This powerful Printed Circuit Board (PCB) design software facilitates not only schematic capture but also simulation and analysis of the circuit. Additionally, *Altium Designer* allows for 3D visualization of the board, provided we have the *STEP (.stp)* files for each component. Following design completion, the necessary production files, known as *Gerber files*, were generated. These files encapsulate all critical PCB information, including copper trace layouts for component connections, layer structures, drill hole specifications, and pad locations. This data serves as the blueprint for PCB fabrication.

The final step of a PCB design is to print the board and weld the components in their respective position. The first attempt to print the board was conducted on a *Voltera V-one* machine [101] available in the INESC TEC labs and represented in Figure 4.12. This machine offers an in-house and quick solution to prototype and assemble PCBs, eliminating the need for external fabrication. It works by printing conductive ink on a board and undergoing a posterior curing process to solidify the traces. In the final step, the machine will dispense solder paste and reflow it to attach the electronic components. This PCB printing method was then compared to the standard method

of board manufacturing by requesting its fabrication from a company dedicated to this purpose, namely *Eurocircuits*. The soldering was done using the recommended alloy (SnBiAg1) and temperature (180 °C) for the *Voltera V-One* printed board and a tin alloy at 300 °C for the ordered PCB.



Figure 4.12: Representation of the *Voltera V-one* PCB printing machine [101].

In parallel, a case to encapsulate the PCB was designed in *Autodesk Fusion 360*. This Computer-Aided Design (CAD) software enables the creation of 3D models that can later be fabricated through in-house 3D printing processes. As a base model to dictate the case dimensions and features, a 3D model (.stp) file containing the board and its components was extracted from *Altium Designer* and imported to *Fusion 360*. Lastly, a support to hold and fix the case to a monitor was also crafted. The parts were printed using the Fused Deposition Modeling (FDM) procedure, a 3D printing technique that builds objects layer-by-layer by extruding melted plastic filament along the programmed paths.

4.1.3 EEG set-up and acquisition parameters

With the intent to study the elicitation of the visual N2pc/N2pcb and P3 ERP components, a focus on the occipital and parietal regions of the scalp was necessary. A good number and distribution of electrodes over the whole scalp can be advantageous if the use of spatial filters is required to perform the intended analysis, such as in the case of Motor Imagery BCIs. However, even though these can still be helpful to the ERP analysis in some specific cases, these would only be complementary to the data cleaning and not the main step in the processing pipeline. Furthermore, using more electrodes equates to an increase in the set-up time for the EEG.

With the option to use a 10-20 system 32-electrode cap or a 10-10 system 64-electrode cap, the recording of the N2pc/N2pcb requires the presence of the intermediate positions offered by the

higher-density cap, namely the PO7 and PO8 channel positions where these components are theoretically maximum [19, 68]. These are absent in the 32-channel cap. Subsequently, the 64-channel cap offered by Besdata [102] was used with a total cover of the Central, Centro-Parietal, Parietal, Parieto-Occipital, Occipital, Temporal, and Temporoparietal regions of the scalp, amounting to 33 electrodes recorded. Even though these are not distributed throughout the whole scalp, this placement still allows for the robust use of spatial filters if necessary due to the high density and number of electrodes used. Figure 4.13 depicts both the cap and montage used to acquire the EEG signals.

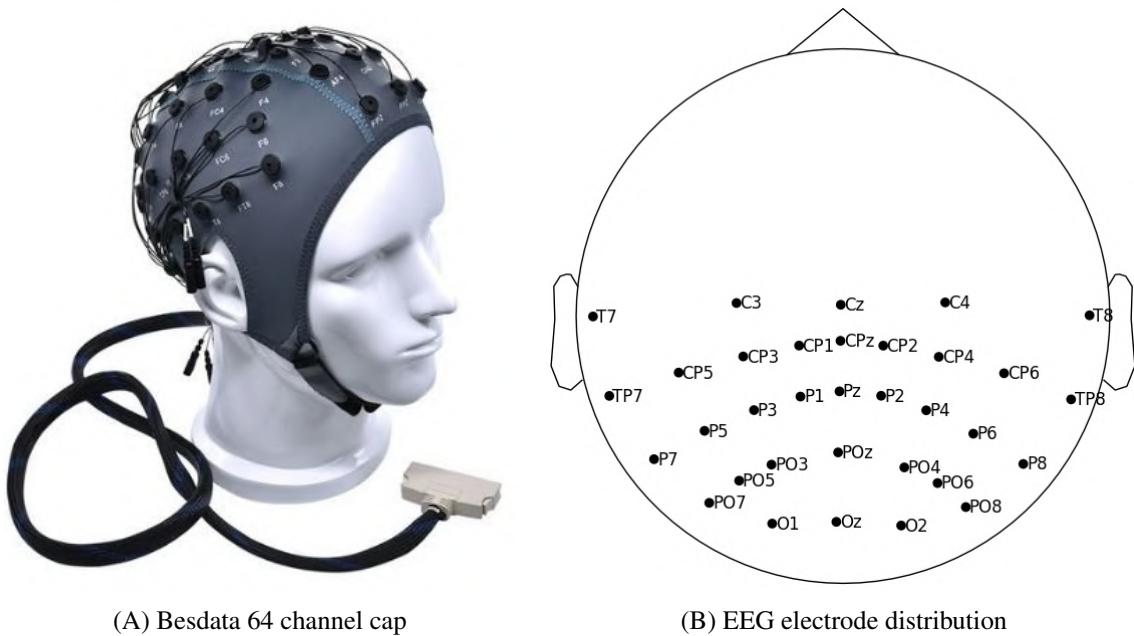


Figure 4.13: The Besdata 64 channel cap (a) and the EEG montage (b) depicting the scalp distribution of electrodes.

Because every EEG system requires a hardware ground and reference channel from which the differential potential is computed, these were placed in the left and right earlobe, respectively, to corroborate with the literature on similar RSVP studies [19]. Analogously, the sampling rate was also matched to 2048 Hz.

Before running the acquisition, an individual impedance check was performed on each participant to measure the impediment to the flow of alternating current between the scalp and the electrodes. This marks a crucial step in the EEG acquisition that ensures the signal acquired does in fact constitute a good representation of the ongoing brain activity. To achieve this, the impedances are reduced with the aid of conductive gel placed between the scalp skin and electrode sites through holes in the cap. This method requires applying priorly a circular motion of the electrode site to create a hair "nest" to hold the gel. Once created, a moderate quantity of gel is applied to the site with a syringe. A circular motion is then applied with the syringe still inside the respective electrode hole and once more after taking it out. This ensures a homogeneous

distribution of gel, improving the signal conductivity. This was then replicated for every electrode of the montage and impedances were mostly kept between $2\text{ }K\Omega$ and $10\text{ }K\Omega$ [103].

Despite straightforward sounding, it's crucial to note that this method requires meticulousness and patience as too much gel will negatively impact the signal, causing electrodes to bridge and smearing spatial resolution. Furthermore, the variability of gel quantity necessary between participants is high due to hair thickness, quantity, and density differences, thus requiring attention and sensitivity to not overcompensate undesired impedance values.

Once the impedance check is finished, the signal is acquired with two computers. The dedicated desktop handles the signal acquisition through the *Micromed* amplifier, including the analog trigger interfaced by the LDR circuit. A laptop is used to display the visual paradigm on an external monitor at the C1 lab in INESC TEC. An illustration of this setup is displayed in Figure 4.14.

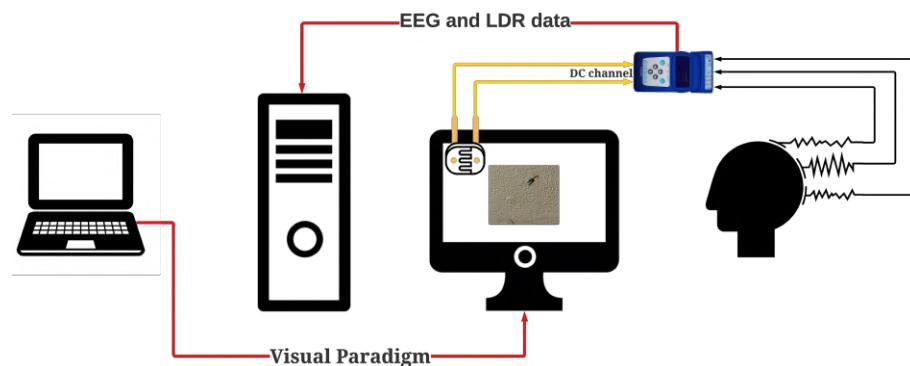


Figure 4.14: The signal acquisition setup using two computers, an LDR circuit, a monitor, and the *Micromed* EEG amplifier.

4.1.4 Participants

All subjects in the study were healthy individuals with no known neurophysiological anomalies, all exhibiting right-handedness. Given the high presentation rates that are used in the RSVP protocol, participants were screened for any personal or family history of epilepsy and rejected if present [19]. Every participant was priorly informed about the experiment's purposes and methods of EEG acquisition and signed a written consent approved by INESC TEC's ethics committee for the data acquisition and its uses in this study. A copy of this document is displayed in Figures A.1, A.2, and A.3 of Appendix A.

All subjects participated in the **Training** and **Testing** protocols, where in the former, a trial block (48 s) was presented to familiarize them with the protocol. On the other hand, the **Testing** paradigm was seen for the first time during the actual acquisition. This is important since participants had access to the full image training dataset, despite their order changing in every trial block ($n = 10$), but not to the **Testing** image collection. This approach was used to track the brain performance to unseen images through time (8 trial blocks) and assess if the performance improved as the brain adapted to the new visual stimuli. Additionally, the first trial block users are shown

serves another purpose, *i.e.*, users are asked to use the first fixation cross to orient themselves to its center through their reflection on the screen, ensuring the visual angles are symmetric throughout the experiment.

In total, 6 subjects viewed both datasets concurrently with a pause between the protocols, amounting to 884 MB of EEG data. The entire subject demographics can be seen in Table 4.1

	Male	Female
Number of participants	5	1
Age	23.2 +/- 1.93	21

Table 4.1: Participant demographics (mean and standard deviation for age).

4.2 EEG Processing pipeline

The scripts developed for the data analysis and processing were implemented in a dedicated environment created in *python 3.10.9*. Programming environments in *python* are pivotal when considering projects of this dimension that require proper dependency management. A dedicated environment enables version control and package isolation, not affecting the system-wide installation of *python* or other projects. This is especially convenient when dealing with dependency conflicts, ensuring these are localized to the corresponding environment. Furthermore, the same environment can be recreated across different machines, making it easier to reproduce bugs and results by sharing the environment configuration files. This guarantees reproducibility but also facilitates deployment, as is required in later stages of BCI systems.

Python, due to being open-source, offers a high amount of versatility in data analysis and real-time applications. *MNE* represents a commonly used dependency that provides tools for exploring, visualizing, and analyzing neurophysiological data [104]. It mirrors the widely known *EEGLAB* for *MATLAB* but in *python*, offering similar but extended data cleaning and processing capabilities with a stronger focus on source localization analysis of MEG and EEG data [105]. Because of this, the *MNE* library was used to analyze and process the EEG signal, in conjunction with the *Sklearn* and *Pytorch* libraries that handled the Classification portion of the pipeline. Libraries *scipy* and *matplotlib* were also used for additional signal processing and plotting, respectively.

An overview of the pipeline is presented in diagram 4.15. The blocks displayed are grouped by color according to the type of signal conditioning they represent. Because of the intricacies of data flow in some regions of the diagram, the details of the processing pipeline are discussed in the next subsections for each group and not by following the chronological order of each block. As a result, the signal quality estimation portion will be discussed first, followed by the signal processing techniques, ERP analysis based on the visual angle of each target centroid for the N2pc/N2pcb and target vs non-target responses for the P3, and the feature vector formation with subsequent classification.

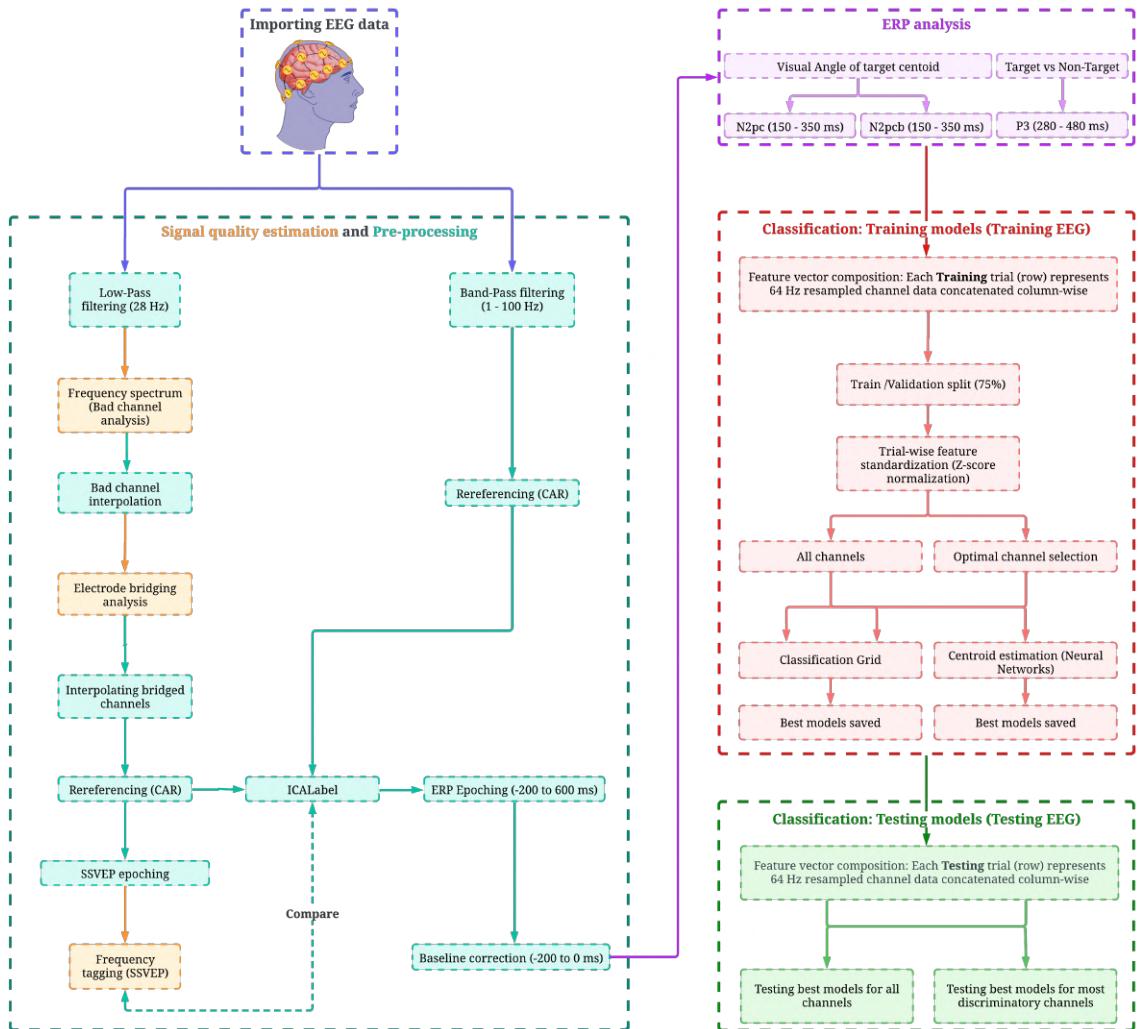


Figure 4.15: Overview of the EEG data pipeline. Signal quality estimation groups are displayed in orange color and pre-processing techniques are in light blue. A dashed arrow connecting the frequency tagging and ICA blocks indicates that data did not flow directly between these two processes but the information extracted in the frequency tagging analysis guided the ICA supervised decisions in component removal, hence the "compare" keyword. The ERP analysis is illustrated in purple, the feature vector composition with subsequent model training in red, and the best model's testing in green. Each one of these groups of data conditioning is detailed in the following sessions. Note that on the classification block, the Neural Networks were only trained and tested with all channel features, whereas the classification grid was approached with both an optimal channel selection and all channel information.

4.2.1 Estimating signal quality

As already described, the EEG is highly susceptible to various artifacts that naturally occur in parallel to the brain activity and are picked by the scalp sensors, overlapping and blurring the signal of interest. As a result, to ensure the neural activity is properly manifested in the acquired signal, a few metrics were developed to assess the signal quality and from this, guide the best technique to reduce the artifact's representation on the overall signal.

4.2.1.1 The frequency spectrum

The best initial procedure to analyze the EEG quality is computing the frequency spectrum of the signal using the *Fast Fourier Transform* (FFT). The intuition behind this algorithm is relatively simple, *i.e.*, the entire signal is multiplied iteratively using the dot product by a real and imaginary sinusoidal concomitantly from the lowest frequency value of the signal to the highest (*Nyquist frequency* = $\frac{\text{Sampling Rate}}{2}$ Hz), measuring the similarity of the EEG to each frequency unit [106]. The dot-product of the real domain sinusoidal with the EEG outputs the magnitude of that specific frequency in units of $\frac{V}{Hz}$, while the dot-product with the complex sinusoidal delivers the phase information of that same frequency in *radians*. Transcribing this behavior into the mathematical language results in the equation 4.3, where $X(k)$ is the frequency domain representation of a discrete signal $x(n)$ of length N [106]. The variable k ranges from 0 to $N - 1$, representing each frequency component and matching the number of time samples of the original signal (N), demonstrating that the FFT generates a perfect representation of the signal on the frequency domain with no loss of information. This means its inversion through the inverse FFT will produce an exact copy of the original time-domain signal [106]. Finally, the $e^{-j\frac{2\pi}{N}kn}$ term depicts the complex sinusoidal basis. With this notion in mind, the dot-product of the EEG signal with a 10 Hz real sinusoidal represents the average EEG energy (amplitude) at 10 Hz.

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j\frac{2\pi}{N}kn}, \quad k = 0, 1, \dots, N - 1 \quad (4.3)$$

The frequency spectrum provides a much quicker glance at the full signal than the time domain due to its characteristic $\frac{1}{f}$ trend, where an acquired signal can be instantly identified as a true EEG and proceed in the analysis pipeline if this shape is present. Moreover, this spectrum can also be used to identify the presence of a potentially bad channel whose frequency trend diverges significantly from the average $\frac{1}{f}$ curve in the same participant. Accordingly, each acquired EEG was initially transformed into its frequency spectrum and evaluated to its overall $\frac{1}{f}$ shape and presence of bad electrodes.

4.2.1.2 Electrode bridging

The increasing higher-density scalp recordings albeit improving the localization capacity of brain patterns also impose some technical limitations. With the necessity of applying electrolyte gel to improve connectivity, overapplication will result in low-impedance links between adjacent

electrodes in the scalp, causing subtle topographic distortion, unpredictable localized artifacts, and the attenuation or insertion of small hemispheric asymmetries [107]. This is especially important to avoid when studying potentials like the N2pc/N2pcb as they rely on hemisphere independence and even subtle intrahemispheric differences of adjacent electrodes when trying to study its modulation to specific target locations [19, 68]. Due to its highly localized nature and size, both in time and amplitude, the N2pc/N2pcb are much more prone to tenuous distortions of this type than the large and widely scalp-distributed P3. In addition, its location on the scalp, namely the ventral occipital region, marks the most likely region where this issue arises due to the vertical orientation of these electrode sites, motivating the importance of a careful and measurable examination of channel bridging [107].

In an attempt to identify the presence of this artifact, a method was developed to objectively quantify electrode bridging by measuring the electric potentials' similarities of adjacent electrode sites. This method is reference-independent and is based on a variation of the *Hjorth Laplacian* algorithm, being denoted as the *Intrinsic Hjorth* where the electrode spatial distances are adapted to non-spatial electric distances [107]. According to this method, the electrical distance is nothing more than the variance of the signal's pairwise subtractions. Channels whose electric activity mirrors another channel will have very low electrical distance, indicating the presence of electrical bridging/smearing [107]. This process is achieved by following equation 4.4.

$$D_{i-j} = \frac{1}{T} \sum_{t=1}^T (P_{i-j}(t) - \overline{P_{i-j}(t)})^2 \quad (4.4)$$

As one can see, the formula resembles an analogous version of the sample variance, where D_{i-j} depicts the electric distance between channel i and j . This value is computed by summing all squared channel differences from the mean $(P_{i-j}(t) - \overline{P_{i-j}(t)})^2$ across the temporal dimension of length T and starting at sample $t = 1$. The resulting value is then divided by the number of samples T . The spatial smearing was evaluated in this manner for each participant.

4.2.1.3 Frequency tagging of the stimulation frequency

The image presentation rate, or in other words, the stimulation frequency of the images passing through the retina, introduces a constant component in the form of an SSVEP to the EEG. However pertinent to some applications, in the context of RSVP BCI this component is a "necessary evil" obfuscating the desired target response in the ERP [16]. Furthermore, the SSVEP has a maximum projection over the occipital region, directly influencing some of the N2pc/N2pcb-targeted sensors. As a result, estimating this component and its harmonics will again provide information relevant to devising ways for its removal, if possible, or at least to account for this "artifact" when explaining the results.

Before computing these metrics, the portion of the signal containing the SSVEP periods was segmented. This period represents the trial blocks time window, specifically 48 seconds for the **Training** and 32 seconds for the **Testing** protocols, respectively. The *steady state* response can

take a while to stabilize with the transient phase in the time window's beginning potentially provoking a distortion of the signal estimates [73]. For this reason, the first second of these periods was excluded from the analysis establishing a more accurate portrayal of the SSVEP component.

Once segmented for all 10 and 8 epochs corresponding to the number of trial blocks in the **Training** and **Testing** sessions, respectively, the *Power Spectral Density* (PSD) and *Signal-to Noise ratio* spectrum were extracted for all participants. The PSD is an equivalent measure to the spectrum described above. In fact, it is calculated using the same algorithm, the FFT, but produces a measure of power ($\frac{V^2}{Hz}$) instead of energy, exacerbating the sometimes tenuous differences in the most prominent oscillatory components over the remaining spectral values. Mathematically, obtaining the PSD (S_{xx}) translates into squaring the magnitudes of the FFT results ($|X(k)|$) and then introducing a normalization factor according to the number of samples ($\frac{1}{N}$), as depicted in equation 4.5 [108].

$$S_{xx}(k) = \frac{1}{N} |X(k)|^2 \quad (4.5)$$

A *boxcar* (rectangular) window was used to encapsulate the signal, which is equivalent to no window at all as regular windowing (e.g., Hamming) causes the consistent frequencies intrinsic to SSVEPs to no longer have constant signal amplitudes, distorting the frequency spectrum by reducing the power estimate at the stimulation frequency and producing adjacent "sidebands", commonly referred as to as "overspill" [109]. More concisely, this window does not impair the isolation of the targeted SNR frequencies on the PSD spectrum. The power spectrum was then expressed in decibels (dB) as logarithmic representations improve visualization and interpretation of power-based plots according to the formula: $S_{xx}(dB) = 10 \log_{10} (S_{xx})$.

From the power spectrum, the SNR spectrum was extracted as a measure of relative power by following the method proposed by Meigen and Bach (1999) [110]. This method describes the ratio of power in a given frequency bin to a 'noise' baseline defined as the average power in the surrounding frequency bins. The algorithm begins by constructing an averaging kernel to calculate the mean noise level. This kernel includes ones for the specified number of neighboring bins on both sides and zeros for the bins to be skipped immediately adjacent to the target bin. The kernel is normalized by the number of neighboring bins used on each side to ensure it averages correctly. The PSD array is then convoluted with the averaging kernel along the frequency dimension accounting for the signal's edges by applying proper padding, ensuring the resulting signal contains the same length and that boundary effects are minimized. This convolution process computes the mean SNR at a specific frequency by averaging the power of its surrounding bins, excluding the skipped bins. Then, the PSD was divided by the mean SNR baseline at each bin to normalize the spectrum accounting for the $\frac{1}{f}$ decay, thus increasing the relative differences in the dominating SNR bins in comparison to the rest.

Here, the power at each bin was compared with the average power of three neighboring bins and skipping one directly next to it on each side since it provides good isolation of the SNR estimates at the frequencies targeted. This forms a kernel array with 9 samples $K = [1, 1, 1, 0, 0, 0, 1, 1, 1] \frac{1}{2n_{nf}}$,

where nf is the number of neighboring frequencies (1) on each side, hence the multiplication by 2. The kernel convolution (4.6) and SNR estimate (4.7) was then performed in the following manner [110]:

$$\text{Mean Noise}(i) = \sum_{j=-nf}^{nf} K(j) \cdot S_{xx}(i+j) \quad (4.6)$$

$$\text{SNR}(i) = \frac{S_{xx}(i)}{\text{Mean Noise}(i)} \quad (4.7)$$

Finally, the analysis converges to compare if the SNR values at a visual region of interest (vROI) are greater than the rest of the scalp. This vROI was defined as the cluster of channels POz, Oz, O1, and O2, where the SNR resulting from the SSVEP is expected to be greater. The significance of these differences was then evaluated by performing a *paired t-test*, in which the mean SNR of the vROI and the rest of the scalp channels were compared. These differences were then addressed at the stimulation frequency (5 Hz) along its two main harmonics at 10 and 15 Hz. The vROI is illustrated in Figure 4.16.

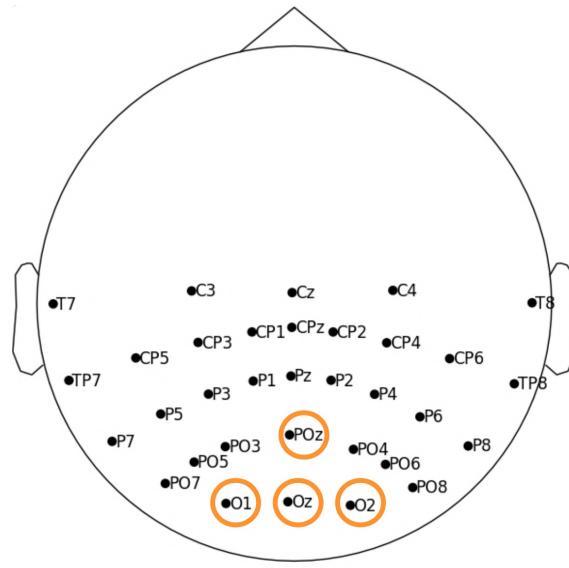


Figure 4.16: The cluster of occipital channels (circled in orange) used to estimate the SNR of the SSVEP caused by the constant image presentation rate (5Hz).

4.2.2 Signal Pre-processing

The artifacts encountered during the signal quality estimation phase were appropriately dealt with by applying signal pre-processing techniques and cleaning the raw signals. These correspond to band-pass filtering, channel interpolation, re-referencing, epoching the EEG, performing Independent Component Analysis, and baseline correcting the signals.

4.2.2.1 Band-Pass filtering

When extracting the recorded EEGs from the Micromed BRAIN QUICK software, the signal comes automatically band-pass filtered between 0.1 Hz and 512 Hz. Despite the lower pass-band (0.1 Hz) tackling the slow drifts associated with lower frequencies by removing the signals DC component, the upper band requires some reduction. Hereby, a Low-Pass filter with a cutoff frequency of 28 Hz was applied at the beginning of the pipeline. This removes the higher frequencies and most of the noise associated with them (*e.g.*, line noise and muscle artifacts) while maintaining oscillatory components from the cognitive processes of interest, namely the N2pc/N2pcb and P3 ERP components [19, 68]. Its frequency and phase response are depicted in Figure 4.17.

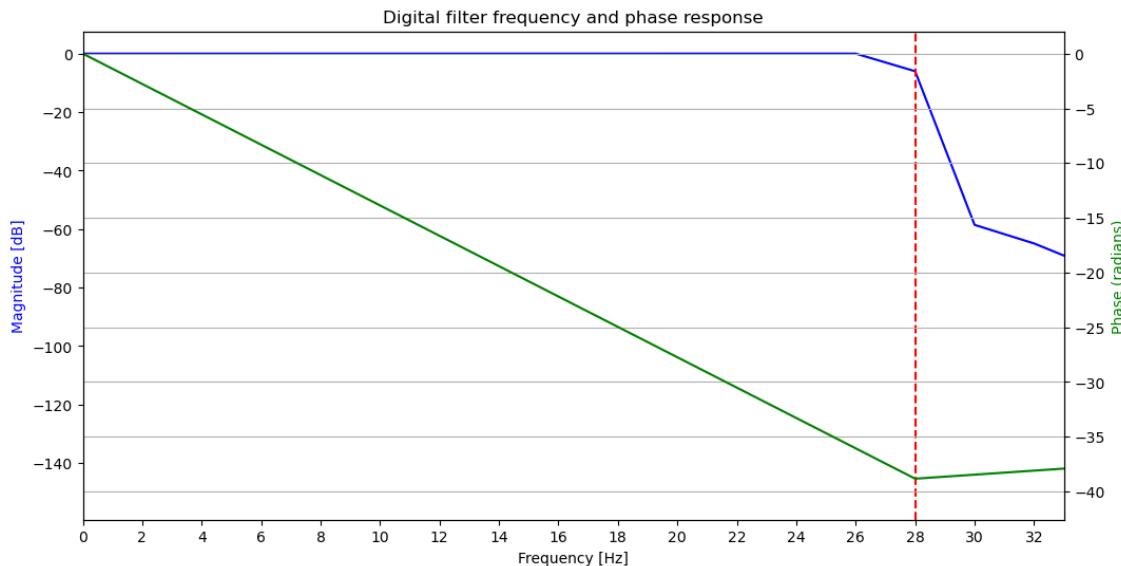


Figure 4.17: The frequency (blue) and phase (green) response of the low-pass filter applied at 28Hz on the EEG signals.

This filter was designed with a few considerations in mind so as to not distort the EEG. It fits the class of a Type I Finite Impulse Response (FIR) with 5001 number of coefficients (order). The high number of coefficients was specifically designed for *offline* signal processing as filters of this size are computationally heavy and impractical in real-time applications. However, because no time constraints are applied during *offline* processing, this length guarantees a steeper transition band of 4Hz, attenuating a shorter range of spectral components around the cut-off frequency, where the signal has a standard attenuation of -3 dB. Frequencies at 30 Hz are therefore fully attenuated (-60 dB) with attenuation beginning at 26 Hz (0 dB), as can be visualized in Figure's 4.17 blue line.

The FIR design, as opposed to the Infinite Impulse Response (IIR) format, is also a standard in EEG pre-processing [58]. In comparison, IIR filters provide even better transition bands (steeper) but at the cost of not being as stable and producing non-linear phase distortions. This means, the signal will present greater phasing distortions at some frequencies than others, making it difficult to fully restore distortions. This is a must to avoid when designing time-domain analysis of EEG

signals. FIR filters ensure a constant/linear phase shift, as indicated by the green line in Figure 4.17, providing a zero-phase response when the filter is applied a second time backward with the same number of coefficients, as it is done here. This zero-phase property is also assured by the Type I format imposed by the odd number of coefficients exhibiting symmetry around the filter signal's center and no group delay. Furthermore, Type I provides no constraints on the placement of zeros ($z = 1$ and $z = -1$), meaning these can be freely positioned to meet the filter design specifications.

Finally, a *Hamming* window was used to encapsulate the filter coefficients. The full mathematical equation is described below for the *Hamming* window (4.8) and posterior filtering (4.9), where $w(k)$ depicts the *Hamming* window coefficients and M the number of coefficients with k ranging from 0 to $M - 1$ [58]. Applying $w(k)$ to the filter coefficients b transforms these into windowed version \tilde{b}_k , preparing it for the convolution summation with the signal x at sample n and resulting in the filtered signal y .

$$w(k) = 0.54 - 0.46 \cos\left(\frac{2\pi k}{M-1}\right) \quad (4.8)$$

$$y(n) = \sum_{k=0}^{M-1} \tilde{b}(k) x(n-k) \quad (4.9)$$

4.2.2.2 Bad and bridged channels Interpolation

During some acquisitions, it was noticed that channel O1 was not responding normally to impedance checks caused by a faulty contact on the cap. Instead of removing this channel from the analysis and because this only happened during 3 of the total 6 acquisitions, it would not be sensible to discard a fair amount of clean data. Furthermore, in addition to being a useful approach to keeping cross-subject dimensionality, O1 is contained in the ROI for the N2pc/N2pcb analysis [19, 68]. In such cases, channel interpolation is often performed.

Interpolation of EEG channels is done using the spherical spline method, which projects the sensor locations onto a unit sphere and interpolates the signal at the bad sensor locations based on adjacent good channels [111]. This algorithm works by projecting the good and bad channels onto a unit sphere ($r = 1$), computing a mapping matrix that maps N good channels to M bad channels, and finally using this mapping matrix to compute interpolated data in the bad channels.

Spherical splines [111] assume that the potential $V(r_i)$ at any point r_i on the surface of the unit sphere can be represented by:

$$V(r_i) = c_0 + \sum_{j=1}^N c_j g_m(\cos(r_i, r_j)) , \quad (4.10)$$

where the $C = (c_1, \dots, c_N)^T$ are constants which must be estimated. The function g_m of order m is given by:

$$g_m(x) = \frac{1}{4\pi} \sum_{n=1}^{\infty} \frac{2n+1}{(n(n+1))^m} P_n(x), \quad (4.11)$$

with $P_n(x)$ being *Legendre polynomials* of order n , a sequence of orthogonal polynomials widely used in problems involving spherical coordinates. To estimate the constants C , two equations must be solved simultaneously:

$$C_{ss}C + T_s c_0 = X \quad (4.12)$$

$$T_s^T C = 0 \quad (4.13)$$

where $G_{ss} \in R^{N \times N}$ is a matrix whose entries are $G_{ss}[i,j] = g_m(\cos(r_i, r_j))$ and $X \in R^{N \times N}$ are the potentials $V(r_i)$ measured at the good channels, in this case, all except O1. $T_s = (1, 1, \dots, 1)^T$ is a column vector of dimension N . Equation 4.15 is the matrix formulation of equation 4.10, whereas equation 4.13 works like applying a common average reference to the data, further contextualized in the next section (4.2.2.3). From equations 4.15 and 4.13 results:

$$\begin{bmatrix} c_0 \\ C \end{bmatrix} = \begin{bmatrix} T_s^T & 0 \\ T_s & G_{ss} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ X \end{bmatrix} = C_i X \quad (4.14)$$

with C_i being the same as matrix $\begin{bmatrix} T_s^T & 0 \\ T_s & G_{ss} \end{bmatrix}^{-1}$ but with its first column deleted, therefore giving a matrix of dimension $(N+1) \times N$.

Now, to estimate the potentials $\hat{X} = G_{ds}C + T_d c_0$ at the bad channels the following equation is followed:

$$\hat{X} = G_{ds}C + T_d c_0 \quad (4.15)$$

where $G_{ds} \in R^{M \times N}$ computes $g_m(r_i, r_j)$ between the bad and good channels. $T_d = (1, 1, \dots, 1)^T$ is a column vector of dimension M . Plugging in equation 4.14 into equation 4.15 the final bad channel interpolation formula is obtained:

$$\hat{X} = \begin{bmatrix} T_d & G_{ds} \end{bmatrix} \begin{bmatrix} c_0 \\ C \end{bmatrix} = \begin{bmatrix} T_d & G_{ds} \end{bmatrix} C_i X \quad (4.16)$$

Channel O1 was interpolated for subjects 2 and 3 when the faulty connection was noticed and repaired. However, as a result of needing a wash after every acquisition, the cap's connection was impaired once again and channel O1 was interpolated for subject 6 as well.

For bridged channels, a variation of this process was implemented. The signal interpolation occurring for O1 discards its data and reconstructs the extremely noisy signal from scratch only with data corresponding to the surrounding locations. Contrastingly, bridged channels contain smeared but biological data, so a virtual channel is produced midway between the bridged pairs

and used to aid in interpolation rather than completely discarding the data from the two channels. This should only be applied, however, in situations where a low number of channels are bridged. Otherwise, interpolations would depend on channels that are also bridged. With this in mind, a threshold of 8 connections, not channels, was defined to decide if the data was interpolated in this manner. Therefore, if a lesser or equal amount of connections arose, channels were interpolated. This was done right after channel O1 interpolation to prevent using the noise from this channel in the calculations.

4.2.2.3 Common Average Referencing

When any electrical potential is measured, it is always relative to another potential. With the EEG, this notion is maintained. However, the way EEG data is referenced severely impacts its interpretability, with some types being better suited than others to evidence a particular form of neural activity relevant to the application being designed.

In this project, the EEG was referenced to the ear lobes since they do not capture brain data while providing a good Common Mode Rejection Ratio (CMRR). This emerges due to consistent skin contact in electrodes placed on the scalp and earlobes, with noise caused by impedances on the skin forming the Common Mode and canceling each other in the differential amplifier, creating a more neutral baseline that enhances the quality of EEG data. However, after the acquisition, the data is typically re-referenced with the most commonly used method in EEG research being the Common Average Reference (CAR) [58].

CAR can only be computed after completing the acquisition and is calculated by subtracting the mean potential of all electrodes for each time sample to each electrode individually, resulting in an average scalp potential of 0 throughout the signal. Its mathematically described in equation 4.17 in which V^{CAR} depicts the re-referenced signal at all channels, V^{ori} the original signal at all channels, N the number of channels, V_j the signal at each electrode j with j running from 1 to N , and t each time sample. It stems from the assumption that any electrical current sources coming from inside the skull, its net current potential on a 3-D spherical surface representing the head is 0, providing good noise reduction [58]. However, an important consideration is to carefully examine focal sources of electrode noise when computing CAR, since artifactual data will potentially propagate to every electrode recorded after its computation, contaminating the overall brain signal. This is the reason why re-referencing was done after filtering and electrode interpolation in the pipeline.

$$V^{CAR}(t) = V^{ori}(t) - \frac{1}{N} \sum_{j=1}^N V_j(t) \quad (4.17)$$

The best and most versatile re-referencing method is still debated but its wide adoption in EEG research makes CAR the best option to directly compare results to the literature. Moreover, CAR has been used in studies targeting the N2pc through RSVP paradigms [19] and for this reason was used here as well.

4.2.2.4 Epoching the data through edge detection on the LDR channel

In order to segment the EEG signal, an edge detection algorithm ran through the whole LDR signal and based on defined ranges, mapped each instance to a target or non-target presentation period. However, the signal requires some cleaning so that no erroneous spontaneous edges that emerge throughout the protocol visualization are detected, mainly during resting periods when subjects may be moving and so may be the system cables. In doing so, it must be ensured that the filter used does not interfere with the edges themselves so as to not distort synchronization. In Figure 4.18, a comparison between the distortion effect of different filters was compared, namely two moving average filters with two distinct signal lengths, and two moving median filters with the same separate lengths.

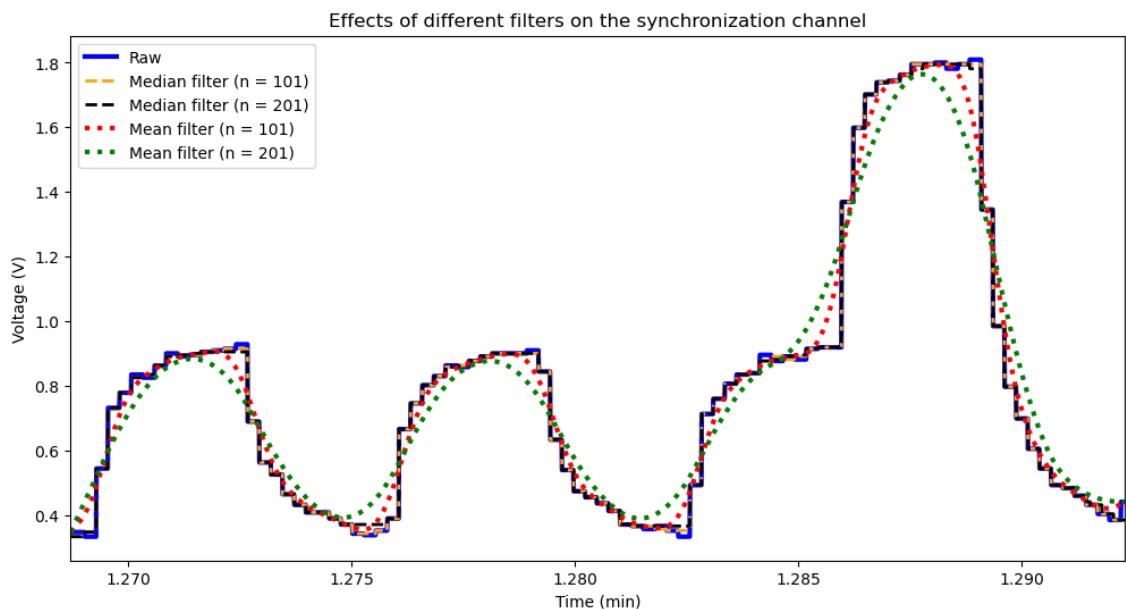


Figure 4.18: Comparison between the different filtering approaches to the LDR channel.

As shown in Figure 4.18, both moving average filters distorted the signal and changed the onset of its edges. In comparison, for the same filter length, the median average maintained the edges onset while eliminating spontaneous peaks. This emerges from the fact that the median is a more robust metric when compared to the average. Moreover, both signal lengths on the moving medians performed similarly in not affecting the phase with the distinction that the wider filter attenuated more significantly the transient peaks during resting periods on the paradigm. This improved the edge detection algorithm by preventing non-relevant edges from being found, which is more relevant for the lower ranges defined.

Regarding its inner workings, the algorithm initializes variables that intermittently toggle rising and falling edge detection preventing two consecutive slopes of the same type from being detected. These variables were defined as *lastlow* and *lasthigh*, initialized to *True* and *False*, respectively, to prioritize rising edge identification. The LDR signal is then iterated for all its values and when a value crosses from below to above the threshold, a rising edge is detected. The first

rising edge of each **Training** protocol corresponds to the onset of EMG recording whereas for the **Testing** session, the first rising edge marks the beginning of the first trial block. Once discovered, variables *lastlow* and *lasthigh* switch their binary value to *False* and *True*, indicating the edge detection algorithm to find falling edges, defined as when the signal value crosses from above to below the same threshold.

The threshold values were established based on their intersection with an amplitude equidistant between the minimum and maximum amplitudes of each class on the LDR channel. This ensures that the Target and Non-Target instances will have similar phases. A threshold of 1.4 V was established for the target instances while a 0.66 V defined the Non-Target threshold. To segment the SSVEP periods, a threshold of 0.2 V was used, meaning the only rising edges found correspond to the instance of a trial block start and the falling edges to its end. Figure 4.19 illustrates these thresholds intersecting the LDR channel classes segmented.

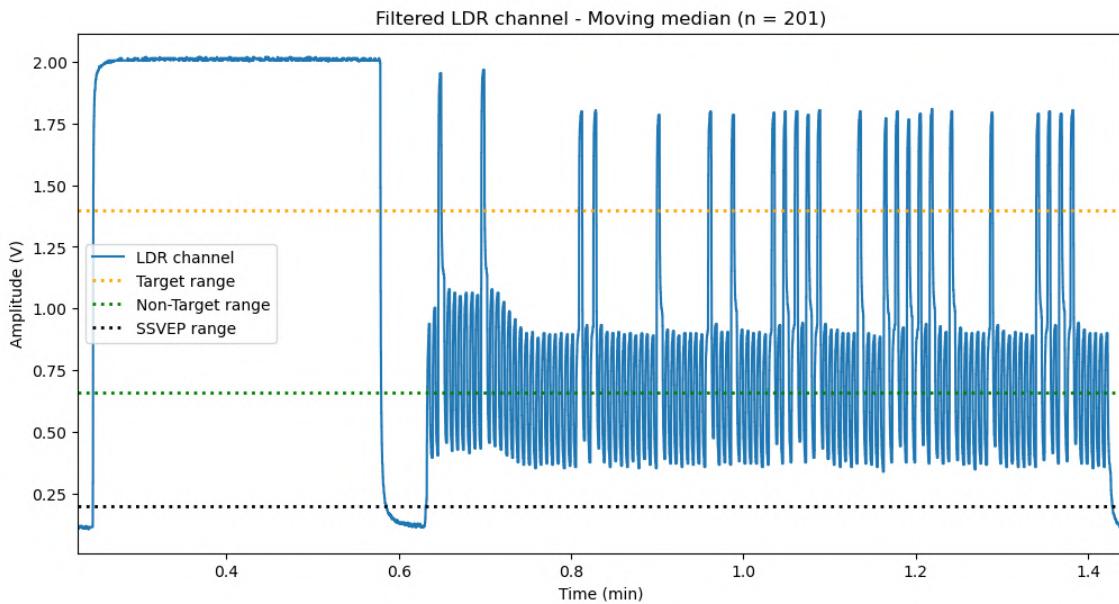


Figure 4.19: The defined thresholds for the target (orange), non-target (green), and SSVEP time windows, intersecting the filtered LDR channel. Notice the equidistant characteristic of each threshold relative to Target and Non-Target stimuli instances.

Finally, the edges marked for the Target and Non-Target classes of stimuli were used to create the ERP analysis period by correlating the sampling frequency with the samples found, established between -200 ms and 600 ms relative to stimuli onset. The obtained period mimics the epoch size on similar studies, with each trial comprising 1640 samples [19]. This period encompasses the N2pc/N2pcb and P3 elicitation periods [58]. The negative temporal offset (-200 ms) is required to perform a standard pre-processing technique for ERP analysis, detailed in subsection 4.2.2.6.

4.2.2.5 Independent Component Analysis (ICA)

ICA is defined as a blind source separation technique that is able to untie independent sources linearly mixed in several sensors, transforming a multivariate signal into its additive subcomponents [112]. It's categorized as a blind source separation because no other information is used to achieve this other than the sensor's information.

Its intuition can be driven by placing several microphones spatially distributed in a noisy room containing several mixed speech sources, commonly known as the "cocktail party problem" in auditory signal processing [112]. Sound proves to be a good analogy to brain data in the context of ICA application since the different sources of signals add linearly, an important assumption for the algorithm application [112]. This technique can decompose the combined speech noise into several signal components with the most influence in each microphone. ICA will make it possible to understand what is being said on each microphone from the indiscernible noise captured by all microphones, as illustrated in Figure 4.20. Analogously applying this rationale to the EEG, electrodes can be thought of as microphones picking different signal sources distributed through the scalp that can ultimately be separated and removed from the signal of interest, namely common artifacts such as frontal dominant ocular movements (EOG) and the most lateral temporoparietal (TP7 and TP8) dominant neck movements (EMG). Through volume conduction, these are propagated to adjacent channels, smearing the brain data and distorting the analysis, which ICA helps eliminate mostly without disturbing the neural signal.

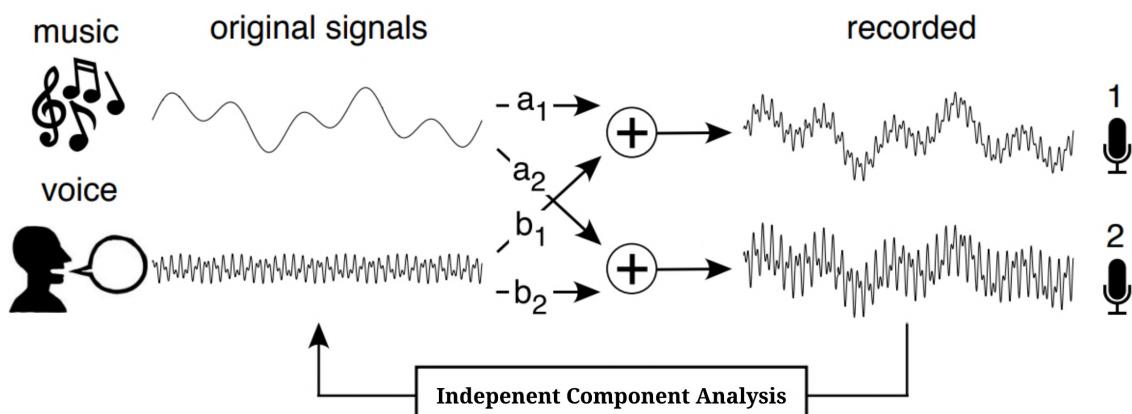


Figure 4.20: Representation of the unmixing ability of ICA to decompose signals from different sources linearly summed and captured by two microphones. The linear weights for each microphone (a_1, b_1 and a_2, b_2), derived solely by the sensor data, reflect the proximity of each speaker to the respective microphones. By unmixing the signals captured with these weights, the original sources can be obtained [112].

The first step in ICA-based algorithms is whitening (sphere) the data by removing signal dependencies. ICA algorithms assume the data has independent and identically distributed sources. Whitening helps achieve this by transforming the data to have zero mean and unit variance. This normalization makes the ICA algorithm's calculations more efficient and reliable. A geometrical interpretation is that it restores the initial “shape” of the data and that then ICA must only rotate the

resulting matrix. The whitening matrix ($W^{C \times C}$) represents the result of an eigenvalue decomposition of the covariance matrix of the observed data $X^{C \times T}$, where C depicts the number of channels and T the number of time samples. Multiplying these matrices results in the whitened data matrix Z with 0 mean and unit variance with the same dimensionality as X , as equated below (4.18).

$$Z = WX \quad (4.18)$$

ICA then performs the rotation by minimizing the Gaussianity of the data projected on both axes, each corresponding to a different signal. By rotating the axis and minimizing the Gaussianity of the projection, ICA is able to recover the original statistically independent sources [112]. This process can be visualized in Figure 4.21 for a whitened data distribution comprising two linearly mixed signals with a Gaussian distribution on both axes (blue bins) (A), and the same data after applying ICA minimizing their Gaussianity (B).

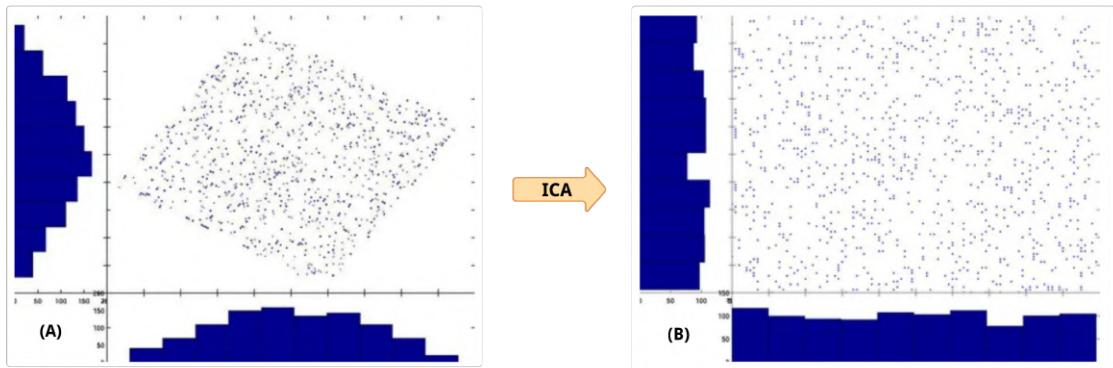


Figure 4.21: Minimization of the Gaussianity in two mixed signals through ICA to recover their statistically independent sources [112]. Image (A) depicts the data distribution of both signals projected on each axe before ICA and image (B) represents the transformed data with ICA.

Mathematically, after finding the unmixing matrix U that contains the separability of these independent sources through optimization criteria and iterative processes too complex to fully describe here, the cleaned EEG data regarding the number of components defined to remove is given by equation 4.19, with S representing the resulting unmixed signal.

$$S = UZ \quad (4.19)$$

An ICA solution was found for each participant and each component produced was analyzed individually to determine its removal since brain data can sometimes crossover to these components. If simply removed, these can eliminate real brain data from the signal. To decrease the algorithm running time, a period between the EMG recording start and the last stimuli appearance on the EEG was used for the **Training** session, with the EMG recording being used to target the most relevant noise source for the EEG captured in this project, as a result of no frontal channels being used.

In practice, the ICA algorithm applied by *MNE* consists of a more powerful version (ICALabel) of the standard ICA insofar as integrating a classifier model that predicts the likelihood of that artifact source, namely EMG, EOG, heartbeat, line noise, and channel noise [113]. It also estimates the probability of that source being mostly brain data, significantly speeding the analysis of the components. Because of this, some specifications must be met to ensure a good independent component's (ICS) determination and subsequent classification. These consist of applying a band-pass filter between 1 and 100 Hz and re-referencing the data with the CAR [113]. A copy of the raw signal was pre-processed this way and used to estimate the components. If one of these was marked to be removed, its elimination occurred in the low-pass filtered at 28Hz, bad channel interpolated and re-referenced version of the data.

The algorithm then ran with the maximum number of components possible indicated by the number of channels ($C = 33$) for a more accurate solution. The ICs found are then ranked based on their explained variance in the signal, meaning the more meaningful source components are displayed first, with these affecting the signal the most. Following this notion, out of the 33 components estimated, only the first 11 were carefully analyzed.

4.2.2.6 Baseline correction

When analyzing and plotting the ERPs, a factor has to be weighted regarding the trend of activity that is built up during the trial presentation. Because the brain activity is responding to closely presented stimuli in different situations (*e.g.*, target vs non-target), the average activity after the different stimuli types are expected to be different and so is the activity before their onset.

In order to visualize differences between the classes of ERPs being studied, the period before stimuli onset has to be factored in and normalized, ensuring a more fair and statistically significant comparison between the peaks that arise after stimuli onset. This technique represents baseline normalization, in which these differences in the DC components present right before image onset between the different classes of stimuli are removed. These periods typically revolve around a 100 ms or 200 ms time window in which the signal average is subtracted from the whole signal and is centered on 0 V. Once again, the methodology applied to the signals in the online dataset was used, describing a time window between -200 and 0 ms relative to stimuli onset to compute the baseline [19]. The average signal voltage in this period was then removed from the entire ERP period, normalizing the signals. This process is shown in image 4.22.

4.2.3 Event-Related Potential analysis

With the EEG free of artifacts and the epochs properly segmented and normalized, the ERPs were computed and analyzed according to the literature guidelines. Despite both being ERPs, there are some marking differences in the methods used to both obtain and analyze the N2pc/N2pcb group and the P3. These are unraveled below.

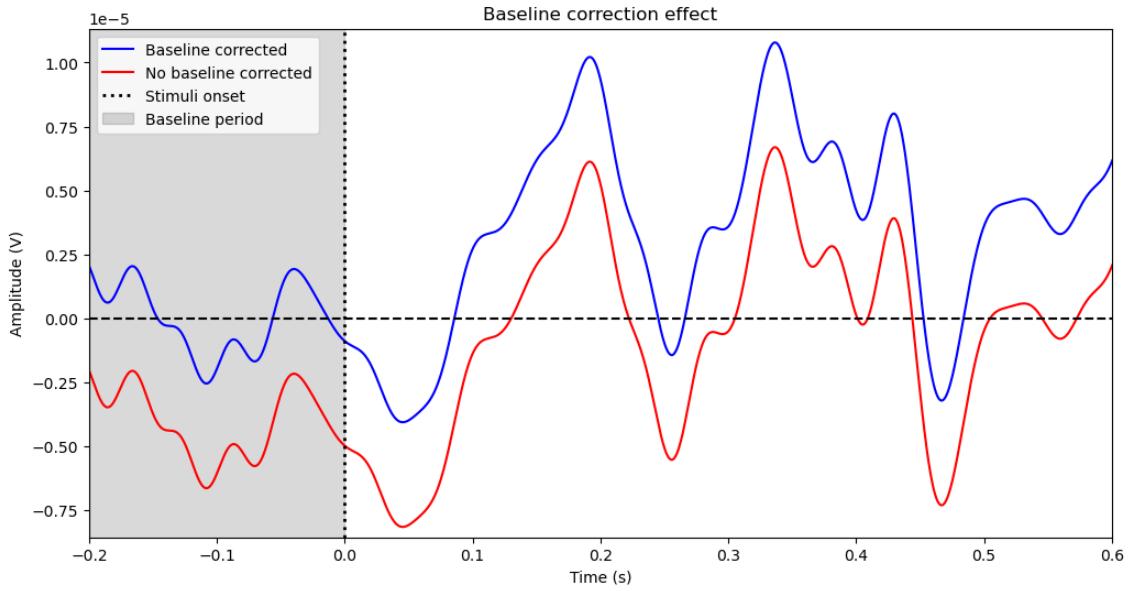


Figure 4.22: The effect of baseline normalization/correction (blue) on the same non-normalized single trial at channel CPz (red). In the baseline period (-200 to 0 ms), the blue signal has a mean of 0, indicative of this normalization.

4.2.3.1 Computing the ERPs

Obtaining the P3 waveforms is straightforward, *i.e.*, they are computed by collapsing/averaging the epoched signals across the trial dimension over the same channel for each class, as indicated in Figure 4.23. Posteriorly, the grand averages were obtained, referring to the average activity across all subjects, as is standardly done in ERP analysis [58]. This was done for the P3 by separately averaging all epochs to both Target and Non-Target trials in the Training and Testing image presentation sessions. For the *Training* dataset, this corresponds to 240 Target signals and 2160 Non-Target signals averaged separately, and 128 Target signals and 1472 also averaged separately.

In contrast to the P3 component, the N2pc/N2pcb require additional computations before obtaining the grad averages. These are necessary to overcome the limited discriminability emerging from their reduced size. For the N2pc, the ipsilateral activity was subtracted from the contralateral activity depending on the lateral region of the human target appearance, namely the Right Visual Field (RVF) or Left Visual Field (LVF) [19]. For instance, for RVF targets, the N2pc is obtained by subtracting channel activity contained on the right parieto-occipital scalp (ipsilateral) from the equivalent region on the left hemisphere (contralateral) and vice-versa for the LVF stimuli.

Regarding the N2pcb, its computation comprises the difference between the average bilateral activity (both scalp hemispheres) elicited by a target appearing either on the Top Visual Field (TVF) or the Bottom Visual Field (BVF) and the average ipsilateral activity that emerges from all lateral targets combined (RVF and LVF) in the same N2pc scalp regions, as defined on the literature [68].

The P3 scalp ROI comprised the most centroparietal portion of the scalp, defined as the activity evoked and captured in channels Cz, CPz, Pz, CP1, and CP2, indicated by the red circled

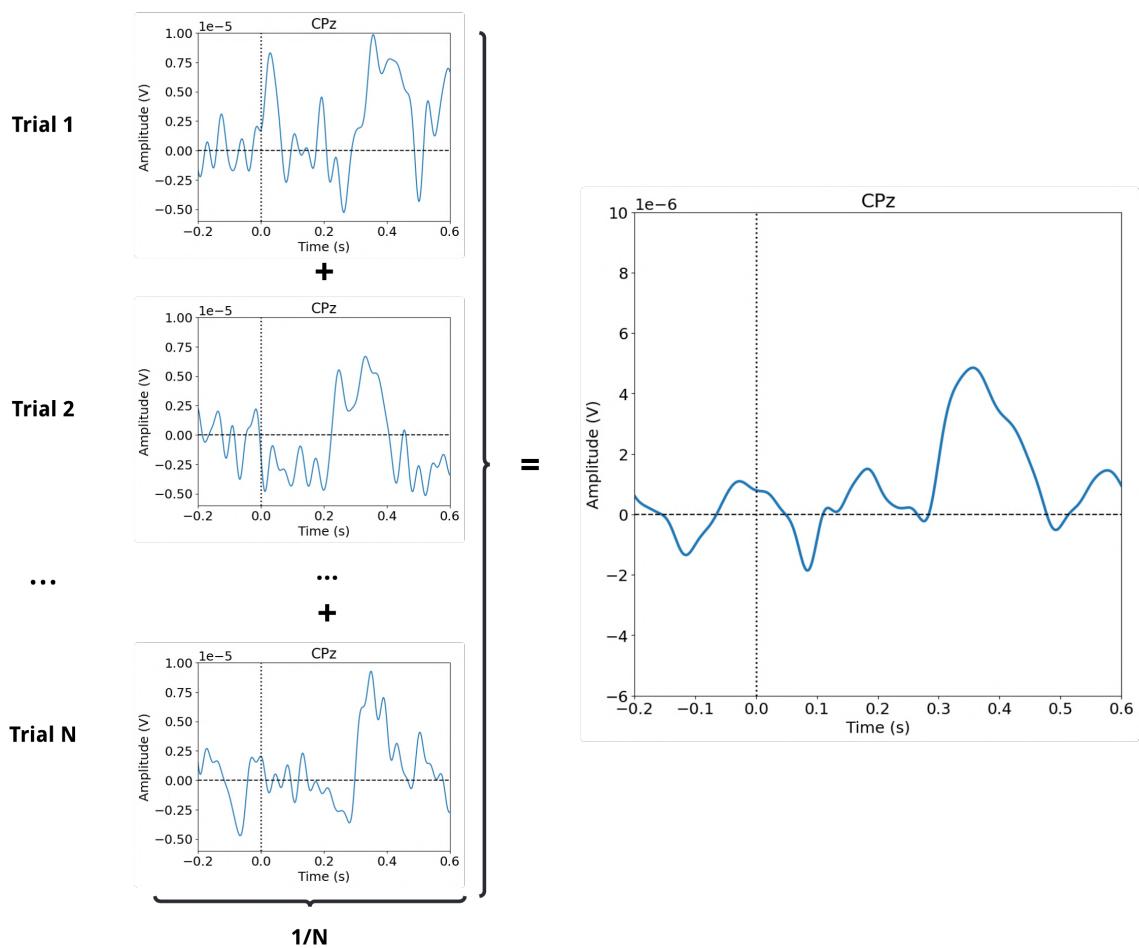


Figure 4.23: Illustration of ERP extraction method. Individual trials (left) are averaged together for each electrode individually, resulting in a waveform comprising only the target stimuli processing while eliminating all spontaneous and variable neural activity associated with individual responses (right).

electrodes in Figure 4.24. As for the N2pc/N2pcb group, channels P7, PO7, PO3, PO5, P3, P5, and O1 on the left hemisphere and their analogous position on the right hemisphere, specifically the P8, PO8, PO4, PO6, P4, P6, and O2, were used, as depicted by the green circled electrodes in Figure 4.24. Exemplifying a single-channel N2pc trial calculation, for an RVF target, the epoched activity from channel P08 (ipsilateral) was subtracted from the epoched activity in channel P07 (contralateral). Approaching the same trial and channels for a TVF N2pcb, its calculation comprises the subtraction of P07 and P08 activity combined ($P08 + P07$) with the ipsilateral activity averaged across all trials for lateral targets over the same P08 and P07 channels, obtained during the N2pc computations. From this, the grand averages for the N2pc and N2pcb were obtained.

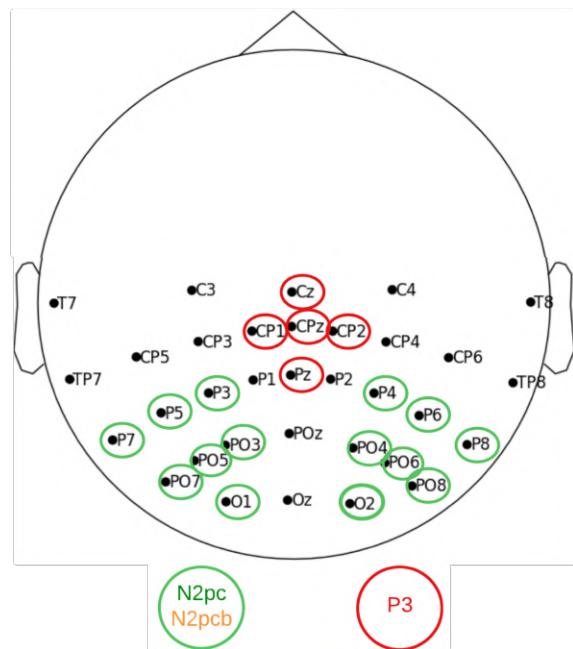


Figure 4.24: The regions of interest targeted for the N2pc/N2pcb (green) and P3 (red) ERP components. The color scheme defined here will be maintained for all figures depicting these ERPs. In other words, N2pc and P3 plots will always be represented by the green and red color schemes throughout the document, respectively. As for the N2pcb, it will always be plotted referencing the orange color.

Once obtained, the temporal emergence of these ERP components was analyzed to select their most discriminatory time windows within the -200 to 600 ms ERP epoch. The inter-subject average period encompassing the N2pc/N2pcb was defined as the window between 150 ms and 350 ms from image onset and 280-480 ms for the P3. A summary of the concepts discussed here to obtain the ERP waveforms are schematized in diagram of Figure 4.25.

4.2.3.2 Visual Angle and the N2pc/N2pcb

The visual angle corresponds to the entire area of vision captured by the eyes and perceived by the brain, both horizontally and vertically [114]. Research states that the vertical dimension of reduced ability in comparison to the horizontal visual field, where the former is able to capture

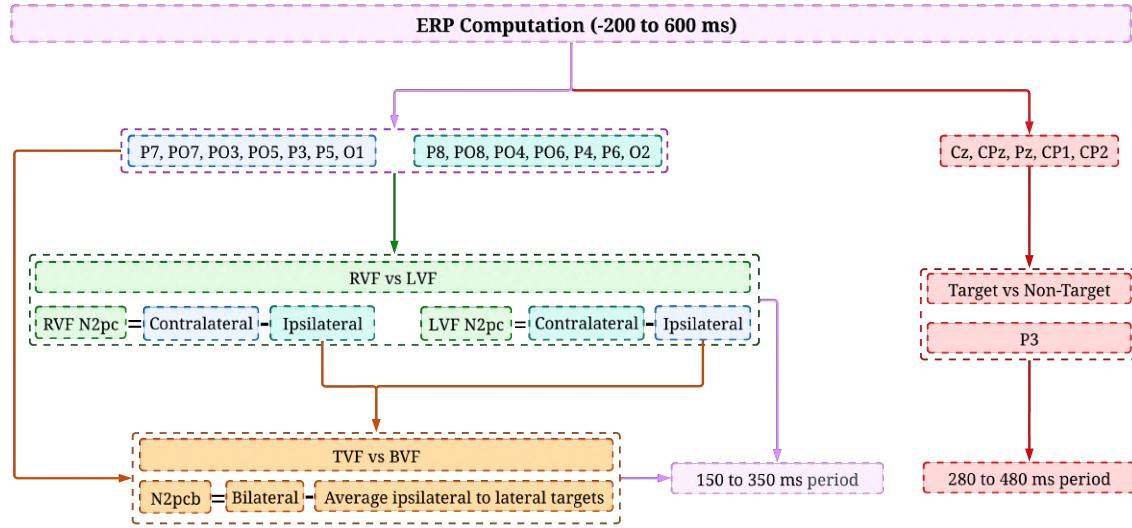


Figure 4.25: Summary of the conceptual computations required to obtain the N2pc/N2pcb ERP group through electrode pair subtractions. The contralateral and ipsilateral components of the N2pc change depending on RVF or LVF appearance, as indicated by the color code of each electrode hemispheric region. Left hemisphere channels are depicted in light blue and right hemisphere ones in light green. On the N2pcb, bilateral activity refers to the sum of electrode hemisphere pairs. Both channel and temporal parameters for all ERP analyzed are displayed.

130° of the environment while the latter captures 180° [115]. These definitions are important in the context of this project to define the range limits of the image presented on the computer screen and to establish the angle thresholds that characterize the distinction between the RVF and LVF stimuli, horizontally, and between the TVF and BVF vertically.

When targeting the N2pc ERP, its definition assumes that there are hemispherical differences in response to lateral targets. If a target appears laterally very near the vertical midline within the image, the bilateral channel difference is not significant and no N2pc is produced [19]. The same rationale applies to the vertical targets close to the horizontal midline and the N2pcb [68]. For this reason, a visual angle threshold must be used to overcome targets that overlap these regions near the center of the image.

The visual angle α , is the angle that sends light from scene s through the lens onto the surface of the retina, with a distance d from lens to scene, as shown in Figure 4.26. Its calculation follows a simple trigonometric approach that correlates these parameters, as expressed in equation 4.20. However, some transformations were required to properly calculate the visual angle of the images on-screen through this equation. Firstly, visual stimuli were presented through a 27" (inch) full HD (1920 x 1080 px) screen. In order for the expression to work, a unit conversion regarding the image size in pixels to its millimeters occupying the screen was necessary. For this, the screen dimensions both in pixels and millimeters were used. Posteriorly, equation 4.20 was applied and the number of degrees that each pixel on the image represents was obtained by dividing the image width in degrees of visual angle by the image width in pixels. This unit essentially represents the resolution of the image on-screen in visual angle per pixel format ($\frac{\text{deg}}{\text{px}}$). Finally, the visual angle

of each target object in the images was calculated by determining the distance of its centroid from the image center in pixels and scaling it with the visual angle resolution ($\frac{\text{deg}}{\text{px}}$). This was done when loading the ordered centroid data saved during the *Psychopy* sessions and matching the order of targets with the LDR channel segmented ERP periods, matching the centroid occurrence with the corresponding trial. The visual angle determination was done parallelly for the horizontal and vertical image dimensions.

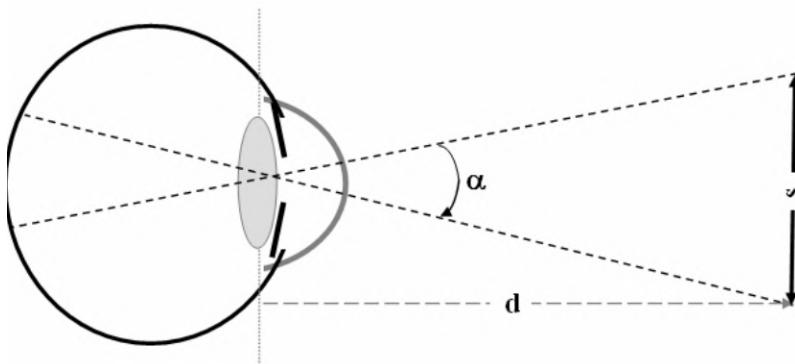


Figure 4.26: The visual angle conceptualization [114].

$$\alpha = 2 \arctan\left(\frac{s}{2d}\right) \quad (4.20)$$

The visual angle calculation also fully contextualizes the final image dimensions (672 x 576) chosen and described in section 4.1.1.3. With this scaling, a 13.94° visual angle contains the image horizontally from left to right and 11.8° vertically from top to bottom. This was done to match the vertical visual angle used here with the horizontal visual angle of the images presented to participants on the literature in which similar analysis was conducted (11.8°) [19]. The horizontal visual angle defined here, however, was pushed a bit more than 1° in total to see if the increased region would impact the performance of the N2pc. The fit of visual angles between the paper and this project was done by not only defining specific image dimensions but also by manipulating the distance of the participants to the screen, which was 800 mm on the paper [19] and 1000 mm here. It is possible to infer that this adjustment resulted from the difference in screen size used here (27") and in the study (24"), even though this information is not provided. The measurements between the participant's shoulders and the screen were done right before the **Training** protocol presentation for each participant.

Once properly calculated for all target images, the visual angle threshold for targets appearing near the image center was established by once more following the paper guideline for this parameter. The referred dataset study defined a horizontal threshold of 1.2° on each side to determine if a target object is considered LVF or RVF [19]. Putting it differently, if the visual angle of a target object is less than -1.2°, the target object is classified as LVF, whereas if the visual angle of a target object is more than 1.2°, the target object is RVF. Here, the same threshold (1.2°) was applied for both horizontal and vertical dimensions.

The removal of this region of image space results in a different number of trials to average regarding the RVF, LVF, TVF, and BVF conditions in comparison to the P3 Target vs Non-Target. The targets centroids tracked can be seen distributed for the **Training** (A) and **Testing** (B) protocol sessions in Figure 4.27. The visual angle thresholds are also plotted with their corresponding value in pixels. The human target stimuli centroid distribution had to be slightly manipulated prior to EEG recording by rotating 180° a copy of some images and eliminating others to balance the classes of centroid distribution. For the **Training** protocol, 100 trials well distributed across each of the 4 regions of visual angle stated above were computed, while 72 were used for the **Testing** dataset separately.

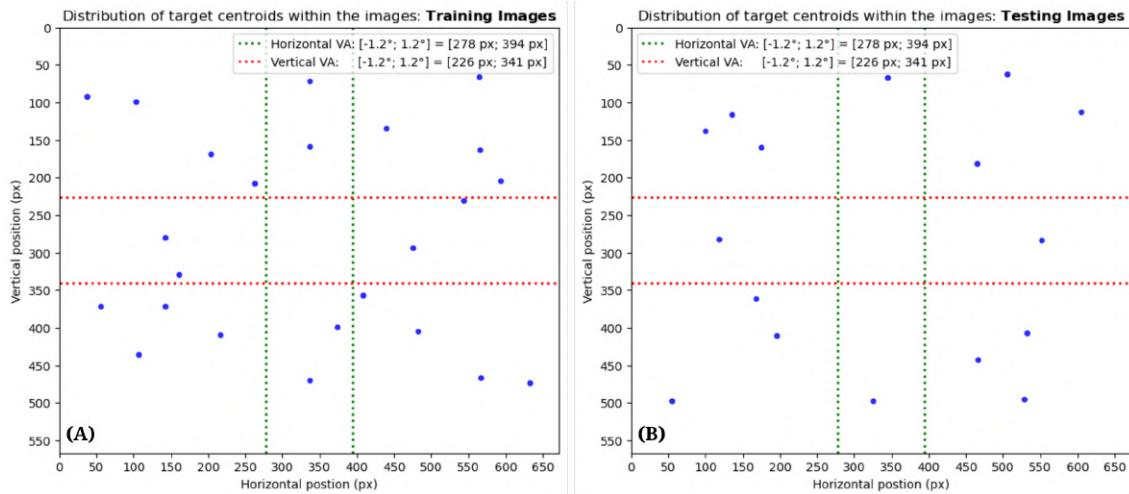


Figure 4.27: Distribution of target centroids within the images presented during the **Training** (A) and **Testing** (B) sessions. The lateral (green) and vertical (red) 1.2° visual angle (VA) thresholds are also exhibited. Notice that the vertical centroid position starts (0 px) from the top as was registered during the centroid tracking algorithm used when preparing stimuli for the *PsychoPy* sessions.

4.2.4 BCI actuation structure: Feature vector composition and class discrimination

After defining the main signal characteristics, namely the ERPs, the single-trial data had to be leveraged to construct models that could actuate specific outputs based on the classes of stimuli. The classification tasks were grouped into six main groups. Following the chronological order of brain processes and thus starting with the N2pc/N2pcb group, these were used to handle the 4 main regions of centroid appearance discrimination. The N2pc offered the binary lateral (RVF vs LVF) region discrimination and the N2pcb the vertical (TVF vs BVF) counterpart in parallel. Additionally, an N2pc-based model was tested to attempt 4 quadrant classifications (RVF vs LVF vs TVF vs BVF).

Following the region-specific classification tasks, models were also built to provide discriminatory outputs between target and non-target stimuli. Such discrimination was approached in three

different ways, specifically by using the N2pc and P3 ERPs both individually and together in three separate model types.

The models described so far all share the characteristic of being subject-specific, meaning that each discrimination task was treated individually and optimized for each participant, resulting in 6 models per classification group. Such design was done to escape the BCI illiteracy challenge mentioned in section 2.5. This problem is especially relevant and documented regarding MI-based BCIs but also emerges in SSVEP and ERP systems, with studies targeting all 3 BCI approaches for the same subjects suggesting a rate of BCI illiteracy for MI paradigms of 53.7%, 11.1 % for ERP systems, and 10.2 % for SSVEPs [53]. Albeit with limitations, some studies have been able to generalize BCI systems to many subjects [22]. With this in mind, some models here were designed to tackle this problem, specifically for Target vs Non-Target discrimination by using all subject data combined during training. Therefore, a Target vs Non-Target subject-generalized model was built solely using the N2pc, another using the P3, and a third one combining both N2pc and P3 visual processing information. This means that three additional subject-generalized models were evaluated on top of the six subject-specific classification groups.

The feature vectors composed for each classification condition were similarly obtained but were dependent on the ERP used. For the N2pc/N2pcb group, the 7 channel pairs described previously and illustrated in Figure 4.25 were used. However, here a challenge arises. This is because the target localization information cannot be assumed during real-time scenarios to correctly apply the N2pc channel pair difference convention. Strangely, this is assumed on the literature and the contralateral minus ipsilateral calculation is done priorly to classification depending on the lateral appearance of the target, both during training and testing, in studies motivating the discriminatory ability of the N2pc for RVF and LVF targets in BCIs [19]. This introduces a bias that cannot be replicated in real-time cases where this target laterality cannot be determined beforehand. Additionally, other studies conveniently leave the information regarding channel difference computation ambiguous for the N2pc during classification [76]. Here, a way to discriminate the target position (RVF vs LVF) within the image was tried blindly without biasing the N2pc computation. Many variations were tried but using all 14 channels relevant for computing the N2pc without subtracting channel pairs, *i.e.*, 7 on the left hemisphere and 7 on the right, yielded the best results. As for the N2pc models attempting to distinguish target from non-target stimuli, these 14 channels were directly compared between these two stimuli types during classification. As for the N2pcb, this is not an issue since its computation is already intrinsically blind to such a bias and is treated the same way for TVF and BVF stimuli. Therefore, the 7 channel pairs subtraction was performed for forming the N2pcb feature vector as stated in Figure 4.25.

The feature vectors are all composed of the temporal information in which each ERP occurs. The period encapsulating the N2pc/N2pcb (150 to 350 ms) was segmented for each channel pair from the ERP period (-200 to 600 ms) with posterior resampling to 64 Hz, resulting in a 14 sample per-channel signal and corroborating similar methods in the literature [19]. The 28 Hz low-pass filtering done at the beginning of the pre-processing pipeline with a full cut at 30 Hz (-60 dB) ensured that no signal aliasing occurred during the resampling since the *Nyquist theorem*

threshold is respected (32 Hz). Once segmented, all resampled channel pairs were concatenated column-wise, giving rise to 196 features per trial. This was performed for every single trial with these 196 feature samples being concatenated row-wise. In the analogous N2pcb models, this resulted in 98 features per trial. The same rationale was used for the P3-based models, where the time-window segmented was encapsulated between 280 and 480 ms, but because only 5 channels were used, a single trial only contains 70 features. As for the N2pc and P3 Target vs Non-target model, the 14 N2pc channels and P3 electrode information were combined with a single trial displaying 266 features.

As opposed to the models attempting to discriminate visual angle regions, the Target vs Non-Target conditions are not balanced due to the 10% ratio introduced to correctly elicit the *oddball* effect. To mitigate this, an equal number of Non-Target (240 for the **Training** and 128 for the **Testing** datasets) stimuli were randomly selected from the Non-Target pool to produce balanced results.

With the intention of finding the best possible model for each task, a classifier grid was designed comprising some of the most common classification architectures. The BCI-favorites SVM and Linear Discriminant Analysis were used, but also the Random Forest, XGBoost, and AdaBoost. Each classifier had its own grid of hyper-parameters to tune to, ensuring that not only the best architecture was chosen, but also its best version. These are illustrated in Figure 4.28. The different hyper-parameter combinations for each architecture were compared based on the average 10-fold validation performance metrics.

Some hyper-parameter combinations within the same model are not compatible. These refer to the SVM and LDA models, resulting in having to treat them as separate models entirely within the grid. For instance, the *degree* hyper-parameter is only relevant for the *poly* kernel on the SVM. Once the grid ran through all its iterations, the best architecture was selected based on the best inter-subject average performance (validation accuracy) for the subject-specific models, meaning there is architecture coherence for the same discrimination task within participants, even if some particular subject achieved slightly better performance with another architecture than the averaged best. However, once the architecture was defined, the models were retrained with all training data with the best hyper-parameter combination for each participant within the selected architecture. As for the subject-generalized models, since there is only one possible best model, the best architecture and hyper-parameter combination were directly selected.

Each architecture and hyper-parameter combination was approached in two different variations of the feature vectors, with a few exceptions. One of them used all the channel information described above. The other used a method of feature trimming to select the most discriminatory subset of channels for each classification, described previously and denominated as Mutual-Based Information (MBI). Figure 4.29 describes in flowchart format all the feature vector composition and classification pipelines for complete analysis. All models, although different with respect to their decision outputs and feature parameters (ERPs), all share the same feature vector composition pipeline, hence the convergence to the common colored processes in purple. However, notice an asterisk (*) on the right corner of the "All channels" process, indicating that models with this

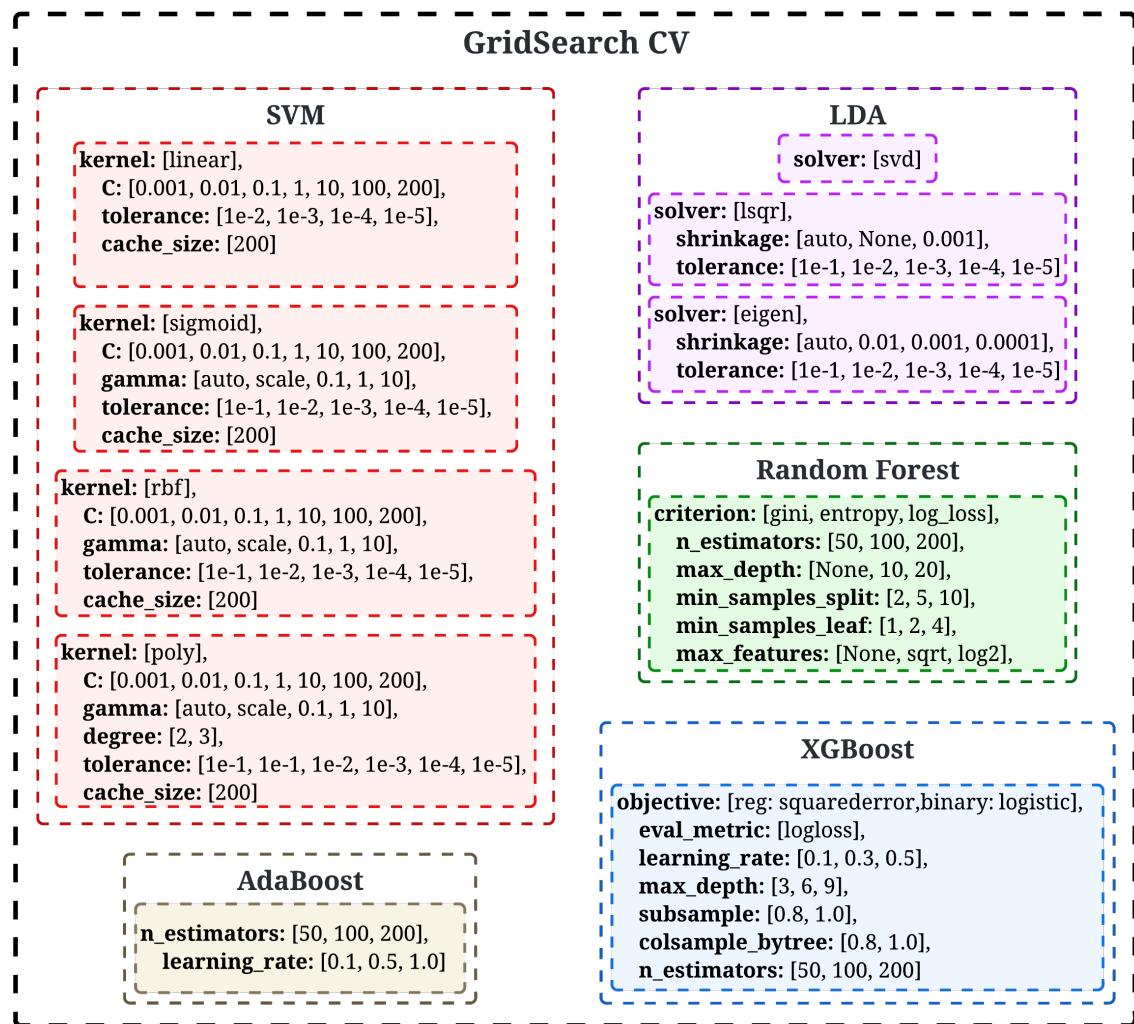


Figure 4.28: The Grid Search CV that encapsulates all hyper-parameter and architecture model combinations.

annotation were solely tested by feeding All channel data and no feature selection (MBI) variation was addressed. These correspond to all Target vs Non-Target generalized models and the subject-specific Target vs Non-Target P3 models. Moreover, the MBI algorithm, when applied, picked the best 7 channels for all single N2pc-based models, 4 for the N2pcb models, and 9 for the N2pc and P3 combined model for target vs non-target classification.

When applying MBI to the feature vectors composed here, it is important to note that the channel selection was not done directly. This is because MBI treats each sample as a single feature without knowing its sample context within the concatenated channel information. To mitigate this, in models where the 7 most discriminative channels were to be selected, MBI was defined to pick the most discriminative 98 samples (*14 samples per channel × 7 channels*). Then, the 7 channels with the most discriminative number of samples were selected and all 14 samples within these were used instead of using the features directly selected by MBI. The same notion was applied for the other models where a different number of optimal channels were selected, with the number of features scaling according to this number.

To validate the models, the trial-wise shuffled **Training** EEG data was used with a 75% Train/Validation trial split in a 10-fold Straight Cross-Validation loop [19]. All models, including the subject-generalized and subject-specific types with both all channel and best channel feature vector variations, were evaluated according to the mean Cross-Validation accuracy score (75%), and both mean validation accuracy and Area Under the Curve (25%) metrics. These three metrics were used to assess and identify the best model for each problem, with more weight in this decision given to the mean validation accuracy. The feature vectors were all Z-score normalized trial-wise before being fed to the learning algorithms by subtracting the trial-specific mean and dividing by the trial-specific standard deviation. This eliminates the problem of larger features dominating the class discriminability and improves learning efficiency.

Once validated and chosen, the best models were retrained with all the trial-shuffled **Training** dataset and their performance was evaluated on the unseen and also shuffled **Testing** EEG data for all trials according to the testing accuracy and AUC. Furthermore, the same metrics were addressed on the trial-ordered **Testing** dataset to track the performance evolution of the brain to unseen images, in an unseen environment with a different target. All pseudo-random selection and shuffling done in every point of the pipeline were specified with a 42 random seed to ensure reproducibility, including the model algorithms that provided a random seed specification.

4.2.5 Horizontal centroid estimation: Using Artificial Neural Networks to describe Biological Neural Network activity

Albeit discriminating regions within the images offer a piece of good complementary information to the Target vs Non-Target classification, extending this approach to pinpoint exactly within the image where the target of interest is located at high speeds poses a more interesting BCI application. For this, an Artificial Neural Network architecture was trained to build a regressor that would estimate a centroid based on the N2pc modulations to different target positions within the image.

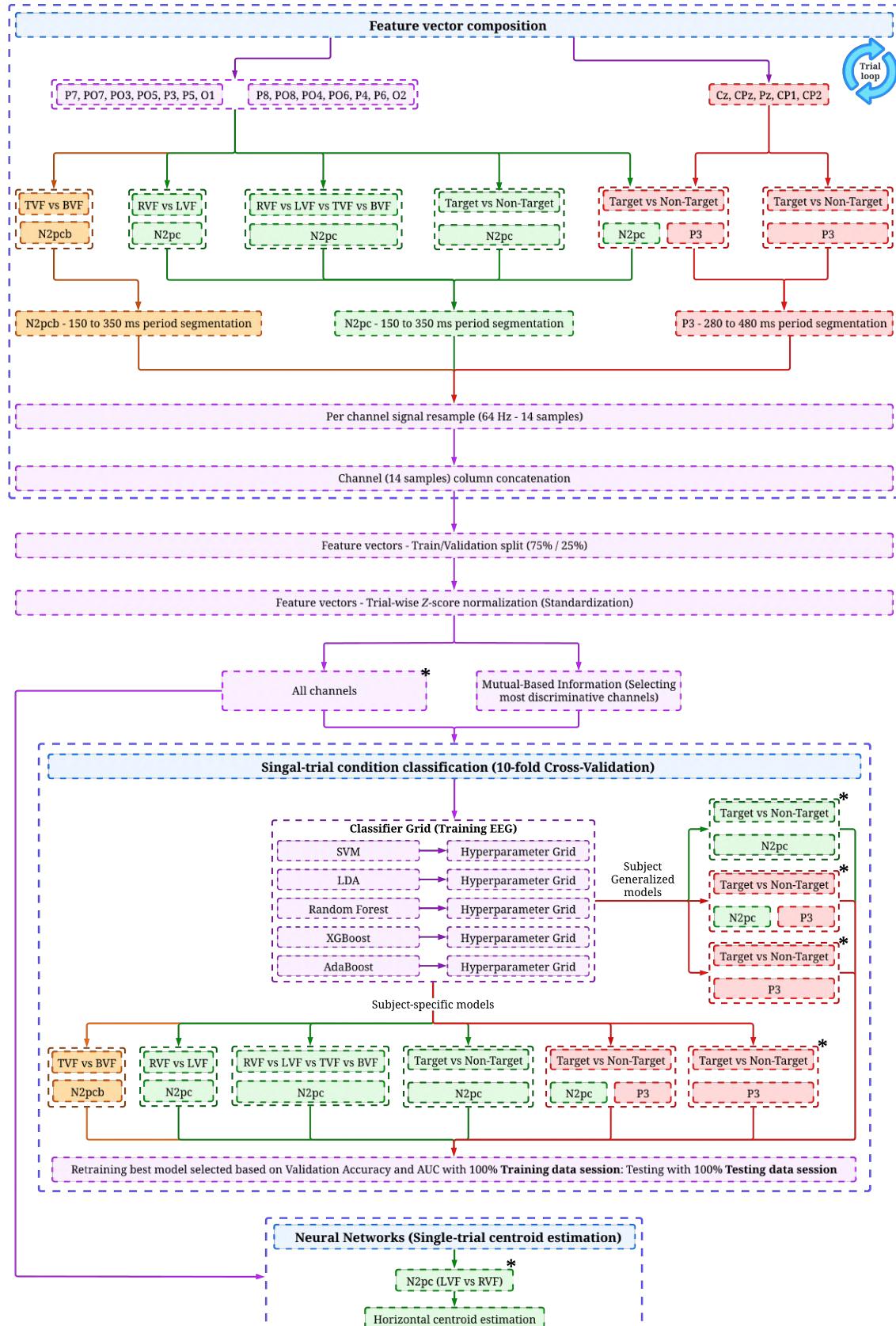


Figure 4.29: Flowchart of the feature vector composition and subsequent classification and centroid estimation. Once again the green color depicts N2pc, orange the N2pcb, and red the P3-based data flow. The purple color represents the data flow common to every model analysis, excluding the channel data pairs at the start of the diagram that are only common to the N2pc/N2pcb group.

A subject-specific approach was used, meaning that 6 neural network architectures with the same hyperparameters were built and trained on the data from each subject. The same Train/Validation split was used (75 % / 25 %) and after validation, the models were retrained with 100 % of the **Training** data session and tested with 100 % of the **Testing** data experiment.

In order to achieve better model convergence and to estimate the true ability of the algorithm to output the target's x-axis centroid within the images, the data were structured with the N2pc bias. This means that it was assumed that RVF and LVF discrimination for each subject outputs a 100 % accuracy, as was done in papers attempting to do the same [19, 76]. However, in contrast to the papers, the N2pc channel pair computation was not performed here and the separate 14 channel information was used instead. In order to bias the N2pc in this manner, for RVF target trials, the 7 contralateral channels (left hemisphere) form the left half of the feature vector and the ipsilateral component the right half. As for the LVF targets, the inverse was done, namely the 7 contralateral channels, this time corresponding to the right hemisphere electrodes, were placed on the left portion of the feature vectors and the ipsilateral channel data on the right.

Regarding the specific NN architecture used here and starting with the input layer, the number of neurons placed here are strictly dependent on the amount of features used. In this regression problem, all 14 channel data relevant to the N2pc were used, composing 196 features. With this, 196 neurons in the input layer were established. Because this is defined as a regression task, only one output is required, *i.e.*, the lateral centroid estimation, and thus, only one output neuron was placed. Because this is supposed to work within a BCI system, these architectures need to compute an output very fast and are expected to do this without the requirement of powerful machines in order to be transversal to less capable systems during deployment. To fit this need, only one hidden layer composed of 32 neurons was used. All these layers were linear and incorporated a sigmoid activation function between each connection to introduce the required non-linearity.

Regarding now the hyperparameters of this Neural Network, a learning rate of 0.005 and a momentum of 0.095 were implemented. The classic Stochastic Gradient Descent was used in combination with the mean absolute error criterion/loss to handle the update of weights and biases within the network during backpropagation, depending on the deviation between the supposed output and the predicted value produced during forward propagation. Finally, with the data from each subject, the ANNs were trained during 30000 epochs with a batch size of 64 trials.

In order to quantify the quality of the regressor prediction, the Pearson Correlation Coefficient (PCC) was used. This represents the correlation coefficient that measures the linear correlation between two sets of data. It is the ratio between the covariance of two variables and the product of their standard deviations and is the same method used in papers that attempt to estimate a centroid value using the N2pc, allowing direct comparison with the results obtained here [19, 76].

4.3 Establishing a target identification performance ground truth with a pre-trained Deep Learning model

Because this project is based on the limitations of AI to aerial imagery, an algorithm of this sort was assessed on the same images to have an idea of the performance differences between both approaches. For this, a pre-trained Deep Learning model from *OpenCV* available on a *github* repository, tuned to find and delineate a skeleton on top of a human body contained in images or even video by analyzing individual frames [116].

This model type is typically nominated as Human Pose Estimation. The model predicts a separate heatmap for each body part (*e.g.*, shoulder, elbow, etc.), where hotter areas indicate a higher likelihood of that body part being there. Based on the heatmaps, the model pinpoints the most likely location (keypoint) for each body part. It then connects these key points based on predefined anatomical connection pairs (*e.g.*, Head to Neck) to create a visual representation of the human skeleton in the image. A key aspect of this model is that it can be specified to identify multiple bodies or just one. Because the target human is lying down compared to distractors appearing standing up from the aerial viewpoint, the former is the only case where more body parts are visible, and thus, more key points can be more easily identified. Based on this notion, the model was set up to identify the most probable full pose which happens to always be the targeted human. This was tested for the target discrimination.

4.4 Chapter 4 summary

This Chapter defines the materials and methods used to build the BCI system, starting with the aerial images acquired to build the image datasets that enabled the design of the RSVP *oddball paradigm*. The stimuli were captured with a drone at two independent environment types simulating search and rescue scenarios, resulting in two distinct image datasets. In addition to the environment, the Target human is different between both datasets. One of these datasets, comprised of 65 raw images, was used to Train/Validate the models with the produced EEG responses (**Training paradigm**), and the other was used to Test them (**Testing paradigm**), comprising 61 raw images in total.

Before being presented to participants, these raw images required some preparation. In order to restrict the visual information, the raw images were cropped, with images containing a human lying down within the environment being defined as Target and images missing this characteristic being defined as Non-Target. From this process resulted 313 sub-images in the **Training** paradigm and 261 in the **Testing** paradigm. Posteriorly, these were labeled and listed in an *Excel* sheet for each experiment according to their directory path, condition (Target vs Non-Target), and if the human is present, its centroid position along both image dimensions.

Posteriorly, the visual paradigm was developed using *PsychoPy software* attending to the experimental parameter specifications of each ERP targeted (N2pc/N2pcb and P3). The participants

were instructed to visually search the human target and mentally count each instance of its appearance to maintain focus. In both paradigms, a proportion 10% of Targets to Non-Targets images was defined, with 240 and 160 images being shown in the **Training** and **Testing** experiments at 5Hz (5 images per second).

In total, 6 subjects participated in the experiment with all signing a written consent approved by the institutional ethics committee. For the EEG acquisition, 33 electrodes covering the centroparietal, occipital, and temporal regions of the scalp were recorded, with the system's ground and reference being placed at each earlobe and all impedances being kept under $10\text{ K}\Omega$. One computer was used to present the visual paradigm and the other to acquire the data at 2048Hz with the usage of the Micromed SD LTM PLUS 64 amplifier.

To synchronize the EEG and map each image instance to the individual EEG response it represents, an LDR-based circuit was developed. The idea is that by placing the LDR integrated into a voltage divider circuit facing the monitor and placing a differently colored square associated with each stimuli type (Target vs Non-Target), the output voltage of this circuit will be different for each stimuli type, enabling the segmentation of the continuous EEG with an edge detection algorithm.

Regarding the EEG processing pipeline, it begins with low-pass filtering of the individual EEG files at 28 Hz. Then, a signal quality analysis is performed for each individual, and additional pre-processing steps are performed based on the signal quality estimation. These comprise bad channel interpolation, which occurred in some subjects for channel O1, electrically bridges channel interpolation, evaluation of ICA components to remove Muscle artifacts, and even the component that the constant rate of image presentation (5Hz) produces in the signals in the occipital channels. This component is denominated as the SSVEP which was not possible to untangle from the EEG using ICA without damaging the image brain target responses. Then, the Common Average Reference (CAR) was applied, and the individual EEG responses were segmented between 200 ms pre-stimulus and 600 ms post-stimulus (edge detection algorithm), with this period comprising the ERP epochs and with the 200 ms pre-stimulus being used to posteriorly baseline correct the individual epochs.

Afterward, the individual responses were averaged based on certain stimuli conditions. The lateral targets were averaged and used to obtain the N2pc component (LVF vs RVF), the vertical targets were used to obtain the N2pcb component (TVF vs BVF), and the Target vs Non-Target averaged waveform was used to access the P3 potential. These averages (ERP) were used to pinpoint the average period across all trials where these responses emerged so that these time windows could be used to segment the individual responses in each condition to train and test the models. For the N2pc/N2pcb this period comprised 150ms to 350ms relative stimuli onset and 280ms to 480ms for the P3.

With regards to the classification structure, the N2pc was used to discriminate lateral Target conditions (RVF vs LVF) and the N2pcb to discriminate vertical conditions (TVF vs BVF). The N2pc was also used to discriminate between 4 quadrants, meaning between the horizontal and vertical dimensions simultaneously. These discriminations were subject-specific. In other words,

a model was trained specifically with data from each participant and tested with data from that same participant.

Then, for the Target vs Non-Target discrimination, both the N2pc and P3 were used, both individually and combined. These models were designed in a subject-specific manner but also in a subject-generalized way, where data from all participants was combined to construct a single discrimination model.

The data used to train these models was the segmented period in specific channel locations associated with each ERP and downsampled to 64 Hz. Then in each discriminative task, a 10-fold Cross Validation was used where data entered a Classifier grid with 5 architectures, *i.e.*, SVM, LDA, Random Forest, XGBoost, and AdaBoost, each with their own classifier grid for hyperparameter optimization.

In addition to the discrimination tasks described, a Neural Network was designed to act as a regressor to estimate the centroid of Human Targets along the horizontal dimension of the images. Both the discriminative task models that enter the classification grid and the regressor model were built using the same data split. Firstly, the EEG single-trial responses generated during the **Training** session image presentation were split into a 75% portion to train the models and validated on the remaining 25% according to the AUC and Accuracy metrics. Based on these metrics the optimal architecture and hyperparameter combination for each discrimination was selected. Following this, the best model was retrained with all **Training** EEG single-trial responses and saved. Then, these models were evaluated with all **Testing** session EEG single-trial responses and the sensitivity and specificity metrics were extracted in addition to the accuracy and AUC.

The number of trials for each discrimination task (100% single-trials in **Training** experiment) is:

- Lateral (N2pc) and Vertical (N2pcb) binary discrimination - 100 Target trials in each condition
- 4 Quadrant (N2pc) - 40 Target trials in each condition
- Target vs Non-Target - 240 trials in each condition

The number of trials for each discrimination task (100% single-trials in **Testing** experiment) is:

- Lateral (N2pc) and Vertical (N2pcb) binary discrimination - 56 Target trials in each condition
- 4 Quadrant (N2pc) - 24 Target trials in each condition
- Target vs Non-Target - 128 trials in each condition

Finally, a classic Computer vision Model (Human Pose Estimation) was used to compare its performance to the BCI proposed here.

Chapter 5

Results and Discussion

This chapter aims to present and discuss all the results achieved through the realization of this project. A first look at the PCB and the 3D printed case designed to encapsulate the synchronization circuit of the EEG will be given in section 5.1. Afterward, the EEG analysis and BCI actuation performance will be detailed in sections 5.2 and 5.3.

5.1 The LDR PCB

The LDR results are divided into the hardware design and the LDR signal quality achieved with it.

5.1.1 The hardware design

The heart of the EEG synchronization system is the LDR and its implementation in a voltage divider. With this, the first design decision was the selection of the LDR component to use. Many model variations of the LDR exist with different resistance ranges and properties in response to light intensities. Any one of these would work if properly projected for the second resistor in the voltage divider. However, a model with the hermetic metal package was selected since in comparison to the epoxy sealing it sustains its characteristic behavior in wider environmental conditions, suffering less fluctuations with changes in temperature, humidity, and air pressure. Furthermore, they allow for better heat dissipation and are an overall more robust and durable design. With this in mind, the *PGM-1201-MP* LDR component was selected, offering a resistance range of 4-10 K Ω to 2 M Ω with a light intensity of 10 Lux and in a light absent condition, respectively, as defined on its datasheet [117]. Furthermore, it allows a maximum DC voltage of 250 V and Power of 150 mW. This particular component is shown in Figure 5.1.

Once the LDR component and its resistance range were defined, the voltage supply could be defined. Because the initial approach was to use a rechargeable battery to isolate the signal from the main power line, a voltage regulator was implemented to reduce its power consumption. The LP5907MFX-3.3-NOPB component is a commonly used voltage regulator in portable medical equipment resulting from its ultra low-noise characteristic output and low load transient response,



Figure 5.1: The *PGM1201-MP* LDR component selected for the circuit developed [117].

offering output stability under sudden supply changes. Furthermore, it fits under the Linear and Low-Dropout (LDO) category, meaning it can operate even when the supply voltage is nearly the same as its output [118]. Appropriately supplied, with a maximum input Voltage of 5.5 V, the device sustains a clean 3.3 V output with a maximum current of 250 mA.

With the circuit supply voltage (3.3 V from the LDO) and the LDR resistance ranges defined, the missing resistor can be calculated by plugging these parameters into the voltage divider equation. The minimum resistance value of the LDR ($4 \text{ K}\Omega$) was used for a maximum voltage divider output of 2.4 V, matching the maximum input of the Micromed Amplifier's dedicated DC channel.

$$2.4 \text{ V} = 3.3 \text{ V} \times \frac{R_2}{R_2 + 4 \text{ K}\Omega} \Leftrightarrow R_2 = 10.67 \text{ K}\Omega \quad (5.1)$$

From this, the minimum output of the voltage divider can be given:

$$V_{min} = 3.3 \text{ V} \times \frac{10.67 \text{ K}\Omega}{10.67 \text{ K}\Omega + 4 \text{ K}\Omega} = 0.005 \text{ V} \quad (5.2)$$

This R_2 value calculated poses a problem, however. Ensuring a $4 \text{ K}\Omega$ minimum resistance is not practical, mainly because the lighting conditions of the monitor where the LDR will sit upon are difficult to set up in a manner that fits this requirement. As a result, a potentiometer was used to fit the place of R_2 . With its variable resistance, the circuit voltage output can be adjusted specifically to the monitor lighting conditions. Guided by the obtained R_2 value, a $20 \text{ K}\Omega$ potentiometer was used, giving more head space for adjustments.

From this, a circuit prototype was built on a breadboard and the noise effect difference between an isolated lithium battery (3.7 V - 1350 mAh) and a Type C connection (5 V) directly plugged into the computer's USB port where the EEG acquisitions took place was addressed. The oscilloscope measurements are depicted in Figure 5.2 and show the Peak-to-Peak Voltage (V_{PP}) and the main frequency of the signals. The voltage divider output produced in Figure 5.2 by placing the LDR on-screen with a white background (1st image row) and adjusted through the potentiometer to generate a maximum voltage of approximately 2.4 V. Then, with these adjustments, a second measurement was obtained for a dark background (2nd image row), essentially representing the voltage divider upper and lower limit on-screen, respectively.

It can be seen in column (A) that the noise captured contains the same frequency component of 100 Hz, representing a second harmonic of the power line noise (50 Hz). Furthermore, in the upper range measured (1st image row), this frequency produces more energy than in the lower range (2nd image row) as expected. The upper range generates a $V_{PP} \approx 100 \text{ mV}$ while a $V_{PP} \approx 75 \text{ mV}$

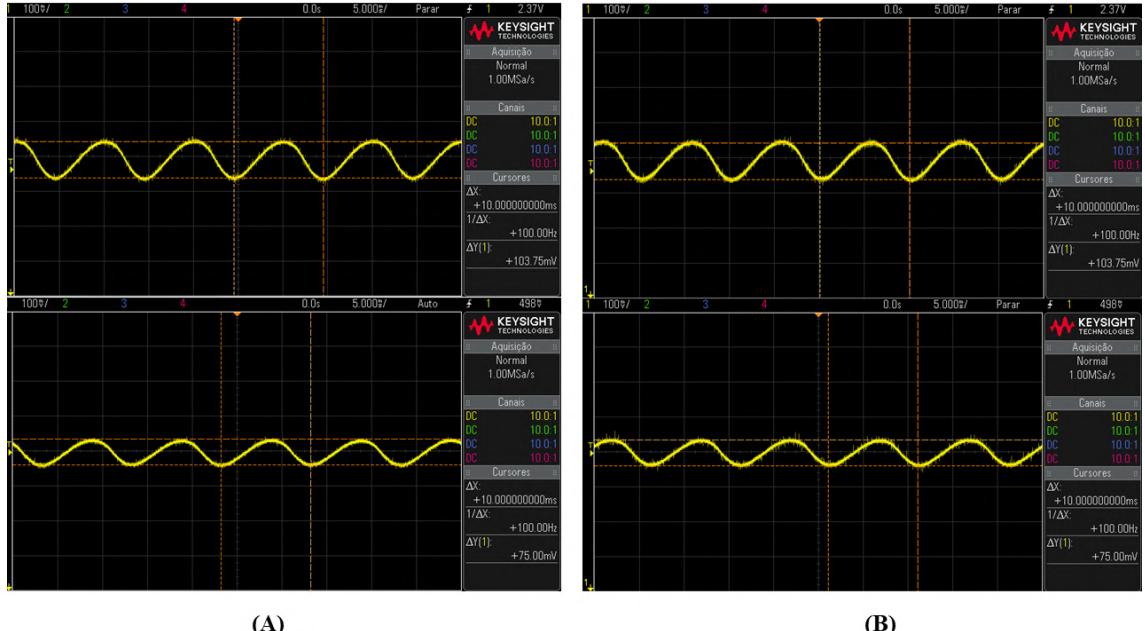


Figure 5.2: Noise measurements on the breadboard circuit prototype compared between using an isolated battery (A) and a USB (B) connected to the computer for an upper Voltage range (≈ 2.4 V) in the first image row and a lower Voltage range (≈ 500 mV) in the images below. The images were generated directly on the oscilloscope and saved on a pen drive through a dedicated USB port.

is obtained in the lower range. This is because the voltage divider reduces the signals at the ratio defined between the two resistors, including the noise component. More importantly, this noise component is identical between the approaches, meaning it will be present regardless of the use of a LiB battery or supplying the circuit with a direct USB connection. Due to convenience, the latter approach was used since no recharging is needed.

Ending the breadboard prototype test, marked the beginning of translating it into a PCB. Two Lighting Emission Diodes (LEDs) were added to the circuit design as a precaution for future debugging. A blue LED was placed between the USB type C port and the LDO and a red LED between the LDO and voltage divider, with a lack of light on the latter representing a problem with the connections after the LDO and a lack of light in both LEDs depicting a problem with the entire supply. The diagram of the whole circuit is illustrated in Figure 5.3, with OUT1 depicting the positive connection to the positive channel of the EEG amplifier's DC channel and OUT2 the negative connection. Additionally, two $5.1\text{ K}\Omega$ resistors were placed in the terminals CC1 and CC2 of the USB Type C port component and two $10\text{ }\mu\text{F}$ capacitors at the input and output of the LDO, as is recommended in their respective datasheet for normal behavior. Furthermore, the USB Type C port used is purposefully directed for circuit supply only, being a simplified version of a USB Type C that contains the additional numerous data transfer pins.

With the circuit schematic fully designed, the PCB layout was drawn to accommodate all components accounting for their realistic size, as represented in Figure 5.4. *Altium designer* does this

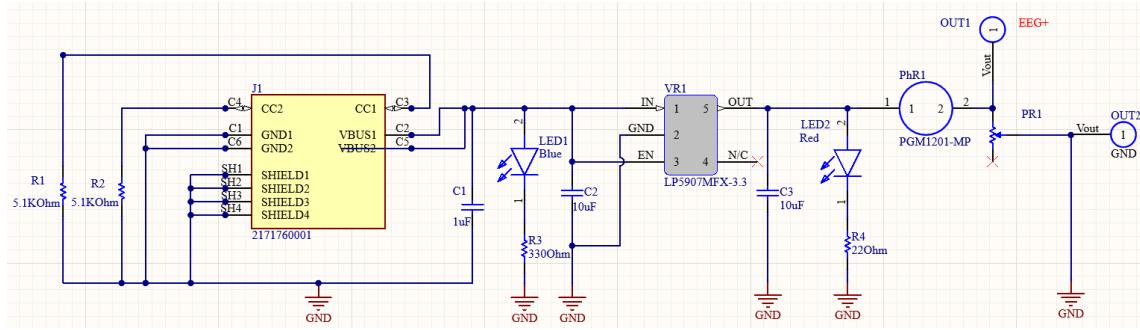


Figure 5.3: The LDR circuit diagram comprising all components. Extracted from *Altium Designer*. The J1 component represents the USB Type C component, VR1 the Voltage Regulator, PhR1 the LDR, and PR1 the $20\text{ k}\Omega$ potentiometer.

by providing ways to import 3D virtual models of the various components with the *pad* sizes and shape information, enabling the design to be perfectly matched in the physical world. However, the LDO model was not found anywhere and had to be carefully designed from scratch following the datasheet dimensions and pad type guidelines.

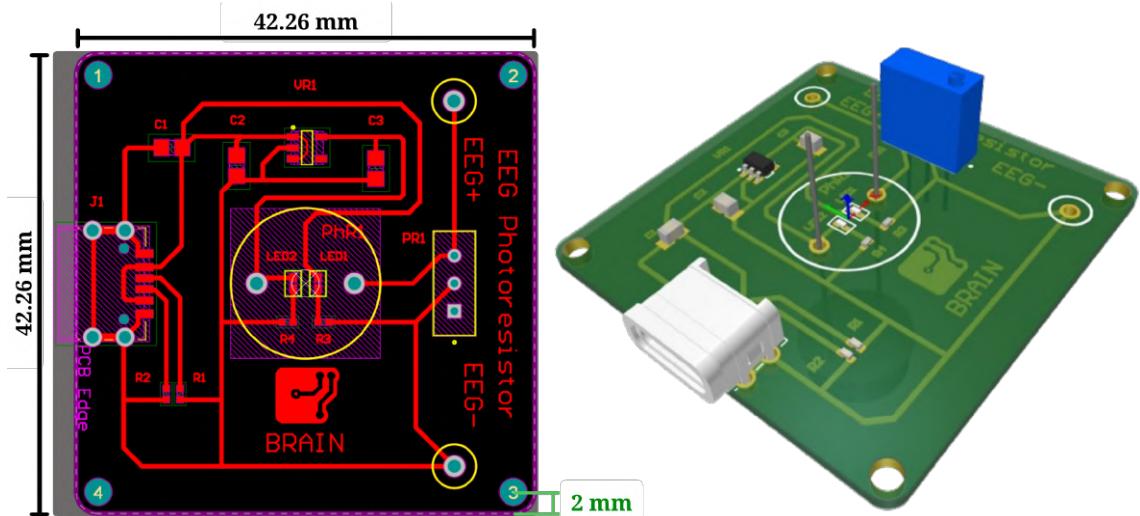
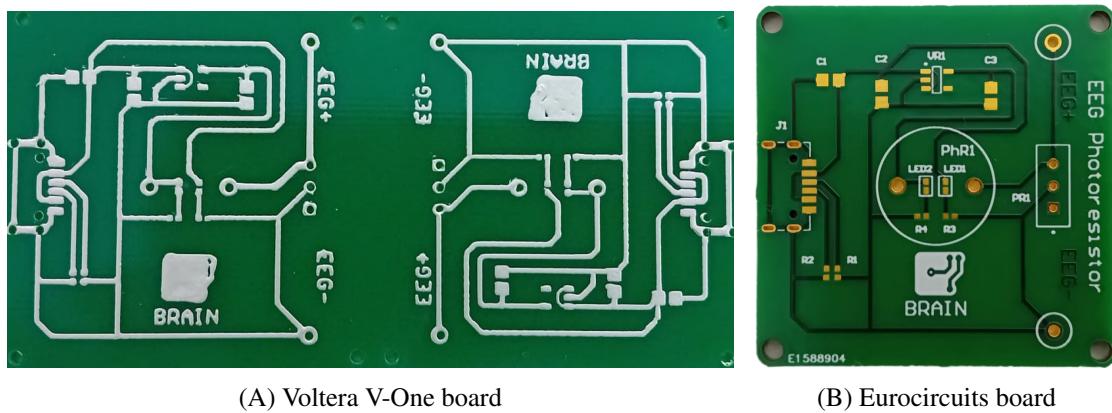


Figure 5.4: The PCB 2D layout design of the LDR circuit and its dimensions on the left. On the right, is an analogous 3D version. Extracted from *Altium Designer*.

The LDR was placed on the opposite side relative to the rest of the circuit, providing light isolation from the debugging LEDs. Additionally, the LDR needs to face the monitor screen while the LEDs can be seen on the other side. The board was specifically modeled to be square-shaped for a cleaner design on the monitor. Its 42.26×42.26 dimensions were chosen based on the *Volterra V-one* board width requirements. The machine can only print circuits on a specific board provided by its company since component mapping is done before placing the conductive ink on the board. The smaller size available has a width of 2×42.26 , allowing for two identical PCBs to be printed on the same board that would be later separated. Furthermore, the four pads placed on each board corner numbered from 1 to 4 have a distance from both immediate surrounding edges of 2 mm

and a diameter of 2.5 mm to accommodate M2 screws, allowing the board to later be fixed within its enclosure.

Despite the undeniable convenience of the easy-to-prototype concept provided by the *Voltera V-One* machine, in practice, the results ended up being underwhelming, as shown in Figure 5.5 (A). Not only was the ink printing sloppy, but its software also suffered from immense incompatibility issues with the *Gerber* files extracted from *Altium Designer*, even though being advertised the contrary. More importantly, when soldering the components by following the *Voltera* guidelines, the ink was easily smudged making it impossible to complete the circuitry placement on the board. For this reason, the final version was ordered from *Eurocircuits*. The final version is evidenced in Figure 5.5 (B).



(A) Voltera V-One board

(B) Eurocircuits board

Figure 5.5: The board resulted from printing on the *Voltera v-One* machine (A) and an analogous version ordered from *Eurocircuits* (B). The difference in quality is very evident.

After soldering the components, the case encapsulation of the PCB was then drawn around the PCB 3D model. In Figure 5.6, four different viewpoints of the case model containing the PCB are shown. The case contains two parts, a lid, and a base, held together by M2 screws. Picture (A) depicts the viewpoint from the top where only the lid is visible, with this being the perspective that faces the subjects when they are looking at the monitor. In the middle, a rectangle hole can be seen, depicting the extrusion from where the debugging LEDs can be visualized from the outside. The blue component peaking on the right represents the potentiometer whose resistance can be adjusted with a screwdriver from the outside as well. The case length and width outer dimensions are of $50.26 \times 50.26\text{ mm}$, with a wall thickness of 3 mm. This gives 1 mm of tolerance on each side between the PCB and the case walls, spanning 44.26 mm from one inner wall to another, as shown in the image (C), also representing a top view but of the base part only.

The case's total height is depicted in the image (B) and corresponds to 20 mm, where a midline cut side-view shows the inner spacing between the case and PCB, including its components. The PCB sits on four supports elevated 6.225 mm from the bottom part inner floor, each with 1.8 mm cut extrusions reaching the bottom floor (C), where two of these are used to fix the PCB inside the case. These will accommodate the M2 screws and the specified cut extrusion diameter allows for the screw to be threaded (2 mm diameter) directly on the plastic with a screwdriver. The circular

holes on the right top corner and left bottom corner (image A) were designed to fit the screw head, having a diameter of 4 mm to accommodate M2 head dimensions ($\approx 3.5 \text{ mm diameter}$). Only two holes were cut extruded on the lid to fix both case parts together, these being diagonally opposite to the circular cut extrusions used to fix the PCB to the base. All component-related cut extrusions are equidistant with a 1 mm tolerance from the specific component, with a circular hole on the bottom allowing the LDR to face directly the screen on the bottom part, as seen in image (C). Additionally, image (D) shows an angled top view of the PCB sitting on the base part without the lid.

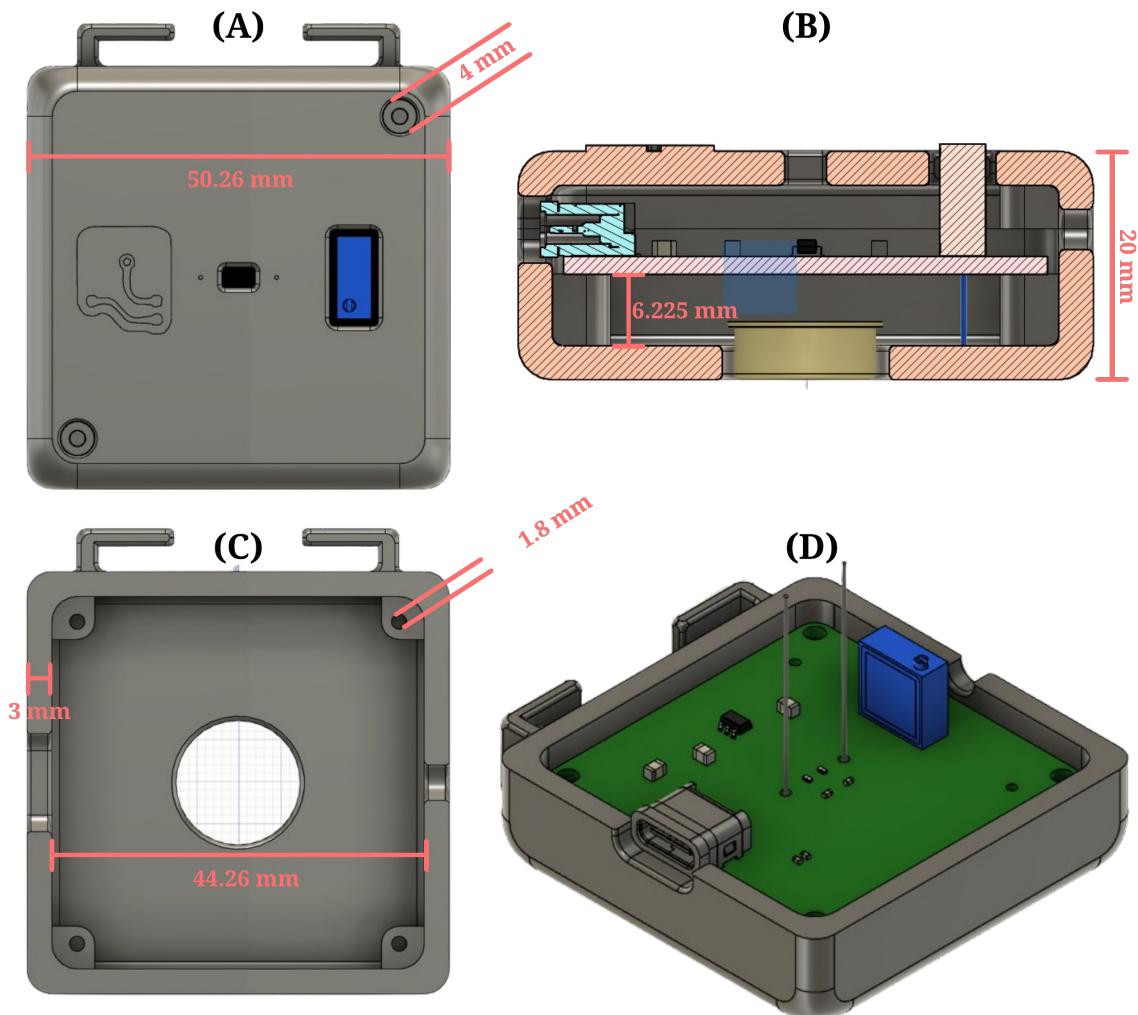


Figure 5.6: Four different viewpoints of the case designed to encapsulate the LDR PCB alongside its main dimensions. Extracted from *Fusion360*.

After 3D printing these parts, the soldered PCB was assembled within the physical model. Figure 5.7 provides four different viewpoints of the enclosure. Image (A) shows a top view of the PCB assembled with two 10 mm M2 screws on the top left and bottom right corners, fixing the PCB in place. The two cable outputs are also displayed, with the green one representing the voltage divider output and the positive terminal of the Amplifier's DC channel, while the black

cable corresponds to the negative terminal (GND). Image (B) provides a side view of the case thickness and demonstrates that the components were positioned in a way that did not extend past the case enclosure. On the other hand, image (C) demonstrates the bottom view corresponding to the face that is in contact with the monitor and where the LDR is arranged. Finally, image (D) shows the top view with the lid assembled by two 18 mm M2 screws on the top right and bottom left corners.

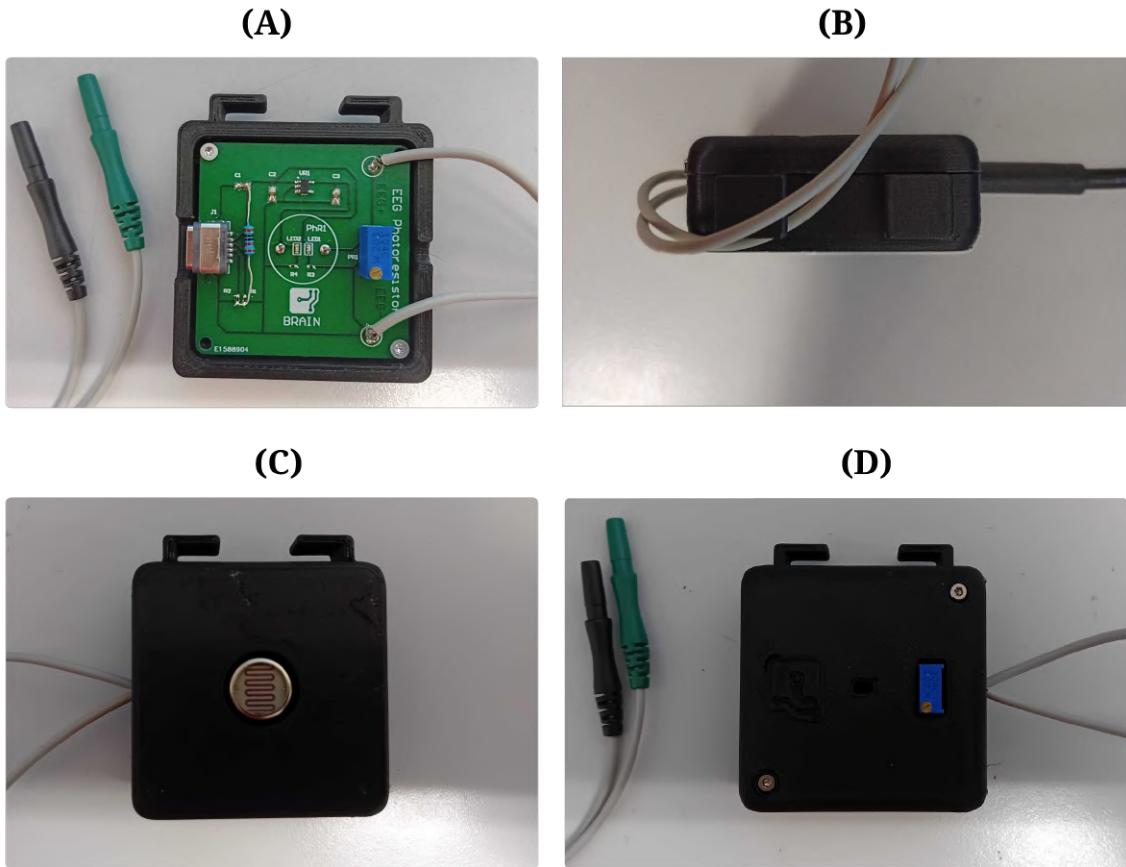


Figure 5.7: Four different viewpoints of the printed case with the assembled PCB.

Now, the only step remaining is to fix the PCB to the paradigm's presentation monitor. For this, the corner-shaped extrusion that comes out of the square-formed case was designed to fit a clamp piece that can fix the case to the monitor, as demonstrated in Figure 5.8. The clamp contains two parts that compress the monitor from opposite sides and are fixed together by a screw and nut through holes in the upper region of both pieces. Furthermore, the dimensions displayed in Figure 5.8 were designed to add versatility and fit various monitor types and sizes.

With all parts designed and dimensions checked within *Fusion360*, the clamp was printed and assembled with the rest of the PCB. Figure 5.9 demonstrates the PCB case held in place by the clamp built. Moreover, it can be seen that the design hardly obstructs the monitor and that the cables interfacing the LDR circuit and the EEG amplifier are easily managed through its structure. Furthermore, the rectangle-shaped cut extrusion on the center is correctly providing the LEDs feedback that the circuit is fully operational. Finally, notice that the PCB case can only be inserted

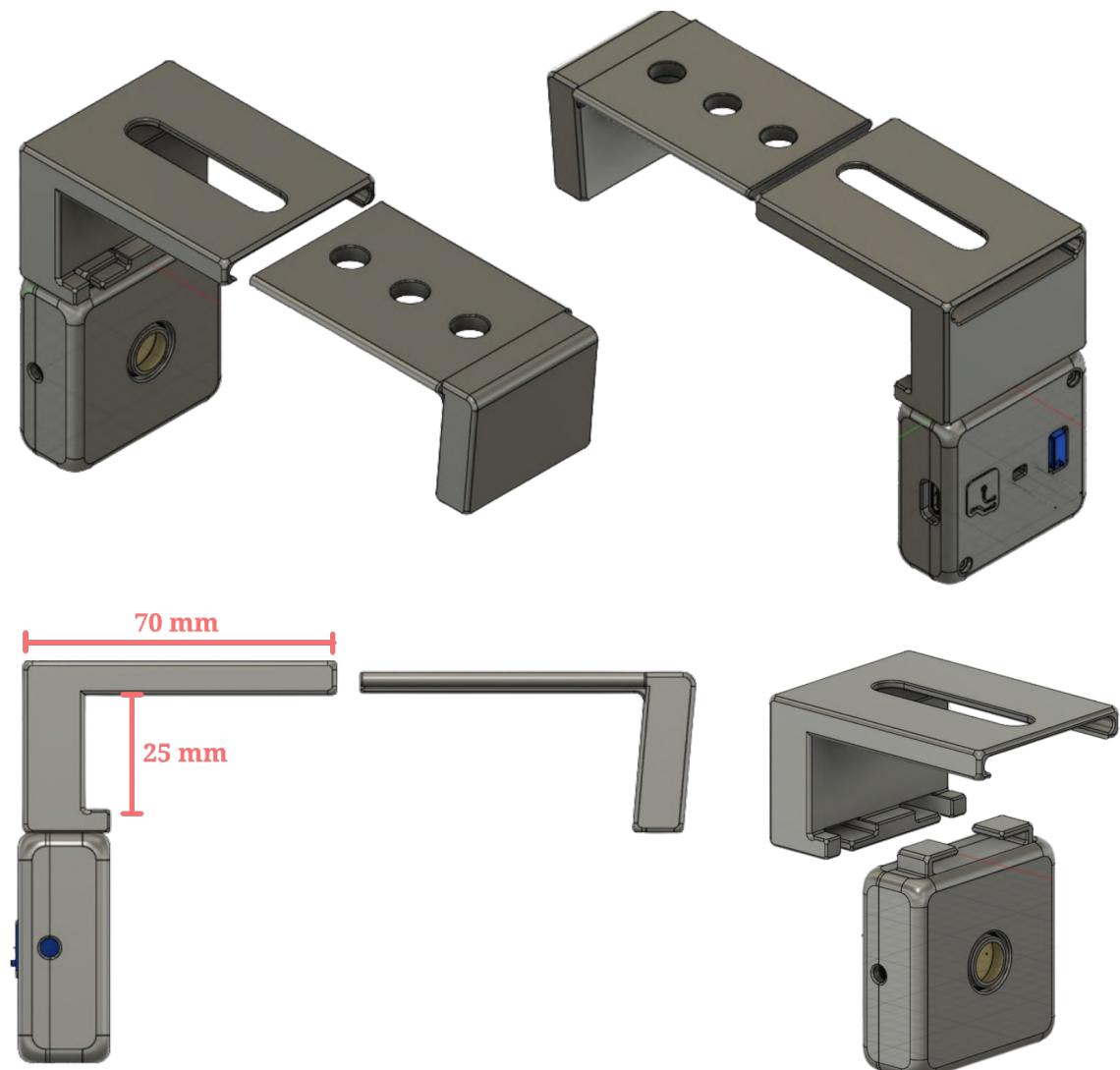


Figure 5.8: Four different viewpoints of the printed clamp designed to fix the LDR circuit to the monitor alongside its main dimensions. Extracted from *Fusion360*.

from the inside of the clamp, preventing the case from falling once the clamp pieces are mounted on the monitor and held together by the bolt.

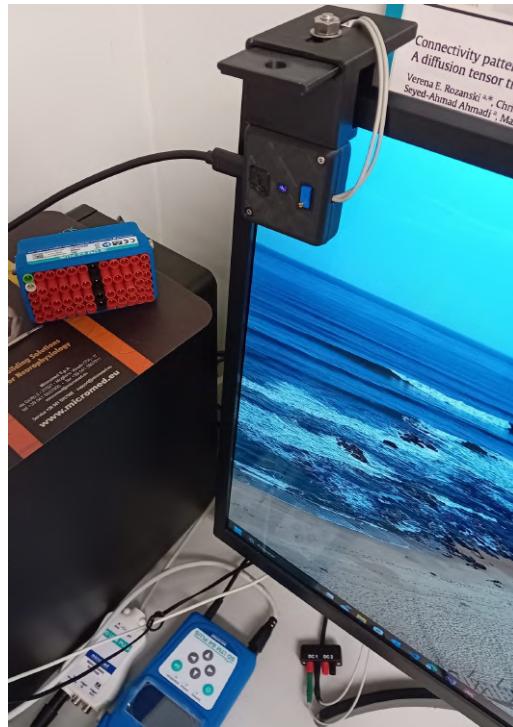


Figure 5.9: The LDR circuit fully mounted and fixed to the paradigm's presentation monitor at C1 lab.

5.1.2 Segmentation of the LDR channel and stimuli labeling

With the circuitry and mounting logistics of the segmentation channel completed, this signal required an assessment of its ability to provide the true labeling of not only the target and non-target stimuli events but also the specific region of the visual angle each target appears from the *excel* files saved throughout the EEG acquisitions.

Before the EEG acquisition phase, the upper threshold of the LDR circuit in response to a white-colored square, corresponding to a target appearance, used during the oscilloscope measurements, was adjusted on-screen using the potentiometer. More specifically, it was reduced from having a maximum peak of around 2.4 V to 2.0 V. This reduction was necessary to prevent a target appearance from overshooting the maximum voltage the amplifier DC channel can handle, resulting from transient noise peaks that may occur during the acquisitions on top of the main line noise component measured previously (Figure 5.2).

The LDR channel of the EEG acquisitions represents the last data column of each *.edf* file produced. A look at the median filtered LDR signal was already given in subsection 4.2.2.4 but no raw signal was shown. The same but raw LDR signal is given in Figure 5.10, representing the fourth trial block of the **Training** EEG acquisition for subject 6.

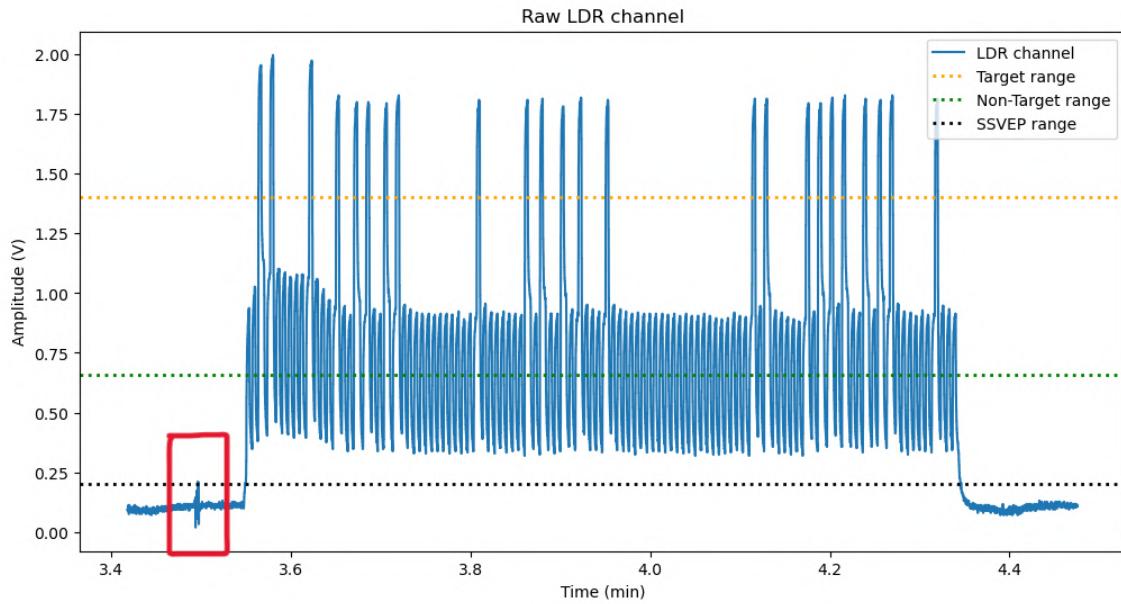


Figure 5.10: Raw LDR channel acquired during the **Training** stimuli presentation for subject 6. The various thresholds for segmentation are once again displayed.

The signal shown here is very similar to the unfiltered version, indicating that the LDR circuit design developed is able to produce a clean signal. However, there is one marking difference highlighted by the red rectangle at the resting period before the trial block starts. This transient peak may have occurred due to movements executed during the resting period. With the SSVEP threshold used, this peak would have messed up the whole signal segmentation since it would be labeled as a rising edge, indicating the algorithm to look for a falling edge, which only happens for this threshold at the trial block's end. Despite being only relevant for the SSVEP threshold, this demonstrates the median filter importance of ensuring good synchronization regarding the unpredictability and sensitivity of EEG recordings.

After smoothing the LDR channel responses, the segmentation algorithm was assessed regarding its ability to find the right epochs with the thresholds specified. A target segmentation for the same trial block represented in Figure 5.10 is shown in Figure 5.11 for a shorter time window.

The target-related rising edges are illustrated in green and the falling edges in red. These intersect the sample immediately following the threshold defined in an ascending and descending manner, respectively. Importantly, the samples found correspond to the target emergence on-screen, concluding that proper target segmenting occurs through the use of the edge detection algorithm.

The same can be said for the Non-Target segmentation, displayed in Figure 5.12. With the target LDR value representing approximately 2.0 V in response to a 100 % white square on-screen, the upper Non-target range of around 0.9 V represents a 37.5 % white square and the lower range (0.35 V) a 12.5 % white square (87.5 % black) defined on the greyscale scale of *Psychopy*.

It is important to note that the samples found through the algorithm do not directly represent

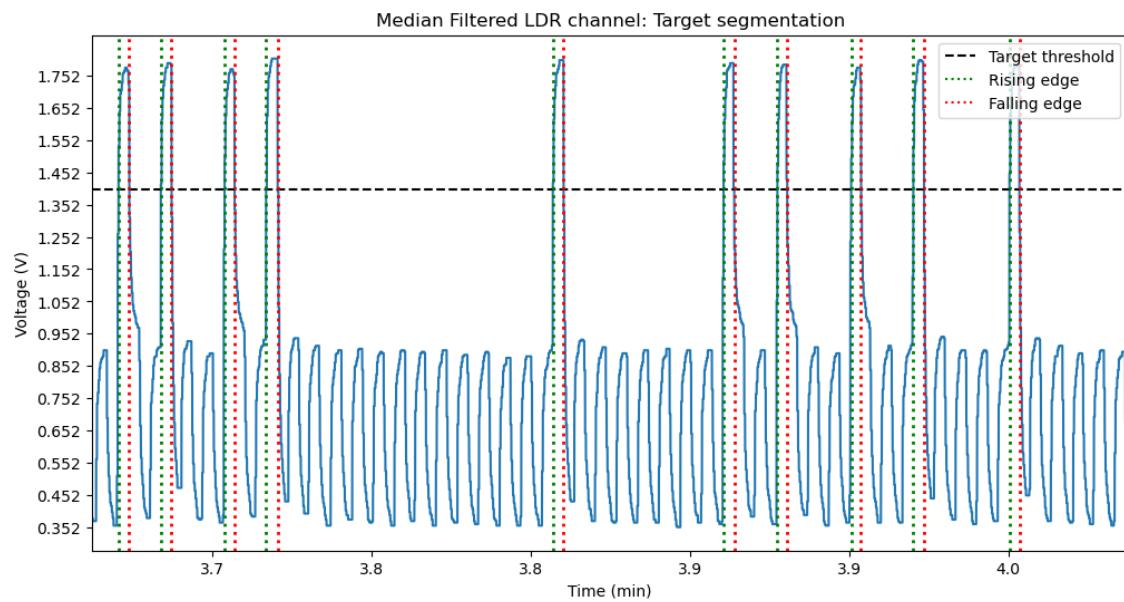


Figure 5.11: Performance of the segmentation algorithm for the target stimuli.

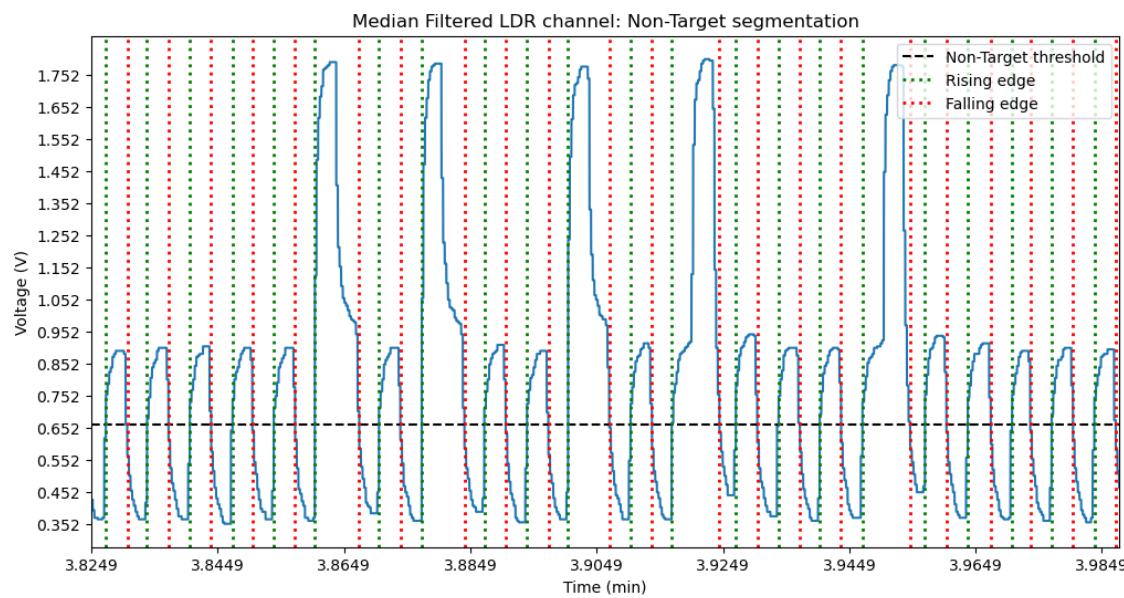


Figure 5.12: Performance of the segmentation algorithm for the non-target stimuli.

the epochs segmented but the stimuli presentation onset, corresponding to $t = 0\text{s}$ on the ERP epochs (-200 to 600 ms). These epochs were obtained referencing these sample instances found in the rising and falling edges.

5.2 EEG signal analysis

Here, the EEG signal analysis will be described thoroughly, starting by displaying the results from the signal quality analysis that drove the choices of signal pre-processing techniques used, followed by the grand-averaged ERP waveforms obtained.

5.2.1 EEG quality estimation underlying the pre-processing choices

Because there are virtually no differences between the signal quality measurements obtained in the **Training** and **Testing** experiments due to these being done concurrently, only the **Training** results in this section are shown. Furthermore, the **Training** protocol results obtained for this analysis are the ones that drove the pre-processing techniques used in both pipelines.

5.2.1.1 The frequency spectrum and bad channel interpolation

Starting with the FFT spectrum in all channels, it can be visualized in Figure 5.13 for subjects 1 and 4, the characteristic $\frac{1}{f}$ distribution of oscillatory magnitude, indicating that a true EEG signal was acquired. Interestingly, in the subject's 1 spectrum, a salient but small peak at the stimulation frequency of image presentation (5 Hz) can be seen but not for subject 4. Furthermore, no big differences are discerned between channels for the same spectrum, indicating no channel appears to produce significant noise relative to others.

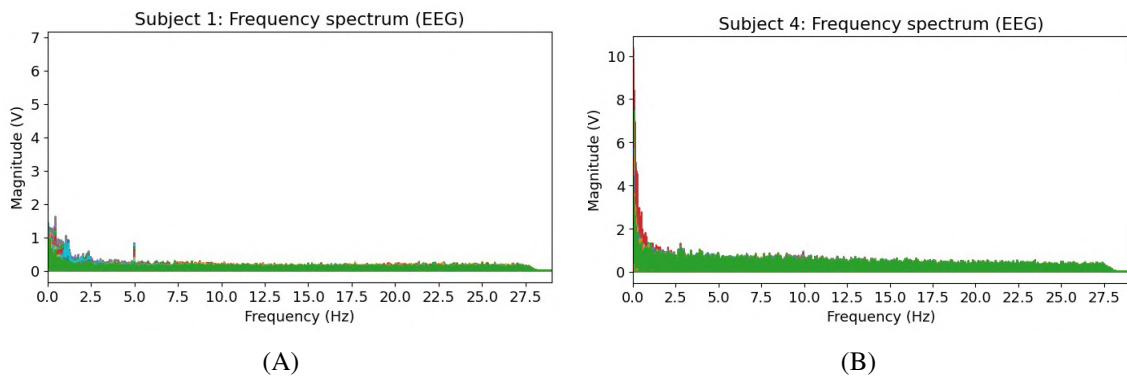


Figure 5.13: The frequency spectrum obtained during the experiment for subject 1 (A) and subject 4 (B). Each color represents the oscillatory data from one electrode.

However, the same cannot be said for the subject's 2 and 3 frequency spectra observed in Figure 5.14. In both spectra, the channel represented by the color red has a discernible energy difference relative to other channels, mainly in subject 2 where the magnitudes reach a $250 \frac{\text{V}}{\text{Hz}}$ in the lower *Delta* range. When trying to find which electrode this red signal represented in both

spectra, channel O1 was found to be the cause. This is corroborated with the impossibility of achieving an impedance check response on this channel during these particular acquisitions. No matter how much conductive gel was used, channel O1 impedance did not go below $20\text{ K}\Omega$, which is the maximum impedance the Micromed system outputs.

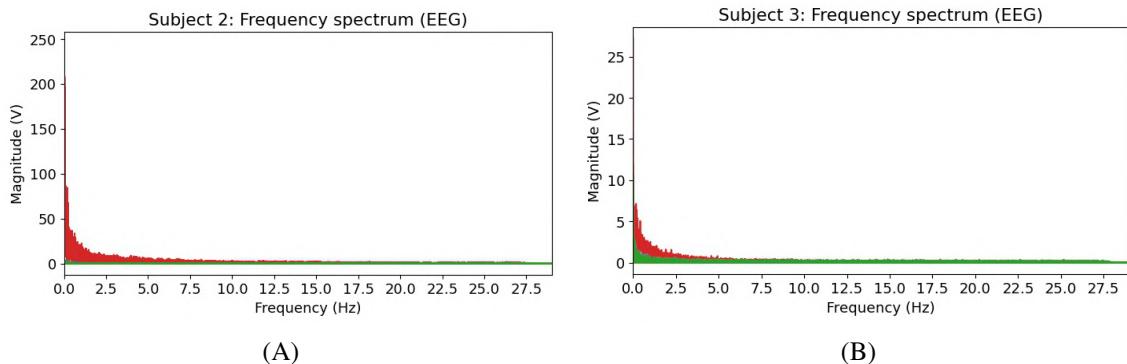


Figure 5.14: The frequency spectrum obtained during the **Training** experiment for subject 2 (A) and subject 3 (B). Each color represents the oscillatory data from one electrode.

One can see the difference in magnitudes between channel O1 (red curve) between subjects 2 and 3 and argue that the frequency spectrum on the latter appears normal, even if producing slightly more energy than on the other channels. However, it is important to note that the analysis of this spectrum has the context of the impedance check done prior to signal acquisition. In both subjects 2 and 3, and later for subject 6, channel O1 was completely unresponsive and hereby, considered unreliable, which in contrast did not happen for the other subjects whose impedance for channel O1 responded normally and was kept below $10\text{ K}\Omega$, as exemplified in Figure 5.15. Hence the decision to interpolate channel O1 in these cases.

The O2 channel signal reconstructed through interpolation can be observed for subject 2 in Figure 5.16. It is possible to see that the reconstructed signal has a much more comparable magnitude relative to the normal channels than before, *i.e.*, the energy on the *Delta* frequency range went from $250 \frac{\text{V}}{\text{Hz}}$ to $2.5 \frac{\text{V}}{\text{Hz}}$. Furthermore, it contains the $\frac{1}{f}$ shape and even an energy peak at 5 Hz, indicative of the SSVEP, assuring that the interpolation produces a reliable EEG signal.

The frequency spectrum for subjects 5 and 6 are depicted in Appendix 1 (Figure A.4).

5.2.1.2 Electrode bridges interpolation

Regarding the channel smearing analysis, the Electrical Distance (ED) metric was computed for each subject to view if the channels were electrically independent or connected through an excessive conductive gel appliance.

This analysis was performed based on three ED data distribution formats, all offering complementary perspectives on bridged channels, as seen in Figure 5.17 for subject 1. Image (A) corresponds to an upper triangle matrix with an increased degree of yellow representing a higher degree of electrical similarity between a channel pair, with both axes depicting a channel index out of the 33 used for recording. This is an upper matrix because ED is computed based on channel

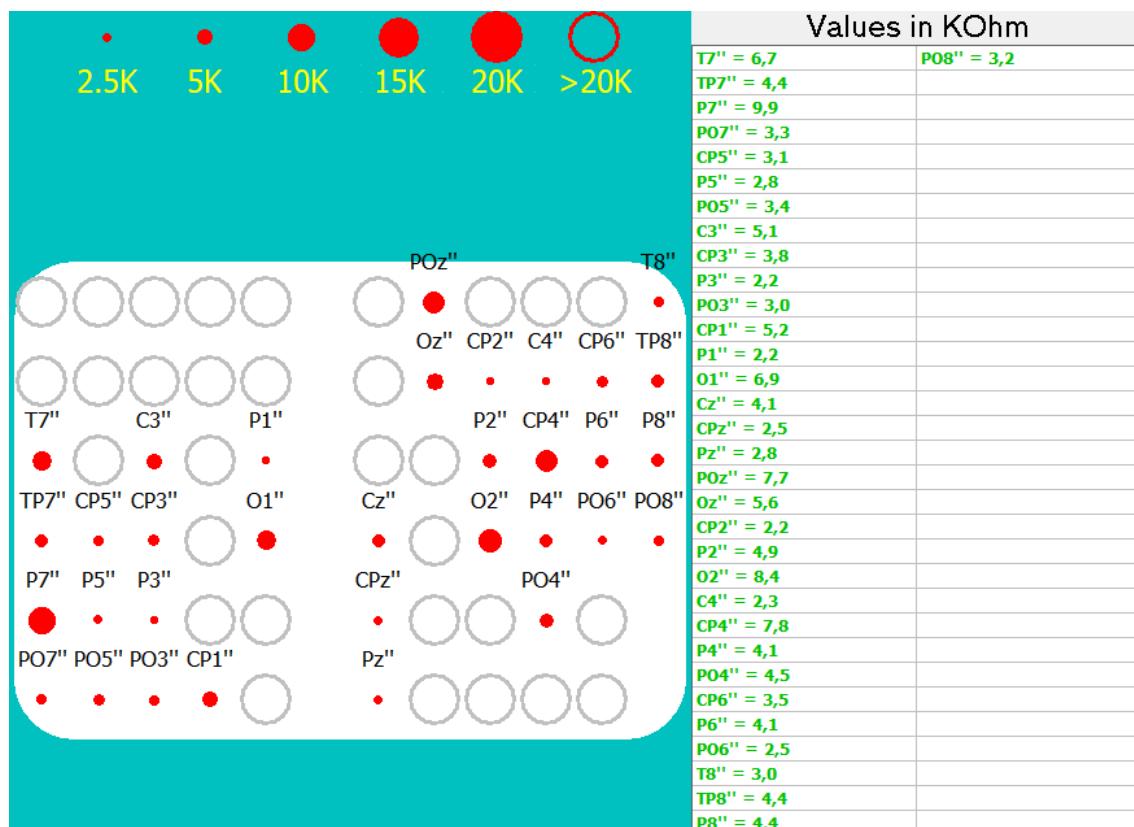


Figure 5.15: Impedance check for subject 1 taken directly from the Micromed BRAIN QUICK software. The ground reference placed on the earlobe measured 2.7 KΩ

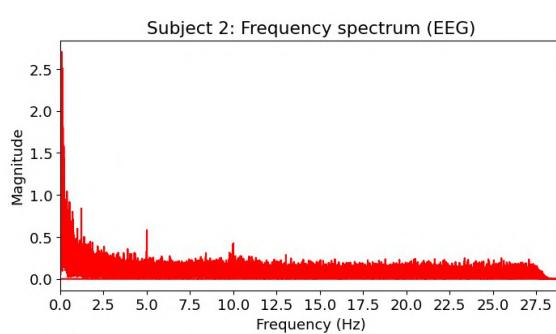


Figure 5.16: Interpolated channel O1 for subject 2. The magnitude is displayed also in Volts (V)

pairs and no metric is obtained between the same electrode. The second perspective of ED is given in the histogram represented in image (B), where the bin width represents a $0\text{-}4 \mu\text{V}^2$ value range of ED, whereas the bin height depicts the count of channel pair connections whose ED value is contained in each bin width range for every 2-second epochs. These ED values counted through epochs may be important to consider since bridging can occur at any time during the recording as a result of the gravity effect on the gel, mainly on the most lateral regions of the scalp. Finally, image (C) offers the most intuitive way to identify channel connections by plotting topographically the distribution of the ED data, providing the scalp view of connected channels. If bridging is identified between two channels, a red line connecting these on the topographic map is drawn, corresponding to an ED value below $1 \mu\text{V}^2$. Images (A) and (C) share the same color scale of ED with green depicting complete electric channel independence.

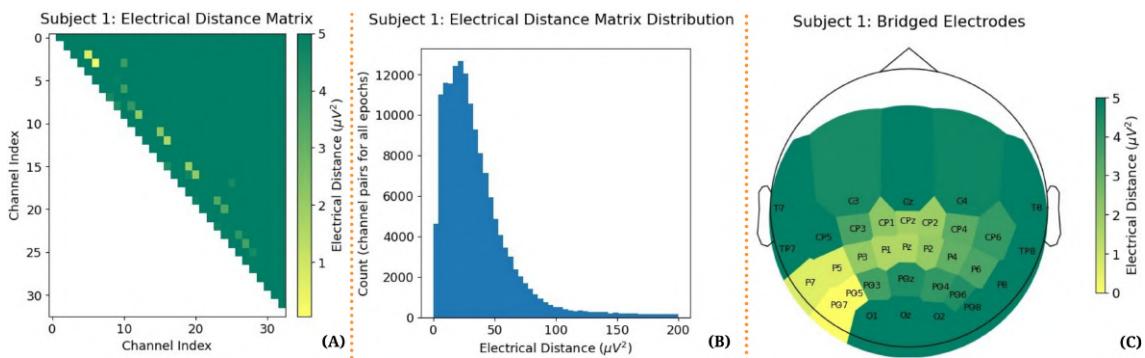


Figure 5.17: Electrode bridging analysis for subject 1.

Firstly, it can be seen through image (C) that no electrode bridging occurred in this particular acquisition, meaning that the gel appliance in the impedance check was correctly done, as indicated by a lack of red lines drawn between two electrodes. However, a more intense yellow color over the left parieto-occipital region, corresponding to channels P5, PO5, P7, and PO8, can be observed, indicating that these are electrically similar but not enough to be considered bridged. In the image (B), these correspond to the most yellow channel index pairs observed within the matrix.

Nevertheless, it is essential to consider that in high-density recordings, adjacent channels (e.g., PO7 and PO5) are expected to produce similar electrical potentials due to their scalp proximity. This is the reason why the histogram bin containing the range of $0\text{-}4 \mu\text{V}^2$ ED has such a high-value count even though no bridging was detected for this acquisition. Such cases may also indicate that bridging did not occur during the impedance check but at any other time of the recording because of gel movement within the scalp due to gravity. To check this, the same metric was computed for the **Testing** experiment but the same results were obtained, indicating that the potentials are indeed just similar or that very little gel crossed from PO5 to PO7.

At the other end of the gel appliance spectrum, subject 2 acquisition suffered with a potential EEG-discarding level of electrode smearing, as shown in Figure 5.18. Image (A) depicts a much more filled ED matrix than the one obtained from subject 1, resulting from many channels being

electrically connected to each other. These yellow regions of connectivity occur on the parieto-occipital regions, with the majority being distributed over the left hemisphere, as observed in the image (C) by the red line drawn between channel pairs. Furthermore, the histogram (B) distribution for this subject shows the channel pair connection count with a prominent peak on the first ED bin ($0-4 \mu\text{V}^2$), as expected for the number of bridged channels observed topographically. In total, 96 connections were generated over these 18 channels, which does not fit the connection number acceptable for interpolation defined in the methodology (0-8 connections).

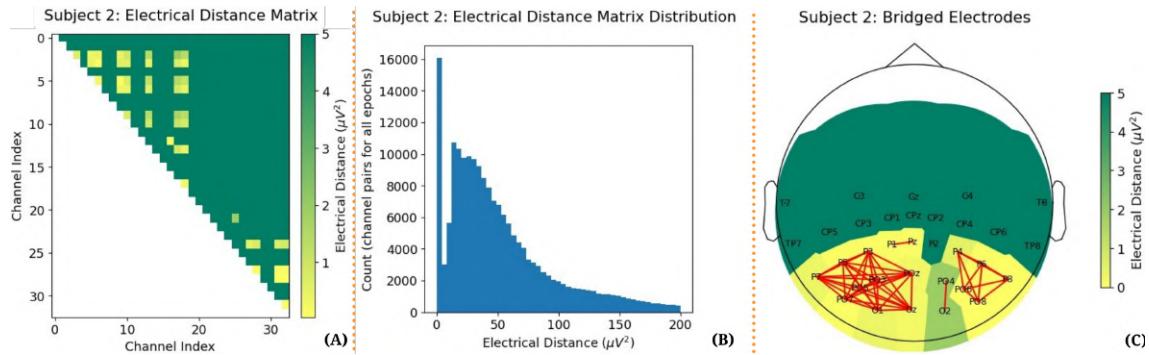


Figure 5.18: Electrode bridging analysis for subject 2.

Despite these results, this acquisition was not discarded due to the lack of inter-hemispheric bridging, not ruining the required hemispheric independence to obtain the N2pc/N2pcb ERPs. Furthermore, other regions of the scalp targeted are not affected, namely the centroparietal locations where the P3 is extracted. However, having this information can provide explanations for when the BCI performance evaluation of this subject is obtained.

As for subjects who were minimally affected by this problem, namely subject 5, interpolation is a viable option to restore electrical independence in these regions. The ED analysis for subject 5 is shown in Figure 5.19. The ED matrix (A) for this subject contains only one connection, as indicated by a single yellow square surrounded by a perfect green distribution across all other channel pair indexes. Looking at the topographic distribution of ED values (C), it can be noticed that bridging occurred between channels PO7 and PO5. Notably, all the other channels are marked as completely independent (green), matching the wide histogram distribution observed in the image (B) and reaching a maximum count of ED values around $60 \mu\text{V}^2$ over all 2s epochs of data. Because of the very focal region of electrode bridging, this channel pair was interpolated.

In addition to subject 5, this type of interpolation was only applied to subject 4, whose EEG acquisition had 6 bridged electrodes between channels PO7 and PO5, P4 and P5, and PO6 and PO8. Subject 3 did not have any electrically connected channels and subject 6 had 10 and thus, no interpolation was performed on the latter participant as it exceeds the defined threshold. Interestingly, the group of subjects whose channel O1 was interpolated is mutually exclusive from the group whose bridged channels were interpolated. The electrode bridging analysis results for these subjects can be found in Appendix A (Figure A.5).

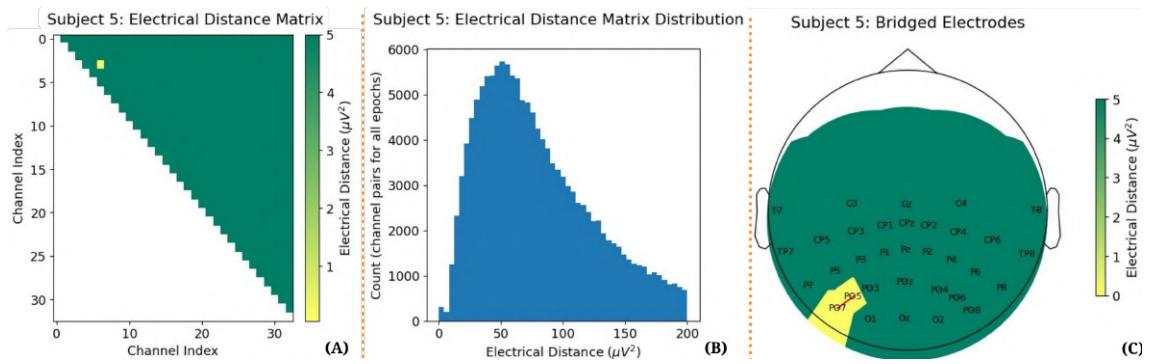


Figure 5.19: Electrode bridging analysis for subject 5.

5.2.1.3 SNR Estimation of the SSVEP component

To estimate the SSVEP contribution to the signal, the PSD and SNR spectra were computed. Figure 5.20 represents these metrics in blue and red, respectively, for subject 3. Both spectra also provide the standard deviation of these metrics for the SSVEP (trial blocks) segmented signal portions in a lighter color.

The image confirms the presence of the SSVEP component at 5 Hz related to the high and constant presentation rate of images shown to participants, here demonstrated for subject 3. When comparing the PSD spectrum to the FFT-obtained one in Figure 5.14 (B), it is possible to observe the similar $\frac{1}{f}$ decay characteristic of the brain oscillatory distribution of energy and power, respectively. However, the SSVEP component is invisible in the former but not on the PSD spectrum, motivating the usage of this metric when estimating the SSVEP. Notably, the 5 Hz stimulation frequency also generated a discernible power burst on its main harmonics at 10 Hz and 15 Hz.

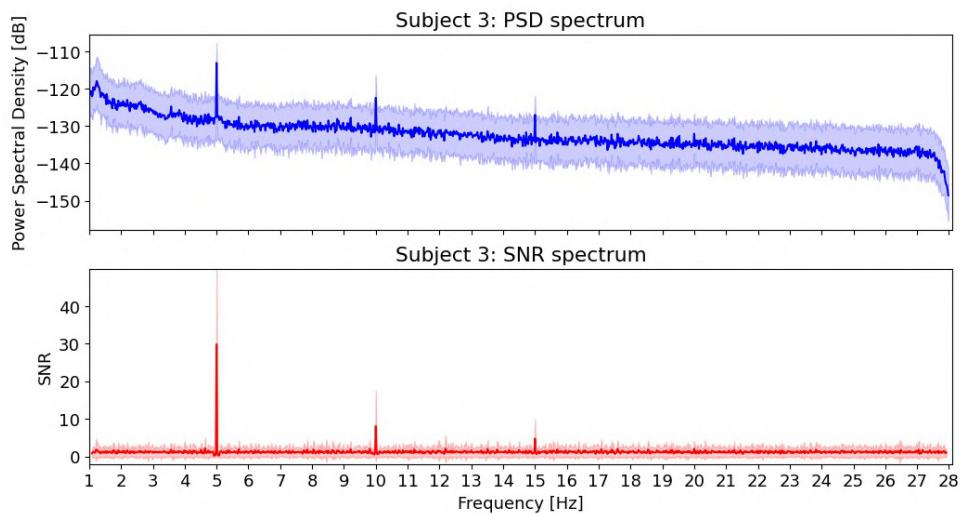


Figure 5.20: The PSD (blue) and SNR (red) spectra for subject 3.

When analyzing the SNR spectrum for this subject, the same oscillatory bursts at the 5 Hz, 10 Hz, and 15 Hz are also very noticeable. This procedure has two advantages over using the raw

PSD insofar as normalizing the $\frac{1}{f}$ power decay and removing wider power modulations normally occurring in the brain. This effect is more noticeable between the power and SNR spectra of subject 4, displayed in Figure 5.21. For this subject, a wide *alpha* modulation is present in the PSD spectrum but not on the SNR one, hence its strength in providing a good estimation of very narrow oscillations such as those formed by the constant SSVEP-related stimulation.

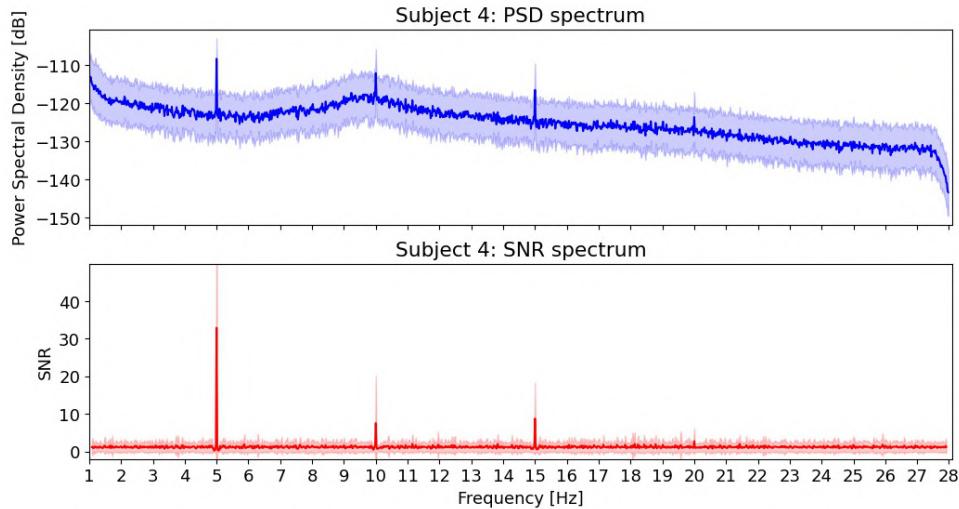


Figure 5.21: The PSD (blue) and SNR (red) spectra for subject 4.

The PSD and SNR spectrum obtained had a very similar shape across all 6 participants, with only the specific PSD/SNR values in each changing for the 5 Hz, 10 Hz, and 15 Hz components. However, in addition to participant 4, a less wide but also present *alpha* power increase was observed in participant 2 PSD, specifically in the upper *alpha*. This may indicate a state of relaxation or "zoning out" of focus during the image presentation for these subjects [119], which must be kept in mind mainly when evaluating the P3 performance with its discriminability being greatly affected by lapses of attention [64]. The rest of the PSD and SNR spectra across all other subjects not discussed in this section are shown in Figure A.6 of Appendix A.

Despite the high SNR values displayed for the 5 Hz, 10 Hz, and 15 Hz, part of the analysis was to see if these components are significantly different between a defined occipital ROI (POz, Oz, O1, and O2) and the averaged SSVEP noise observed across the entire scalp. The *t-test* results are summarized for all subjects in Table 5.1, with the red colored values representing statistical significance between the two groups. In other words, when the *p value* is below the 0.05 threshold.

Firstly, it can be seen that the overall SNR is greater as the image presentation rate of 5 Hz, decreasing with the increase of the frequencies tested, matching the SNR spectra obtained. Secondly, although the overall SNR at the visual ROI is greater than the SNR in the entire scalp, this is not always the case, indicated by a negative *t-value*. This occurred for subjects 4 and 6 at 10 Hz and subjects 1, 4, and 5 at 15 Hz, with these differences being statistically relevant for subjects 1 and 4 at 15 Hz. Although strange at first, it simply means that most of the SSVEP component was not spatially coincident with the visual ROI defined and is occurring in other regions with

Table 5.1: *t-test* results of the SNR between the visual ROI and the entire scalp for all subjects (S). These tests were done for the image stimulation frequency (5 Hz) and its main harmonic components at 10 Hz and 15 Hz. The values marked as red represent the statistical significant SNR values between the entire scalp data and the visual ROI defined within the corresponding frequency assessed.

5 Hz				10 Hz				15 Hz				
SNR		t-test		SNR		t-test		SNR		t-test		
Scalp	ROI	t	p	Scalp	ROI	t	p	Scalp	ROI	t	p	
S1	31.9	33.4	0.325	0.753	7.7	10.0	1.775	0.110	6.5	2.9	-4.743	< 0.001
S2	34.3	41.4	2.146	0.060	14.0	15.2	0.533	0.607	9.0	13.1	3.948	0.003
S3	29.9	36.3	1.409	0.193	8.1	8.9	0.685	0.510	4.7	6.8	2.369	0.042
S4	32.9	34.1	0.379	0.714	7.5	6.6	-0.738	0.479	8.8	6.3	-2.401	0.040
S5	40.6	89.7	6.194	< 0.001	13.0	16.7	4.141	0.003	5.7	5.6	-0.251	0.808
S6	23.3	50.4	4.883	< 0.001	12.9	11.1	-1.129	0.288	8.9	13.8	1.817	0.103

greater power. This is confirmed when plotting the SNR values topographically for these subjects, as shown in Figure 5.22, with the visual ROI being highlighted by a green dot over these relevant channels. In these subjects, the 10 Hz (middle column topographies) and 15 Hz (right column topographies) SSVEP harmonics are indeed present more anteriorly relative to the visual ROI, mostly around the centroparietal region while exhibiting some hemispheric predominance to the left.

Interestingly, the image presentation frequency (5 Hz) only produced a statistically significant SSVEP on the visual ROI for 2 subjects, these being participants 5 and 6, corroborating with these subjects' topographies (left column of Figure 5.22). Participant 2 obtained a *p-value* of 0.060, almost falling in this category. However, this did not happen because this oscillatory component appears to be focused laterally on channel PO3. Regarding the 10 Hz SNR estimation, only subject 5 produced a 1st harmonic SSVEP, being this the only subject with a significant SNR at two of the three frequencies targeted. As for the 2nd SSVEP harmonic, this is where most subjects produced a discernible SNR component.

Notably, most subjects do not present a significant SSVEP component in the visual ROI relative to the rest of the scalp, including its harmonics. However, this results from the fact that the overall SNR at these frequencies across the entire scalp is high even though some narrow regions are clearly dominating the topographies as sources of SSVEP noise. It appears that this may be caused by the re-referencing used (CAR), which is sensitive to noise and through its computation, propagates a focal point of noise to other parts of the scalp. Regardless of the potential prob-

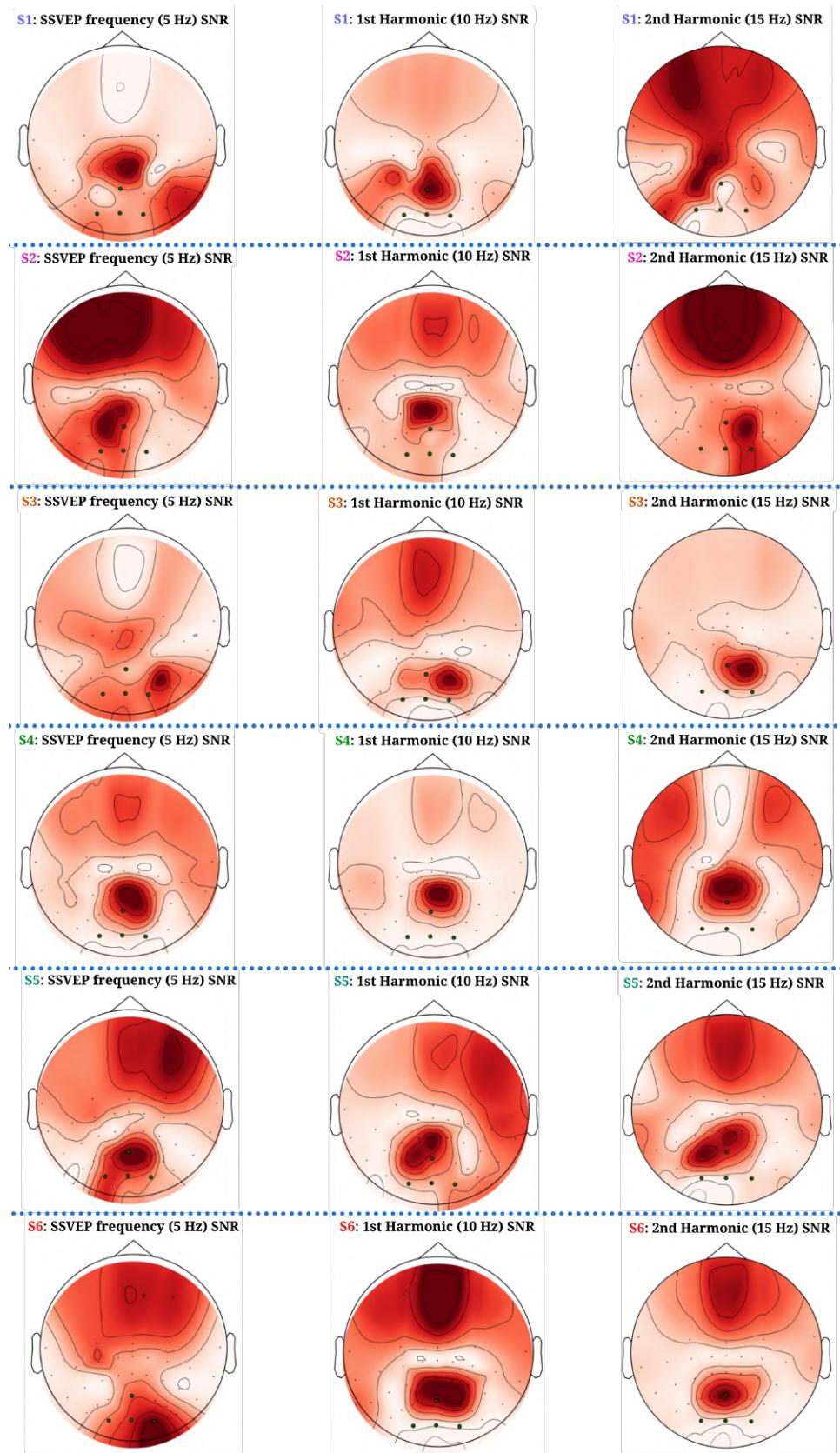


Figure 5.22: From top to bottom, the scalp SNR distribution for subjects 1 to 6 was measured at the stimulation frequency (left column), and 1st (middle column) and 2nd (right column) harmonics. No colorbar is specified because the range is different for every topographic map, ranging from 1 to two times the maximum SNR in each estimation. Note that the anterior regions of the scalp are not covered with electrodes and the activity represented in these regions are artifacts of the *MNE* package trying to fill the entire scalp when computing the topographies.

lems of this re-referencing, CAR is still commonly used in RSVP paradigms, even with higher presentation rates [16].

Nevertheless, the context of the time-domain signal is still required since the PSD and SNR metrics obtained are from the entire trial blocks, which contain target responses mixed in and potentially on top of the SSVEP component. ERPs are low oscillatory responses with the possibility of the N2pc and P3 representing 10 Hz and 5 Hz oscillations, respectively, depending on the experiment's cognitive load [58], meaning that removing the oscillatory information in these targeted frequencies can do more harm than good. Furthermore, if these are coincident with the constant SSVEP, they would be generated on top of this component in the target responses, remaining discriminative anyway relative to the non-target instances. The presence of the N2pc and P3 in these estimations could explain some of the more centroparietal and lateral SNR distributions observed in these topographies instead of the more typical purely occipital SSVEPs (POz and Oz) [73] since their amplitudes relative to Non-Target stimuli would produce more energy in these frequencies. This may also be the reason behind the most discriminative frequency component here evaluated being the 2nd harmonic (15 Hz) for most subjects, where the SSVEP produces less energy, instead of the main frequency of stimulation or its immediate harmonic. In an attempt to separate the SSVEP from the ERP response, the ICA algorithm was used, which may also shine some light on these findings.

5.2.1.4 ICA analysis

ICA was mainly used to target the SSVEP component and spontaneous muscle artifacts that may occur throughout the acquisition. No handling of the ocular movement was necessary since no frontal channels were used. Similarly, no low line noise harmonic (*e.g.*, 25 Hz) was observed on any of the PSD spectrum analyses, excluding the need to also look for line noise artifacts.

The first 10 ICs obtained through the ICAlabel solution for subject 1 are shown in Figure 5.23. It is possible to observe that these topographies captured very focal points of possible artifactual activity within the scalp. These are labeled from 000 to 009 and are ranked based on the ICA model's confidence/probability in identifying each source, meaning that most of the variance in the signal can be explained by that labeled artifact (*e.g.*, muscle artifact). However, the topographies alone are not enough to validate a putative artifact within the data and both the frequency and time-domain analysis of each IC is required to ensure a safe removal without hurting the target responses. ICA components 000, 002, and 004 were inspected more thoroughly to understand their implications on the data and are plotted in Figure 5.24.

These plots all share the component topography (upper left corner), its frequency information (lower left corner), and their emergence throughout the experiment by segmenting the signal fed to ICA in equal time windows, corresponding to an arbitrary ERP (upper right corner), while also providing the component variance throughout each segment (lower right corner).

The ICA solution labeled the first component (Figure 5.24 (A)) as a muscle artifact, being this the most confident labeling out of all components. Combining the scalp distribution and the

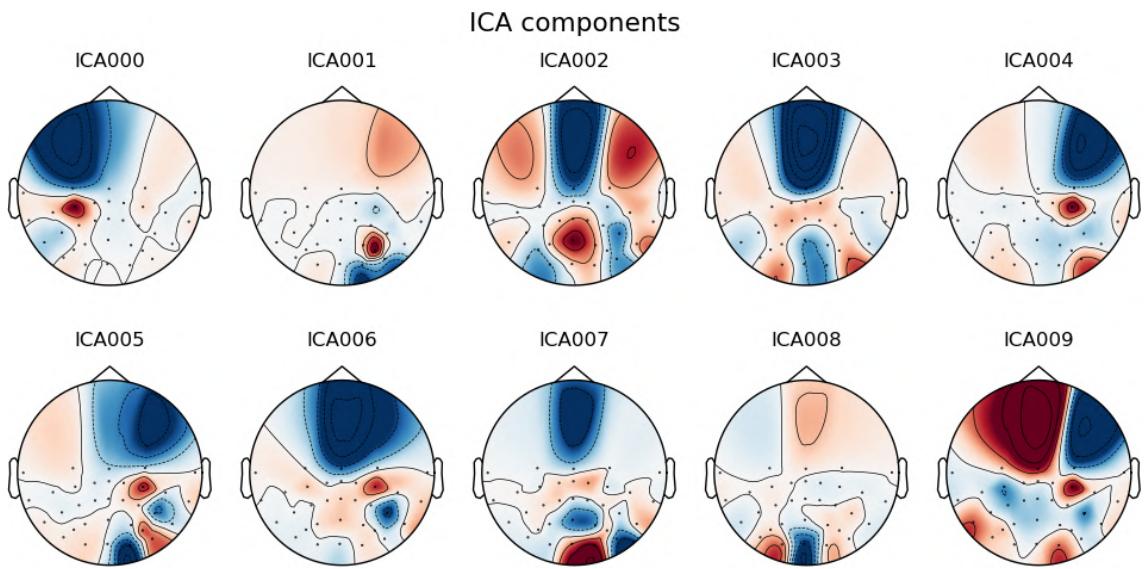


Figure 5.23: The ICA solution for subject 1.

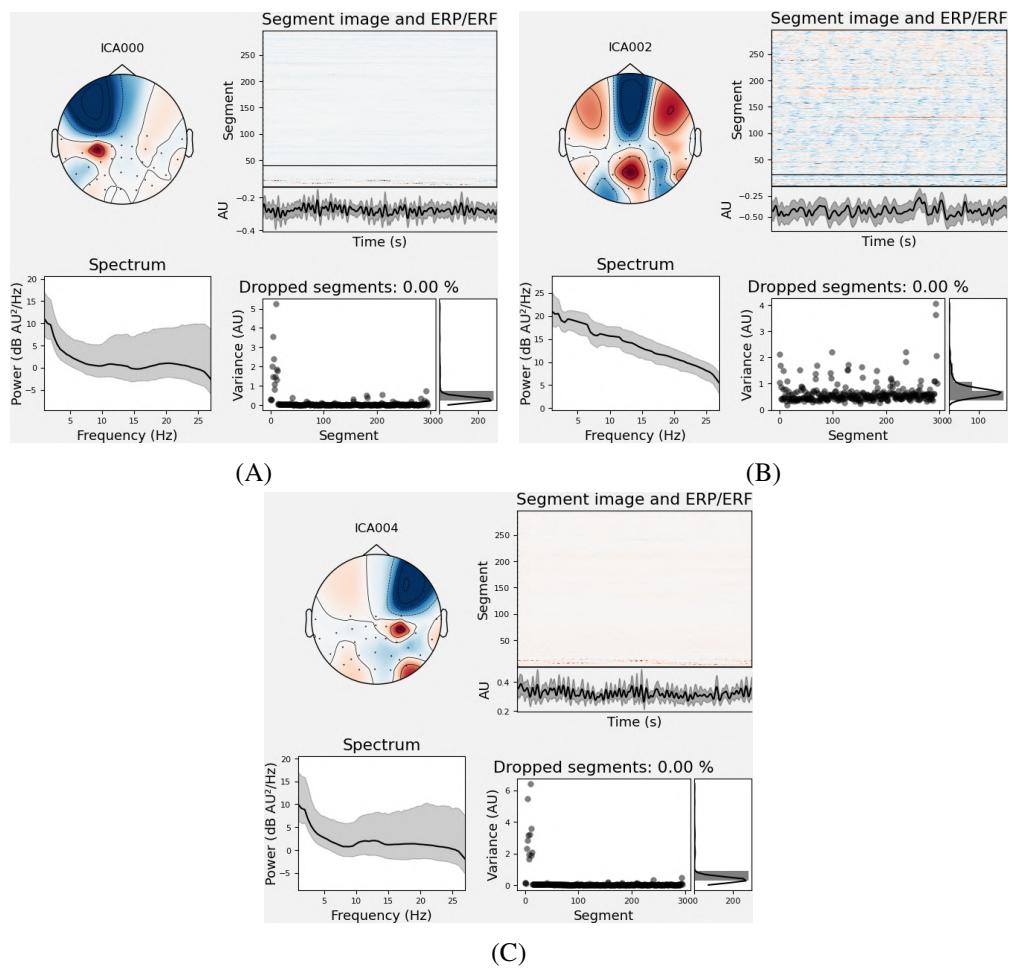


Figure 5.24: Analysis of ICA components 000, 002, and 004 for subject 1, corresponding to images (A), (B), and (C), respectively.

oscillatory information of this artifact, it can be confirmed that this source was correctly identified. In addition to resembling a muscle artifact topography by displaying a very focal energy burst (red) on the left central hemisphere with a subsequent steep drop of energy both anteriorly and posteriorly (blue) within this scalp region, the spectrum also displays an increase in energy around the upper *beta* frequencies, contrasting with the $\frac{1}{f}$ decay of normal neural activity. More importantly, its origination within the temporal segments matches the EMG recording period used to target jaw movements. This is indicated by a more intense red in the initial segments in the upper right corner plot, which fades very quickly in the subsequent segments. Another way of identifying is by observing that the variance (lower right corner) of this component occurs in the initial data segments. Therefore, it can be confidently stated that this component did in fact capture the jaw movement done during the EMG recording phase of the **Training** experiment, more specifically, the *temporalis* muscle activity recruited during such movements, hence the focal energy burst on channel CP3. Analogously, the same can be said for IC004 and the right *temporalis* muscle activity (Figure 5.24 (C)).

It was observed that removing these artifactual components tenuously hurt the target response, more specifically the P3 since this activity emerges near CP1 and CP2 channels, whose data was used to discriminate target vs non-target stimuli through this ERP. This is indicative that some brain activity was captured within this component, confirmed by the $\frac{1}{f}$ decay below 10 Hz on both these ICs spectra. Nonetheless, the explained variance of these muscle artifacts is only relevant for the initial data segments (EMG recording) and not for the actual image presentation phase, so there is no need to remove them anyway.

On the other hand, IC002 captured a source within the signal labeled as "other". This indicates that it is neither a common artifact nor a brain source. Looking at its topography, it can be noticed that it shares some resemblance to some of the topographies analyzed for the SSVEP component for subject 1. Furthermore, this component variance is much more present throughout the signal than the other two discussed while displaying small energy bursts around 5 Hz and 10 Hz. These oscillations are not very discernible but it is expected since this is a normal FFT spectrum and not the PSD. However, the spectrum of this component denotes the normal $\frac{1}{f}$ decay and the energy bursts around 5 Hz and 10 Hz are wider than the estimated SSVEP at these frequencies. This suggests that a lot of brain data is mixed within this component with the wider bandwidths around 5 Hz and 10 Hz perhaps capturing the ERP frequencies. It is normal for the ERPs to display wide bandwidths within this experiment since different levels of cognitive load are recruited depending on how far the target appears relative to the center of the screen, ultimately influencing the visual ERPs amplitudes [58]. With a higher cognitive load, a reduced amplitude is expected, corresponding to a higher frequency oscillation, mainly affecting the P3 [58].

With these observations in mind, the IC002 component was removed resulting in the reduction of the SSVEP component but at the cost of completely obliterating the ERP target responses. This confirms that both the ERP and SSVEP components were captured by this component and more importantly, that these ERPs are indeed operating within this range of frequencies, which makes it particularly difficult to separate both components.

The ICA analysis from other subjects shared the same patterns of ICs found, namely some muscle artifact, line noise, and "other" components with a lot of brain sources being labeled too as the most relevant ICs, resulting in the exclusion of this pre-processing technique from the analysis entirely. The line noise was a common finding but it is important to note that the EEG signal fed to ICA contained frequencies from 1 to 100 Hz since ICAlabel was optimized this way [113]. This means that even though it was captured, this source of noise is not relevant for the 28 Hz low-pass filtered signal.

Despite not being used to remove artifacts, ICA still provided a good insight into the EEG data acquired. The high presence of brain and non-relevant muscle sources implies that the 28 Hz low-pass filter handled all true non-neural artifacts. This is quite important here because ICA is a computationally intensive and time-consuming operation on the data, not optimal for real-time BCIs. By not using ICA, the high-throughput trait of the BCI proposed here is maintained alongside lower hardware requirements to achieve fast responses in the context of deployability.

As for the neural artifacts, namely the SSVEP, its complete separation from the ERP might not ameliorate the discrimination between target and non-target responses, with the possibility of even being detrimental. Zhang *et al* (2023) implemented a technique leveraging the *xDawn* pre-processing algorithm to generate a SSVEP and ERP templates during a 10 Hz presentation rate RSVP *oddball paradigm* in an attempt to separate these components and enhance the ERPs [16]. Although it provided better BCI results in some participants in comparison to 4 other pipelines that did not attempt to unmix these components, this was not achieved for all participants. Additionally, the improvements made on those subjects were not significant, putting forward the question of the necessity of trading the BCI response time performance of an RSVP paradigm for minimal classification improvements. This is because *xDawn* is also a very computationally expensive algorithm. In fact, when attempting this methodology, the computer crashed due to requiring more than 16 GB of RAM to run *xDawn* for one subject. Furthermore, and as already pointed out, the SSVEP is constant for both the Target and Non-target responses, meaning that the target response is superimposed on top of this component anyway. A look at the ERP grand averages in the following section will elucidate this statement.

5.2.2 ERP Grand Averages

The time-domain analysis of the studied ERPs is provided in this section, starting with the Target vs Non-Target differences in the P3 period, followed by the components used to differentiate the regions within the image where the target appears, namely the N2pc and later the N2pcb.

5.2.2.1 Target vs Non-Target (P3)

Starting with the P3, this potential is expected to be different between the Target and Non-Target image presentations with the emergence of this component over the centroparietal channels representing the matching of the human target appearance that participants hold in their working memory. A look at Figure 5.25 evidences that this is the case across all channels picked to extract

these temporal features, where the **Training** experiment responses are depicted and collapsed across all 6 subjects.

The left plot makes the distinction between both P3 subcomponents with the P3a corresponding to the orange rectangle illustrated between 280 to 380 ms, and the P3b the red rectangle depicting the period between 380 to 480 ms, both relative to stimuli onset, represented by the vertical dashed black line in $t = 0\text{ s}$. These subcomponent regions were defined by looking at the individual P3 responses where this distinction is more evident but it disappears during the grand average computation for the **Training** paradigm. In the regions highlighted (280 - 480 ms) the difference in amplitudes between the Target (red line) and Non-Target (blue dashed line) stimuli is very apparent. Furthermore, the Standard Deviation (STD) across all subject responses is also given on the right plot, where it is possible to observe that for most channels, the amplitude variation of the target response amplitudes is greater than on the non-target counterpart, mainly during the period encapsulating the P3.

The P3 here depicted produces its maximum amplitude centrally on channel Cz and decreases following the sagittal plane posteriorly, decreasing tenuously in CPz and reaching its minimum amplitude in Pz. Notably, a small but noticeable difference between the P3 in channels CP1 and CP2 can be identified, with an increased amplitude showing on the left hemisphere. Moreover, the variation in elicitation amplitudes is also greater in CP1 when compared to CP2, as illustrated in the STD plot.

The plots show the similarity between the Non-Target and Target responses from -200 ms to around 280 ms relative to the stimuli onset, only diverging during the P3 period. Two main observations can be made from this. Firstly, the phase correlation between both responses in each channel demonstrates that the thresholds used for segmentation were correctly chosen, ensuring that the segmented time windows used later for classification are in fact corresponding to the same period relative to image presentation in both the Target and Non-Target occasions.

Secondly, this phase alignment in the -200 ms to 280 ms period demonstrates the presence of the SSVEP component in both image class presentations, whose influence in the signals reaches its maximum in channel Cz with an amplitude of roughly -3.5 V and then in channel Pz but positively. These peaks occur every 200 ms, matching the 5 Hz presentation rate of images, with the anti-phase SSVEP observed in channel Pz relative to other channels being behind the attenuated P3 over this electrode. Having the time-domain signal perspective fits the last pieces regarding the puzzle of the SSVEP component and provides answers to the open questions left on its estimation discussion. The P3 elicitation does indeed match the SSVEP phase across all channels and is superimposed on top of it with a clear distinction between the SSVEP Non-Target amplitudes for this time window. More importantly, it roughly spans 200 ms in length, representing an average frequency of 5 Hz and equaling presentation rate. These findings may indeed explain the insignificant differences related to the visual ROI SNR estimation on this frequency, emphasizing the unreliability and possibly even unnecessary separation of the SSVEP from the ERP.

The peaks observed at $t = 0\text{ s}$ are also related to the stabilization of the SSVEP over the trial blocks. However, it can be noticed these peaks fall slightly behind this time point on the plots. This

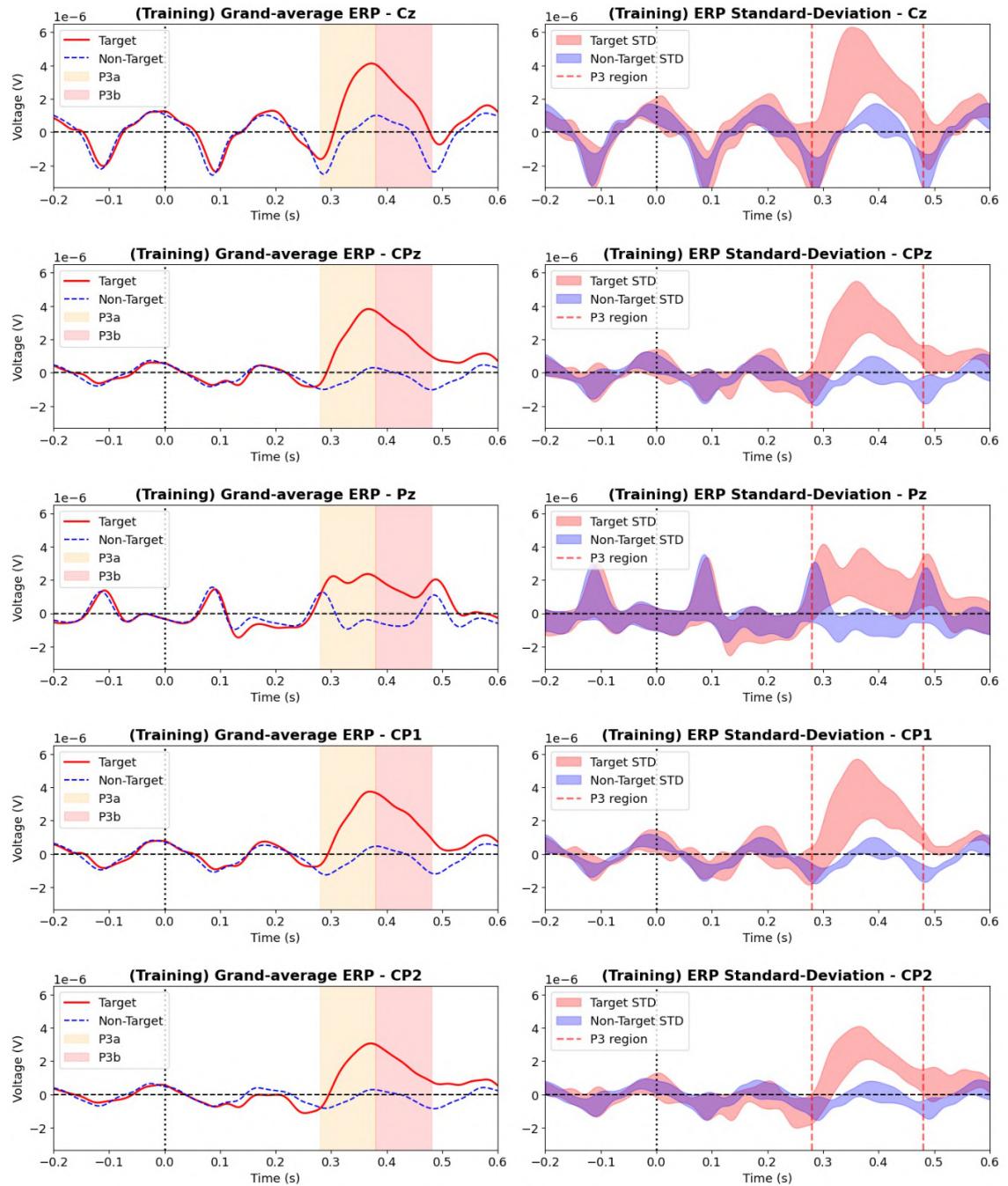


Figure 5.25: The Grand-Average P3 (left) and its standard deviation (right) along the 5 analyzed channels for the **Training** experiment across 6 participants.

can be better observed on the signals displayed over channel Cz, where the first SSVEP negative peak does not match precisely the 100 ms time point when it should. Such phase misalignment corresponds to the LDR component chosen that produces a $\approx 30\text{ ms}$ rise delay [117]. However, this was not taken into account for both classes' epoch segmentation and is not relevant since the phase alignment between both responses is maintained.

Finally, contextualizing all channel information related to the differentiation between Target and Non-Target responses, the topographies across various time windows for these stimuli classes are given in Figure 5.26. Each topographic map represents the averaged activity between 10 ms before and 10 ms after each time window specified for both the Target (A) and Non-Target stimuli with the P3 relevant channels highlighted by green dots. Looking at these topographies it is easy to see why these were the channels picked to discriminate both response classes using the P3, mainly during time windows 0.35 s and 0.4 s relative to stimuli onset. In these periods, the P3 reaches its maximum amplitude for the Target stimuli in the centroparietal channels implied by the red color, whereas in the Non-Target responses, this activity is absent.

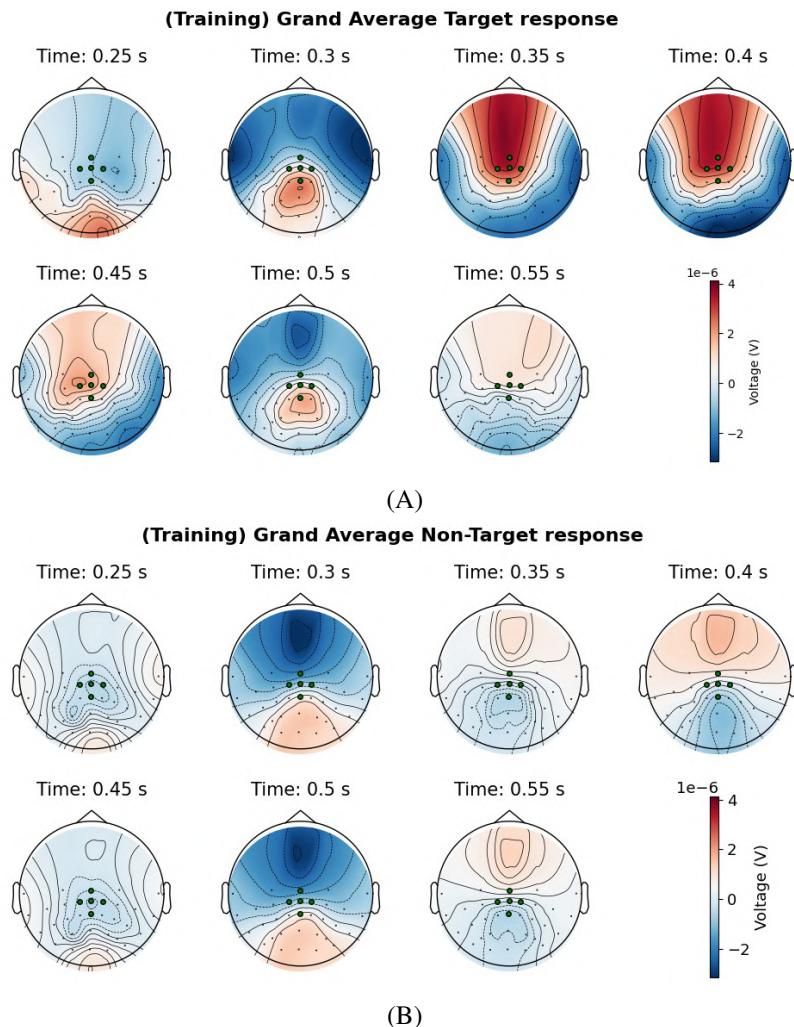


Figure 5.26: The grand averaged Target (A) and Non-Target (B) topographies.

The P3 response obtained during the **Testing** experiment was very similar in comparison to the P3 shown here for the **Training** session. However, a more clear distinction between the P3a and P3b subcomponents was noticed. These are shown in Figure A.7 of Appendix A, alongside its topographies in Figure A.8.

5.2.2.2 Capturing the laterality of target appearance (N2pc)

The N2pc is posed to capture the lateral location in which the human target appears within the images, appearing in the parieto-occipital regions of the scalp more negatively contralateral to the target emergence on the visual field. A look at the grand average N2pc (**Training** session) calculated by collapsing the RVF and LVF responses together shows this response depicted across the 7 channel pairs analyzed by a green line in Figure 5.27. The grand average contralateral and ipsilateral are also shown, illustrated by a dashed red and blue lines, respectively. Furthermore, the standard deviation for these potentials is also provided. The same activity for the **Testing** session can be visualized in Figure A.9.

It can be seen that this potential is much smaller than the previously demonstrated P3, with the amplitudes reaching $-2 \mu V$ to $-2.5 \mu V$ around 150 ms to 350 ms relative to stimuli onset, as depicted by the green region plotted. Furthermore, the contralateral hemispheric dominance of this response is also evident, appearing negatively across all channels, whereas the ipsilateral response shows a positive trend.

A look at its width (≈ 100 to 150 ms) in the time domain indicates that this potential operates within the 7.5 Hz to 10 Hz frequency range, also matching the suspicions behind its possible presence in the SSVEP estimation but not as strongly as for the P3 case. However, despite a great portion of these channels covering the most parieto-occipital regions of the scalp, the SSVEP is much less present than on the P3 analyzed electrodes, being only visible on O1-O2 and PO3-PO4.

Despite its correct representation in the collapsed ERP for both LVF and RVF stimuli plot, a look at the individual N2pc components for the RVF and LVF depicted in Figure 5.28 shows something strange. Most of its amplitude depicted in Figure 5.27 derives from the LVF (right plot) human targets while the RVF (left plot) is much more attenuated. Note that both plots share the same amplitude scale for a direct comparison.

The N2pc has been known to have some functional sensitivity, generating different amplitudes depending on the attentional demands to target stimuli, ranging from $-0.5 \mu V$ to around $-5 \mu V$, with more distant targets from the image center eliciting attenuated N2pc [120, 20]. The N2pc depicted here matches the amplitudes observed in the literature, with the RVF falling in the lower range with amplitudes around $-0.5 \mu V$ to $-1.5 \mu V$, and the LVF responses generating amplitudes of $-2 \mu V$ to $-4 \mu V$.

Evidence suggests that the N2pc contains information related to subject handedness with attempts to discriminate left dominant from right dominant individuals by using this potential, achieving 100 % accuracy across 11 participants in a 5 Hz RSVP *oddball paradigm* [76]. Notably, these differences were found to be more significant for RVF stimuli from which this discrimination was built upon despite no amplitude values being given to perform a direct comparison.

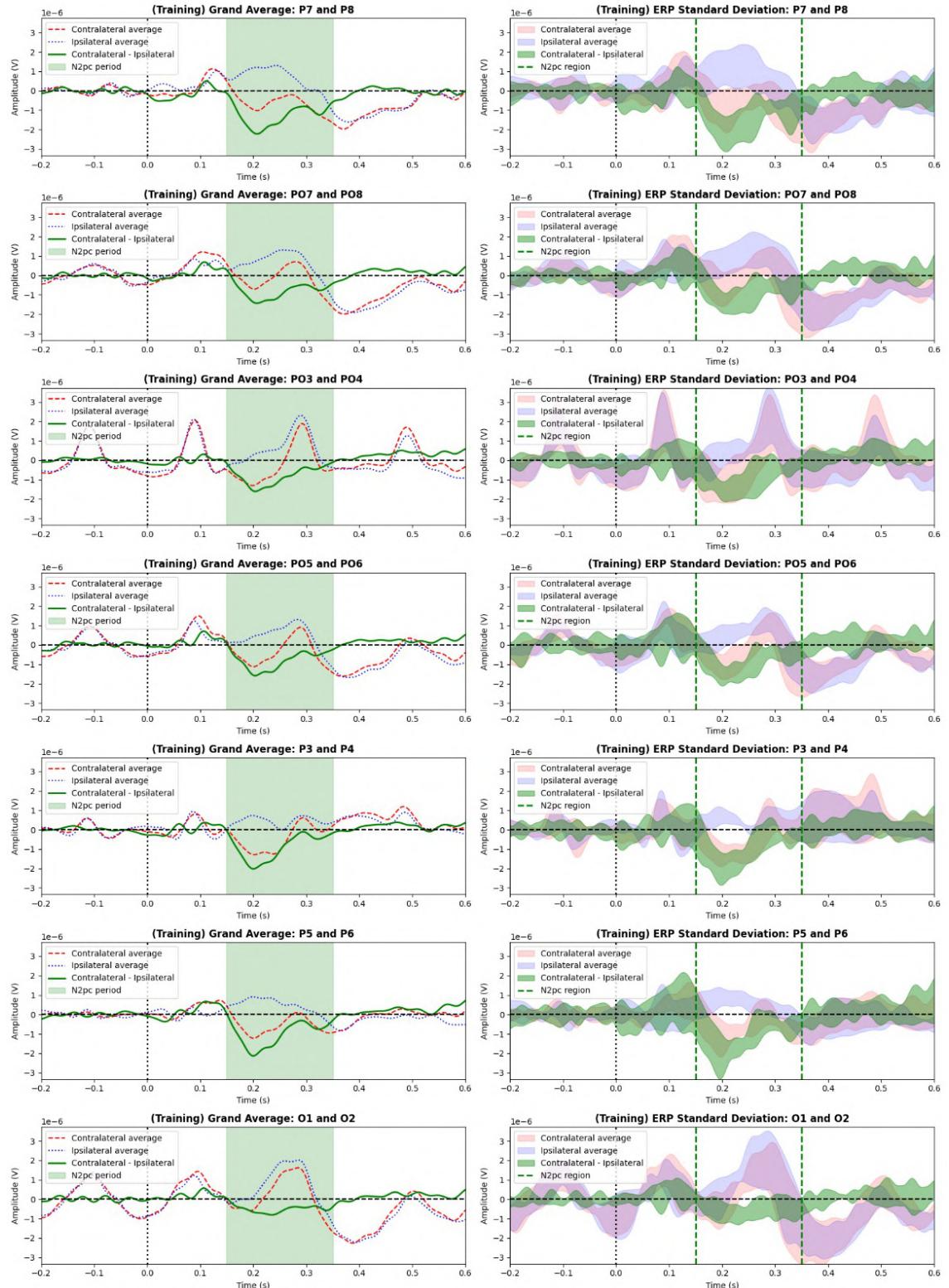


Figure 5.27: The Grand-Average N2pc (left) and its standard deviation (right) along the 7 channel pairs for the **Training** experiment.

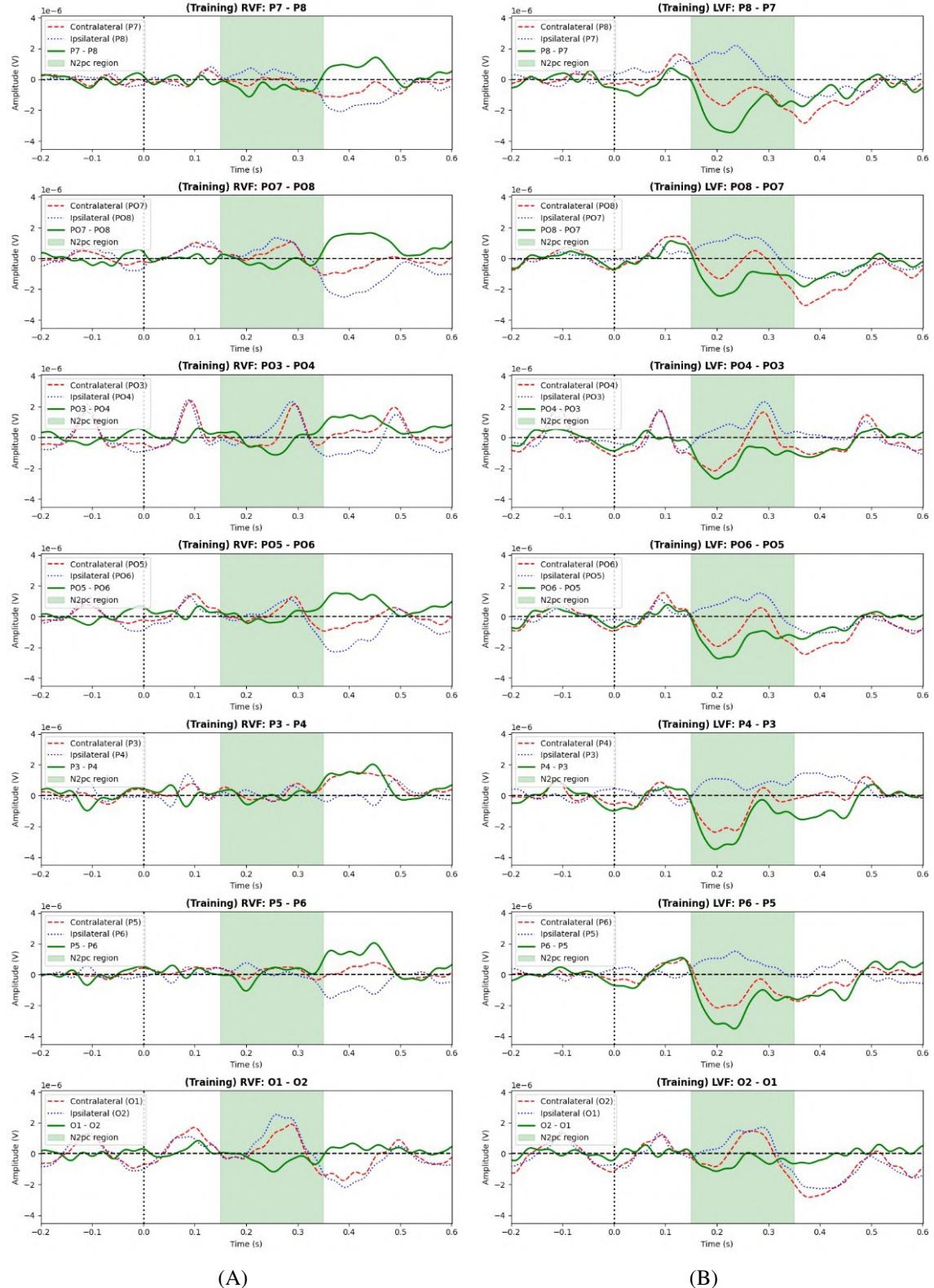


Figure 5.28: The grand averaged RVF N2pc on the left (A) compared to the LVF N2pc on the right (B) across all channel pairs for the **Training** experiment.

Furthermore, it has been known that stimulus-driven orienting of visual attention is lateralized to the right hemisphere, with ERPs emerging earlier on the right hemisphere than on the left [121]. Indeed, when looking at the N2pc elicited in channel pairs PO8-PO7, PO4-PO3, and O1-O2 in Figure 5.28, this RVF potential appears delayed in comparison to the LVF. This phenomenon has been described as the Left Visual Field Advantage (LVFA), with data revealing that this dominance modulating the hemispheric differences in the N2pc depending on stimuli lateralization is independent of handedness in subjects [121]. Despite no data directly presenting answers to the amplitude differences, many experiments demonstrate a positive bias on the orientation of visual attention to LVF targets during time-sensitive attentional tasks, such as during RSVP paradigms [122]. These phenomena could explain the visual field asymmetries on the N2pc elicited here.

Another possibility is related to the target stimuli parameters chosen here. The fact that a human target is shown to participants belly up, implies the brain could pick the target's face and evoke the similar temporal negativity and scalp distributed N170. The fact that this potential is generated negatively around 170 ms relative to stimuli onset [62], indicates that the face processing on the visual field is also picked during RSVP paradigms. However, this poses a hard concept to compare since no papers combining RSVP experiments with facial perception have been found in the literature. More importantly, the N170 reaches its maximum focal amplitudes in channels PO8, PO7, P7, and P8, also presenting a dominance, *i.e.*, *more negativity* toward the right hemisphere [62]. Because of their similar spatial and temporal emergence, these ERPs can be merged together and the right hemispheric biases associated with both potentials could explain the amplitude differences between the RVF and LVF stimuli.

Still regarding the putative presence of the N170 within the N2pc, it can be noticed that the ipsilateral response to RVF targets is very tenuous when compared to the more positive analogous response in the LVF stimuli (Figure 5.28). This could be explained by the RVF ipsilateral positivity competing with the N170 negativity on the right hemisphere. However, the small RVF contralateral response should compile both negativities (N170 and N2pc) in the left hemisphere, which does not happen. This again is only speculation because no study has researched the visual field lateralization effects of faces in the N170 behavior.

If merged together, the parieto-occipital region of the scalp proves to be a difficult region to prove the existence of N170 within the N2pc response due to their temporal similarity. However, the brain response to facial inputs also generates a positive potential of around 170 ms relative to stimuli onset that can be used to confirm the facial recognition stream if performed by the brain, the VPP [62]. This potential is captured at channel Cz and more anteriorly at channels FCz and Fz. A look at the P3 plotted in Figure 5.25 demonstrates a slight amplitude increase during the target responses in comparison to non-target stimuli over channel Cz around 200 ms after stimuli presentation. On the other channels, this deviation at 200 ms between both conditions does not occur. Albeit possibly indicating that the brain does indeed pick and process the human face during target stimuli, during the **Testing** session (Figure A.7) this increased positivity over channel Cz does not appear and channels FZ and FCz would be required to confirm this hypothesis.

Another interesting difference between the RVF and LVF target stimuli arises in the signal polarity transition around 350 ms to 500 ms after image presentation. In the RVF stimuli, emerges a wide positive peak during this period, whereas for the LVF, this temporal region remains negative. This peak has been denominated as the Post-N2pc-Positivity (PNP) and has been recently posed to be correlated with the attentional disengagement from the target stimuli [120]. Furthermore, it has been observed that unlike the N2pc, the PNP amplitude increased with the distance from the center of the image, suggesting a different neural basis despite more studies being required to confirm this [120].

To finalize the N2pc discussion regarding the **Training** session, a look at all the scalp context for the RVF and LVF targets is given in Figure 5.29, obtained in the same manner for the P3 and with the N2pc. It can be observed that during the N2pc elicitation periods, the overall hemisphere differences in the RVF (A) are much more subtle than in the LVF counterpart (B). The negative activity, indicated by the blue color in the right hemisphere of the LVF responses starts more focally in the central occipital region (0.15 s) and over time starts extending toward the parietal and right lateral region of the scalp, illustrating both the ventral and dorsal stream of visual information toward the temporal and parietal regions within the brain, respectively [120]. This negative burst keeps expanding on the right hemisphere (LVF) while rotating contra-clockwise within the scalp until the occipital region is covered with positive activity around 0.3 s. The same happens for the RVF stimuli and the left hemisphere but the overall scale of activity is smaller, as already confirmed by the time plots.

Similar results were obtained for the **Testing** session regarding the N2pc. However, the RVF N2pc in this session appears slightly more discernible across the 7 channel pairs than on the **Training** paradigm. The time plots for the **Testing** experiment are depicted in Figure A.10 while the topographies can be viewed in Figure A.11 of Appendix A, where this increase in N2pc for RVF stimuli can be seen in both plots. Both the ipsilateral and contralateral responses to RVF appear greater than on the **Training** session. Furthermore, the LVFA is also evident with the RVF N2pc reaching its maximum amplitude around 0.25 s relative to stimuli onset while the LVF targets elicit an N2pc with its maximum amplitude around 0.21 s.

This increased N2pc relative to RVF target stimuli in **Testing** session can be explained by the overall increase in salience and color contrast in targets' appearance relative to the environment background [123]. In the **Training** session, the background environment not only is more varied but also the color contrast between the target and background is less evident, mainly when it appears on the sand background due to the yellow color shirt. In contrast, in the **Testing** session, the background environment remains mostly constant with the overall discriminability being easier on the brain. Furthermore, the red sweater provides a greater color contrast to the grass (green) background. Both the overall target salience and color contrast relative to distractors within the environment have been linked to an increased N2pc amplitude [123].

Additionally, during the **Testing** experiment, the human target appears more distanced from the drone with the brain having more difficulty in discerning its face at high speeds, possibly resulting in a more pronounced RVF response than on the first session since the typically positive

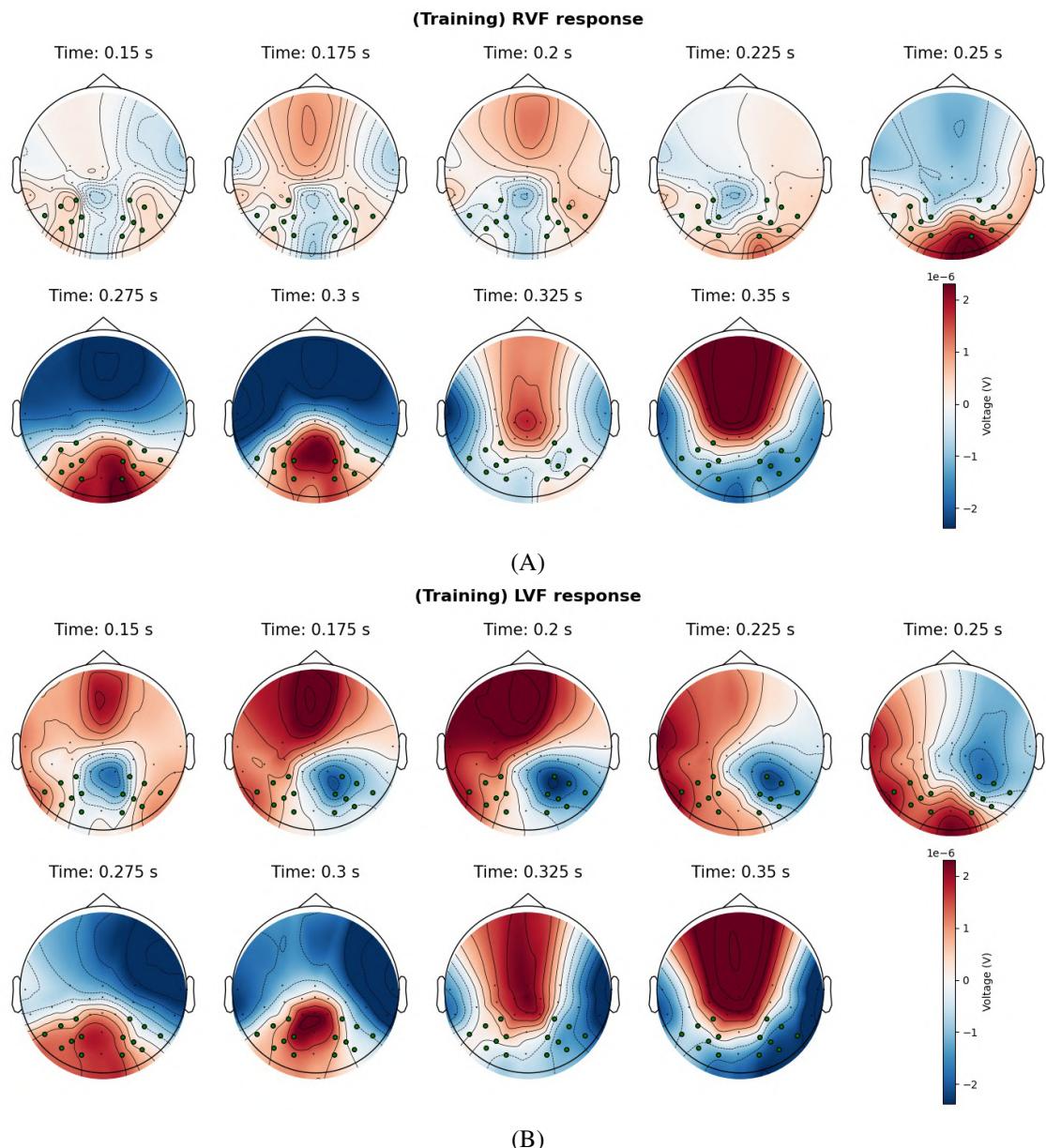


Figure 5.29: The grand averaged RVF (A) and LVF (B) topographies.

ipsilateral response is not competing with the N170 in these channels. Notably, this also could explain the lack of an emerging VPP over channel Cz during the **Testing** session (Figure A.7) but not during the first protocol, where a hint of its presence is shown but not confirmed (Figure 5.25).

5.2.2.3 Capturing the verticality of target appearance (N2pcb and N2pc)

The N2pcb works similarly to the N2pc but tries to capture the target position along the vertical dimension within the images. The grand average N2pcb is represented in Figure 5.30.

When compared to the BVF stimuli, the TVF responses appear to display a more negative N2pcb, namely across the P7-P8, PO7-PO8, PO5-PO6, and O1-O2 channel pairs. This contradicts the only paper that proposes the existence of this component where the N2pcb elicited by targets below the horizontal midline of the image produces a greater negativity [68]. Additionally, no standard deviation is provided here for this ERP because the variations in activity were immense and lacked discernability unlike the ones seen from the P3 and N2pc, indicating that the classifiers that will use the N2pcb to separate TVF and BVF might have a hard time handling the single-trial classification.

Despite these results, it is important to note that the methodology proposed for this ERP in the paper differs slightly from the one applied here. This refers to the fact that the N2pcb described there was applied to vertical targets appearing along the vertical midline by assuming that there is a bilateral N2pc elicited on both hemispheres [68]. Here, however, the N2pcb was computed for all target stimuli appearing above and below the vertical visual angle thresholds (1.2°) while maintaining targets that present a horizontal laterality appearing within the top and bottom regions. This means that the assumption of the bilateral response no longer holds true for the upper and lower stimuli appearance that also displays an RVF and LVF component.

Nevertheless, capturing the modulations of the TVF and BVF stimuli with the N2pcb was still tried on the basis that the analogous N2pc also produces elicitation differences to stimuli appearing above or below the horizontal image midline. It is theorized that this is the result of the anatomical orientation of the neural generators of the N2pc within the parieto-occipital, mainly V4 [120]. The idea is that stimuli on the lower visual field activate more prominently the dorsolateral pathways that are close to the scalp, hence a more noticeable negativity is captured by the EEG for the N2pc [120] and N2pcb [68]. Contrastingly, stimuli appearing above the horizontal midline generate more activity within the ventral pathways that occur more deeply in the brain toward the temporal lobe and are not captured as well, hence a more attenuated N2pc [120] and consequently N2pcb [68].

This can indeed be visualized when plotting the topographical maps, as represented in Figure 5.31, containing the scalp activity related to the TVF (A) and BVF (B) stimuli presentations. When looking at these scalp topographies of neural activity, a more intense, both positive (red color) and negative (blue color) can be seen in the BVF response when compared to the TVF. The LVFA is also present within both scalp topographic maps depicted by a more intense negativity on the left hemisphere below the N2pc/N2bcb target channels (green dots), mainly between the time window between 0.15 s to 0.25 s relative to stimuli onset.

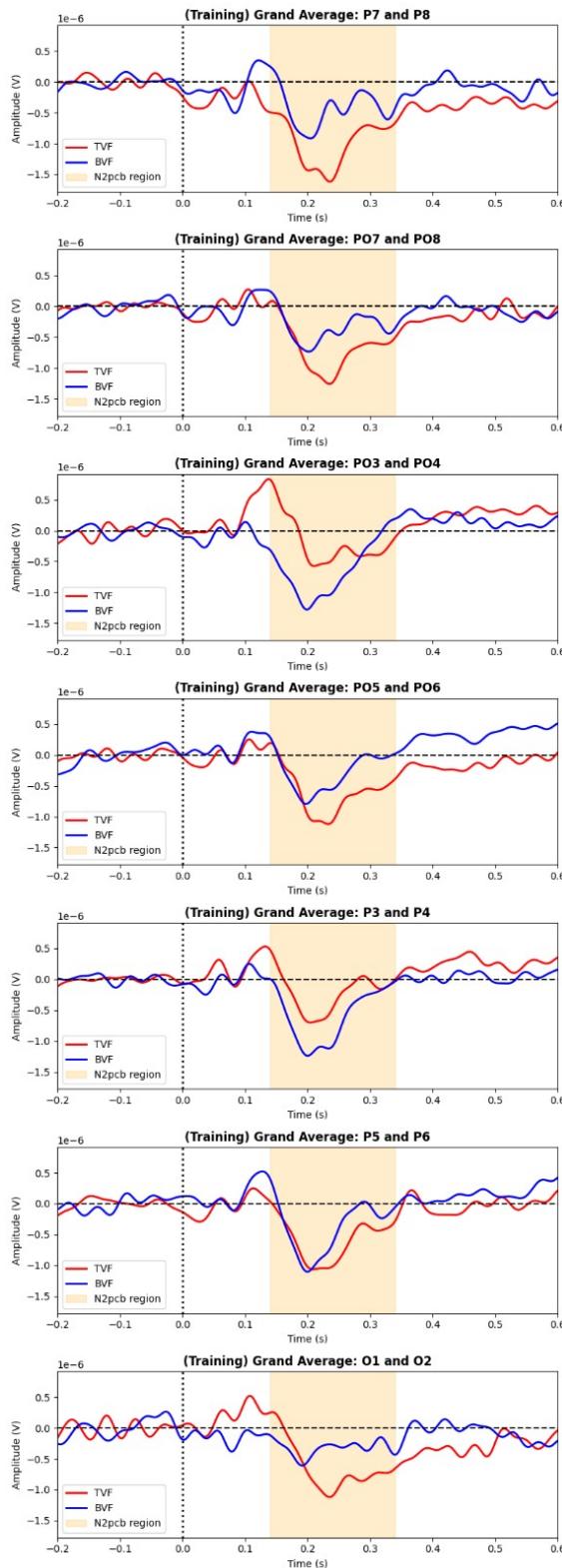


Figure 5.30: The Grand-Average N2pcb along the 7 channel pairs for the **Training** experiment.

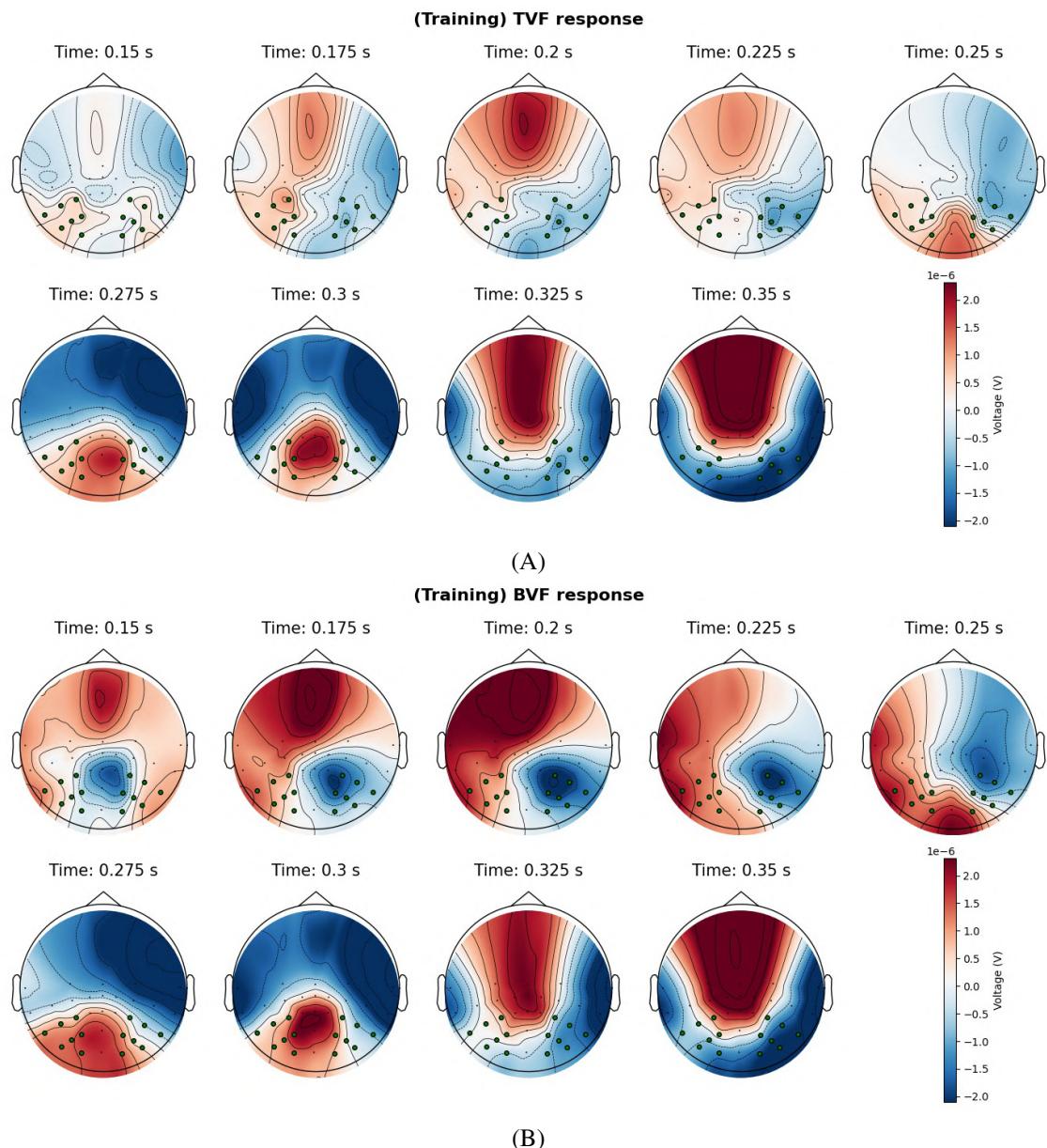


Figure 5.31: The grand averaged TVF (A) and BVF (B) topographies.

Similar results can be observed for the N2pc time plots for the **Testing** image dataset depicted on Figure A.12 of Appendix A. However, during this session, only the P7-P8 and PO7-PO8 channel pairs display a greater N2pc negativity during TVF responses in comparison to BVF target stimuli, with channel pairs P5-P6 and O1-O2 producing similar amplitudes between both classes. The topographic maps for **Testing** session are also displayed in Appendix A (Figure A.13) and exhibit similar focal activity trends with the topographies shown in Figure 5.31 for the **Training** experiment.

Despite these topographies being corroborated with the neural activity produced by both TVF and BVF human target appearances, the inner computations of the N2pc appear to indeed be limited to capturing vertical target stimuli along the vertical midline, which was not attempted here. However, guided by topographies and literature exhibiting different activities for TVF and BVF target responses, a separation of the N2pc produced over the 4 quadrants of image regions was conducted to assess its putative and mixed horizontal and vertical human target modulation. The resulting grand average N2pc waveforms divided into the RVF-TVF, LVF-TVF, RVF-BVF, and LVF-BVF defined by the visual angle thresholds discussed previously (Figure 4.27) are shown in Figure 5.32.

Through this 4 quadrant region separation it can indeed be seen that across all channel pairs in which the N2pc/N2pcb analysis was conducted, the BVF stimuli appearance produces a greater negativity for the same lateral group than the TVF. Putting it differently, the LVF-BVF generates a greater N2pc (negative) amplitude than the LVF-TVF and the RVF-BVF N2pc generates a more negative N2pc relative to the N2pc elicited by the LVF-TVF visual field. Additionally, the LVFA phenomena still holds true, with the LVF group exhibiting an increased negativity in comparison to the RVF group.

It can also be seen that these N2pc waveforms are also more noisy than the ones shown previously for the binary RVF vs LVF discrimination. This is because the EEG acquisition was not intended for such a specific visual angle division of target appearance. While in **Training** session, the binary N2pc visual angle separation resulted in 100 targets in each lateral region, here, each per subject grand average produces only 40 targets across the 4 quadrants. As pointed out in the methodology, the N2pc requires around 100 target trials to exhibit a robust ERP formation [92]. In the **Testing** session, this is even worse where only 24 per quadrant targets are available, producing an even noisier ERP across the segmented epoch. These are represented in Figure A.14 of Appendix A.

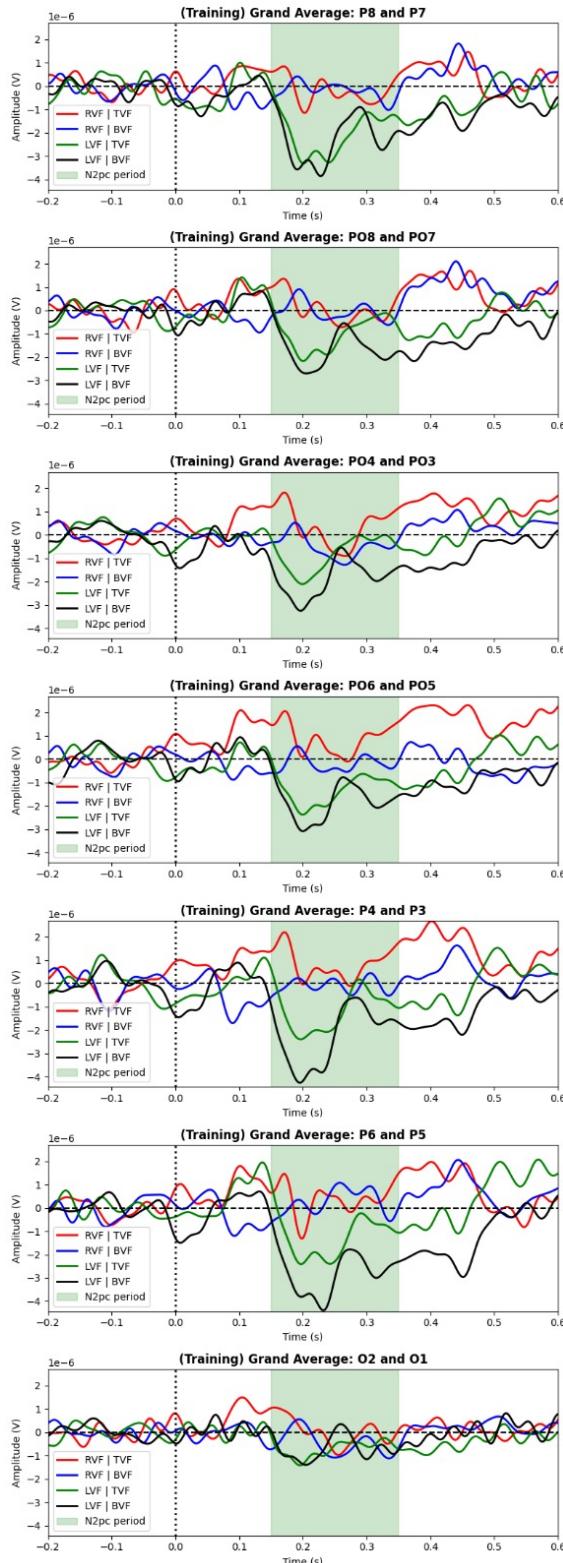


Figure 5.32: The Grand-Average N2pc along the 7 channel pairs divided into four quadrants (RVF-TVF vs LVF-TVF vs RVF-BVF vs LVF-BVF) for the **Training** experiment.

5.3 BCI performance

In this section all the models' performance based on the ERPs discussed are described, starting with the Target vs Non-Target discrimination, followed by the N2pc and N2pcb performance in discriminating between the target's lateral and vertical appearance, respectively. All models were validated through a 10-fold CV loop with a 75 %/25 % split of the single-trials **Training** data session and tested with all the single-trials **Testing** data session.

5.3.1 Target vs Non-Target classification

Despite the N2pc/N2pcb potentially allowing to discern where within the image the target appears earlier on the visual stream processing, the BCI application firstly requires the determination of whether a target is contained in said image, where the P3 may or may not be essential. For this reason, the target vs non-target discrimination is asserted first.

Before discussing the BCI models, the ground truth established with the Human Pose Estimation model is required and the results are not satisfactory. In the **Training** session images, the algorithm correctly identified only 54.43 % of the classes (accuracy). Its bad performance actually derived from many target identification on non-target images. Furthermore, many correctly identified targets also had distorted skeletons. As for the **Testing** session images, the performance was even worse, achieving 45.12 % accuracy. This stems from not only having the human target appearing more distant on these images but also from detecting several distractor humans appearing in non-target images. This indicates the model does not generalize well to different search and rescue situations.

5.3.1.1 Subject-specific (P3)

Starting with the subject-specific models, a look at the P3 discrimination performance between Target and Non-Target stimuli is given in Figure 5.33 for the Train/Validation split. The average performance for each classifier architecture is presented in the image above (A) while the best classifier performance for each subject is given below (B).

Looking at the 5 model architecture performance comparison (A), the CV training results show a similar accuracy distribution collapsed across all 6 subjects, with an average accuracy of 69.41 % in which the Random Forest achieved the best performance of 70.97 %, closely followed by the XGBoost classifier. However, when looking at the average validation accuracy and AUC score per architecture (A), the SVM clearly dominates the correct prediction of target and non-target labels with 72.50 % and 0.77 for both metrics, respectively. Moreover, in the validation performance department, the LDA produced the second best results with an average AUC on the same level as the SVM (0.76) but fell short in terms of predicting the correct labels. For this reason, the SVM architecture was chosen as a clear winner to indicate the presence of a human target within the images by exploiting the P3 potential features.

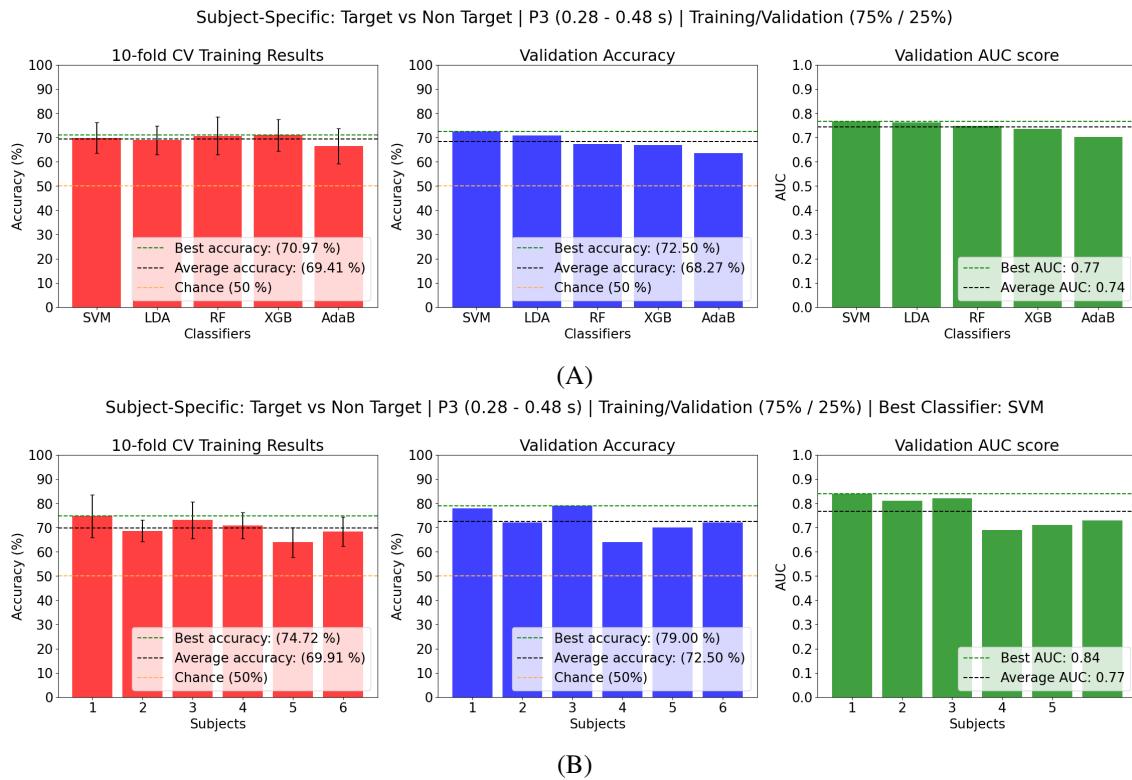


Figure 5.33: The Train/Validation results of the subject-specific P3 model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the SVM.

Looking at the individual subject performance with the SVM (B), subject 1 achieved the best CV training accuracy and AUC score out of all subjects, with a value of 74.72 % and 0.84, respectively. However, it was subject 3 that achieved the highest accuracy with 79 % of all trials correctly predicted, closely followed by subject 1 with 78 % accuracy. Additionally, with the exception of subject 4, all subjects achieved an accuracy above 70 % and an AUC above 0.7 on the validation set, with subjects 1, 2, and 3 providing an AUC above 0.8.

When testing the SVM architecture retrained with all **Training** session data with not only unseen data by the classifier but also unseen images by the brain (**Testing** session data), the overall testing performance dropped slightly, as depicted in Figure 5.34.

After retraining with all data using the best individual hyper-parameter combination, the CV accuracy results of the best subject (1) increased from 74.72 % to 76.04 % but the average CV accuracy across all 6 subjects dropped from 69.91 % to 69.25 %, accounted by the small drop in CV accuracy that occurred mainly in subject 4, whose performance dropped below the subject-average during the retrain but not initially. This may be explained by the division of trials done initially for the Train/Validation set in this subject. Perhaps, trials with lower cognitive load, on average, were assigned to the initial train set (75 %) and higher cognitive load target stimuli were left in the validation portion (25 %). This corroborates with the lower validation accuracy and AUC on this participant out of all subjects but achieving the 3rd best testing accuracy and AUC on the

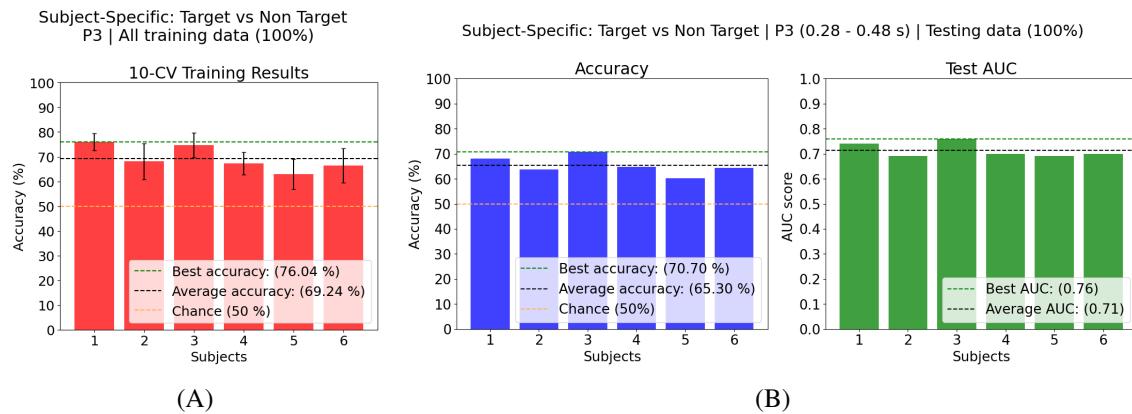


Figure 5.34: The training of the best subject-specific P3 model with all the **Training** session data (A) and model evaluation with all the **Testing** data session (B)

testing EEG data. Notably, this subject achieved an identical performance on both EEG recording sessions as indicated by both metrics, meaning the P3 was similar between both experiments for this individual. This hints at a good generalization performed by the brain on unseen environment backgrounds and human target characteristics, even though the overall performance is not great.

As for the other subjects, an average testing accuracy and AUC of 65.30 % and 0.71, respectively, were achieved, with the overall inter-subject performance trends being similar to the validation set but with lower values. In other words, subject 3 dominated the P3 classification performance, closely followed by subject 1 for both metrics. Furthermore, subjects 2, 4 and 6 accomplished similar results with subject 5 displaying the lowest performance of the cohort.

None of these subjects achieved the 0.8 AUC requirement for BCI application but subjects 1 and 3 did exceed the acceptable AUC value of 0.7, with the other participants almost reaching the later threshold with a minimum of 0.68. These results indicate that the P3 did not optimally discriminate both classes. However, the context of the experiment has to be considered. Three small details can explain this. Firstly, the time when the P3 arises in coincident with the SSVEP component, whose similar peak trend despite the amplitude differences may be influencing the classifiers negatively. Secondly, the cognitive load required to attend the target stimuli was high in general and no target was mutually displayed in the center of the image for the lateral and vertical visual fields. Thirdly and more importantly, the **Testing** session was always conducted after the **Training** one. In addition to the never seen images, participants visually searched 2400 images before being displayed the **Testing** image dataset. RSVP sessions have been linked to lapses of attention due to fatigue after a while, which directly influences the ERPs derived from target identification in visual search and subsequently, participants' performance [33].

Figure 5.35 demonstrates the cohort performance across each trial block of 16 target images. It can be seen that individuals significantly vary the performance across the trial blocks, potentially describing visually the lapses of attention for each participant in each specific trial block. With the exception of subject 5, all subjects were able to achieve accuracies above 70 % in certain trial blocks, with subject 3 achieving 87 % on the second trial block. However, note that no specific

analysis was done to objectively evaluate these affirmations and other subtle factors may influence the subjects' performance in each trial block, including the classifier design. Interestingly, with the exception of participant 4, most participants increase the P3 discriminability from trial block 1 to trial block 2, hinting at the brain adaptation to unseen images.

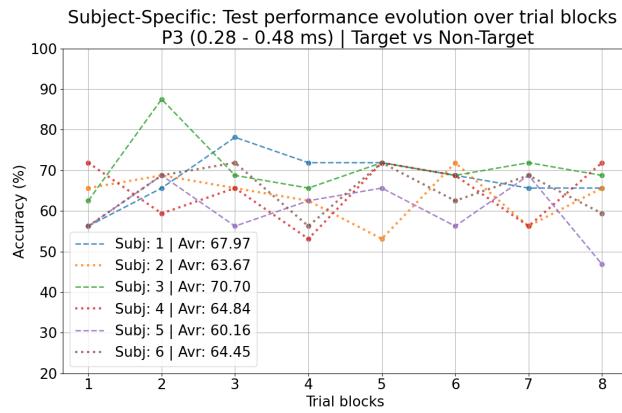


Figure 5.35: P3 performance in discriminating Target from Non-target stimuli over the 8 trial blocks.

Comparing these results to the best studies in the literature using the P3 on BCI-based RSVP paradigms for target vs non-target discrimination, an AUC of 0.856 [76] and 0.876 [16] has been achieved before. However, these studies have all different visual stimuli presentation details that do not allow direct performance comparisons. The former displays 87 targets out of 216 between the left and right 1.2° visual angle threshold, meaning these appear around the center of the image. Here, only 40 and 16 targets appeared within this region for the **Training** and **Testing** paradigm, all above or below the vertical 1.2° visual angle threshold, respectively, posing a higher degree of target attending difficulty. Moreover, in the study where an average AUC of 0.856 [76], only 24 target stimuli were used during the testing phase, whereas 128 were displayed here, increasing the probability of an increased number of misclassifications. Additionally, the target objects displayed here are much smaller than the ones presented in the study and are not shown in greyscale format. Finally, the study used 20 channels to perform the target vs non-target classification for the P3 ERP while 5 were used here, being much more computationally efficient.

As for the later study, no train/split information was provided [16]. More importantly, the brain generalization targeted here was not attempted in these papers, meaning that the testing performance displayed here does not only provide the model performance to unseen EEG data but also to unseen images with different visual properties by the brain, tremendously increasing the task difficulty.

Regarding the portion of trials correctly classified as Target (1) and Non-Target, the specificity and sensitivity results for this classifier are given in Table 5.2. On average, the sensitivity achieved was 0.67 with subject 3 providing the highest amount of Target trials correctly classified (76 %). As for the specificity, the subject average was 0.64 with subject 1 displaying the best performance with 71 % of non-target trials correctly classified.

Table 5.2: The sensitivity and specificity of the subject-specific P3 target vs non-target model. The chance level is 50 %.

	S1	S2	S3	S4	S5	S6	Mean
Sensitivity	0.65	0.66	0.76	0.67	0.66	0.64	0.67
Specificity	0.71	0.61	0.66	0.63	0.55	0.65	0.64

5.3.1.2 Subject-Generalized (P3)

Despite not being impressive, the subject-specified P3 model for target vs non-target discrimination results did not vary significantly and a subject-generalized model was attempted for this classification task. The train/validation (A) and testing (B) results are displayed in Figure 5.36.

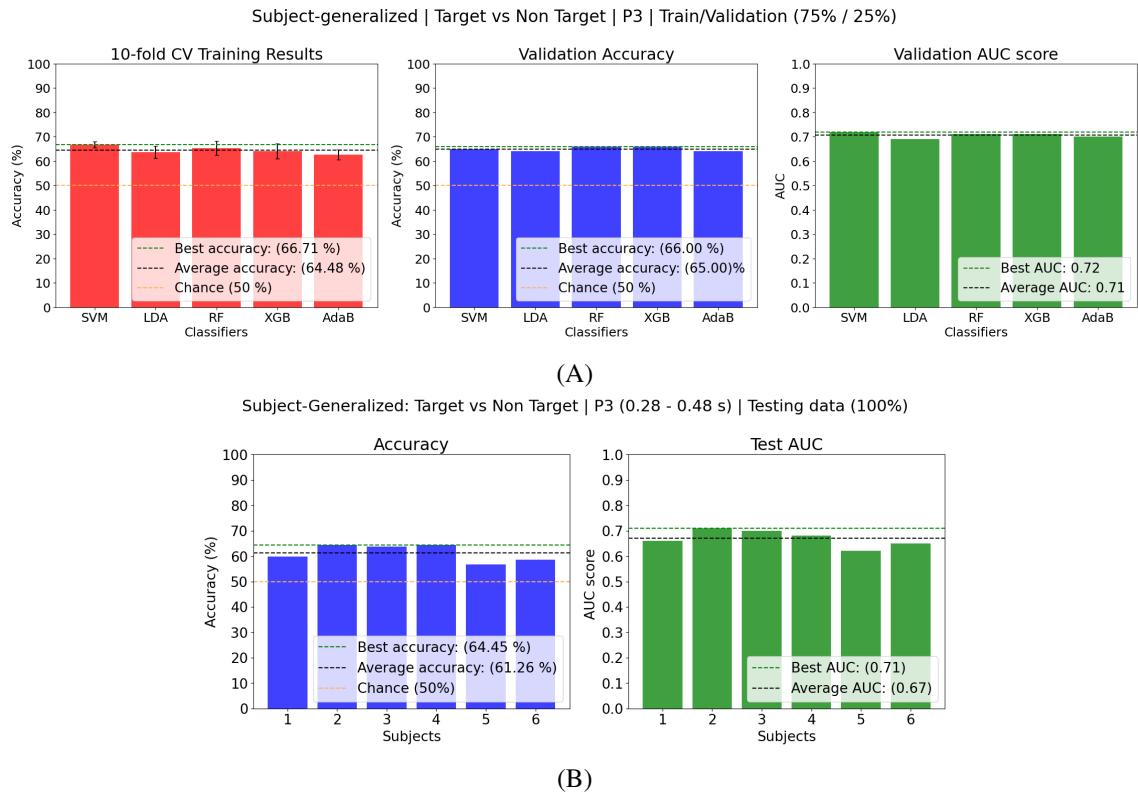


Figure 5.36: The Train/Validation results of the subject-generalized P3 model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the SVM architecture and the model retrained with the **Training** data session achieved 67.4 % CV accuracy with 1.66 % std.

For this classifier, the SVM achieved the best subject-average CV results with 66.71 % accuracy and best AUC (0.72) but not the validation accuracy, obtaining 64.75 % accurate class predictions (A). For this metric, the Random Forest was able to achieve the best validation accuracy of 66 %, closely followed by the XGBoost (65.77 %). However, the combined higher 10-fold CV accuracy and AUC of the SVM architecture hints that this model provides better generalization to unseen data. These three classifiers were retrained with all **Training** session data and their performance was compared with the **Testing** data session.

Indeed, the SVM was able to perform better on the unseen data, resulting in a subject-average CV accuracy of 67.4 %, a testing accuracy of 61.26 %, and a 0.67 AUC (Figure 5.36 (B)). Despite the concomitant low results and the challenge of constructing a hyperplane that would work on different P3 properties associated with each individual, the average results were only slightly lower than the subject-specific averages. However, in contrast to the subject-generalized, the subject-specific results varied more between subjects, allowing for those subjects who produced a more discriminatory P3 to achieve better results. In these subject-generalized models, a subject whose data is not as discriminatory can hurt the ability of the model to learn to separate the classes for all individuals, hence the lower average results.

Interestingly, the best and second best performances obtained were not for participants 3 and 1 but for subjects 2 and 4, respectively. Subject 2 achieved a 64.45 % testing accuracy and 0.71 AUC, exceeding the acceptable AUC threshold through this model. However, this was the only subject reaching this level with all others falling behind. Because this subject-generalized P3 model did not yield relevant results when compared to the subject-specific counterpart, the trial block performance was not plotted.

As for the subject sensitivity and specificity on the testing data, the averages were also lower than subject-specific models, producing values of 0.61 and 0.62, respectively, as represented in Table 5.3. Notably, subject 3, which produced the best subject-specific P3 model for this discrimination task, achieved 80 % of correctly labeled targets, which was even better than the model built only on his data (0.76 sensitivity). However, its overall lower testing accuracy derived from the lowest specificity of all subjects, with 47 % of non-target stimuli being classified as target. In terms of correctly classifying most instances of the non-target through this model, subject 2 achieved the best results (70 %) but missed most target class instances with a sensitivity of 0.59.

Table 5.3: The sensitivity and specificity of the subject-generalized P3 target vs non-target model. The chance level is 50 %.

	S1	S2	S3	S4	S5	S6	Mean
Sensitivity	0.64	0.59	0.80	0.55	0.57	0.52	0.61
Specificity	0.56	0.70	0.47	0.74	0.56	0.67	0.62

5.3.1.3 Subject-specific (N2pc)

In addition to the P3 potential, the N2pc was exploited in an attempt to see if the target vs non-target discrimination could be done earlier on the visual processing stream. This classification was attempted in two feature vector variations, namely one with all 14 channels placed on the parieto-occipital region of the scalp where this potential emerges and another by selecting the 7 most discriminative electrodes with MBI. Starting with the latter, Table 5.4 summarizes the channel selection for each participant.

Out of the 14 channels, the PO7 and PO8 are common to all subjects as two of the most discriminative channels selected, corroborating with the literature for where in the scalp the N2pc reaches its maximum discriminating ability [19]. Furthermore, with an exception for subjects 5

Table 5.4: The 7 most discriminative channels to classify target vs non-target stimuli through the N2pc ERP.

	S1	S2	S3	S4	S5	S6
Left hemisphere	PO7, PO5, P3	PO7, PO3, PO5	PO7, PO5, P5	PO7, PO5, P5	PO7, PO5, P3, O1	PO7, PO5, P3, P5
Right hemisphere	P8, PO8, PO4, PO6	P8, P4, P6, O2	P8, PO8, PO4, O2	PO8, PO4, PO6, P6	PO8, PO4, P6	PO8, P4, P6

and 6, the right hemisphere displays most of the best channels in comparison to the left hemisphere, with MBI perhaps picking the LVFA described previously.

Looking at the classifiers' performance with this selection, the LDA architecture was the clear winner with a subject-specific average of 67.83 % validation accuracy and AUC of 0.76, as illustrated in Figure 5.37 (A). Regarding the subject-specific performance in the Training/Validation set, subject 1 was the only one achieving an accuracy above 70 %, providing 79 % of correctly predicted labels with this model (B). However, in the AUC department, all subjects reached the 0.70 AUC acceptable threshold with subject 1 achieving the highest score of 0.87.

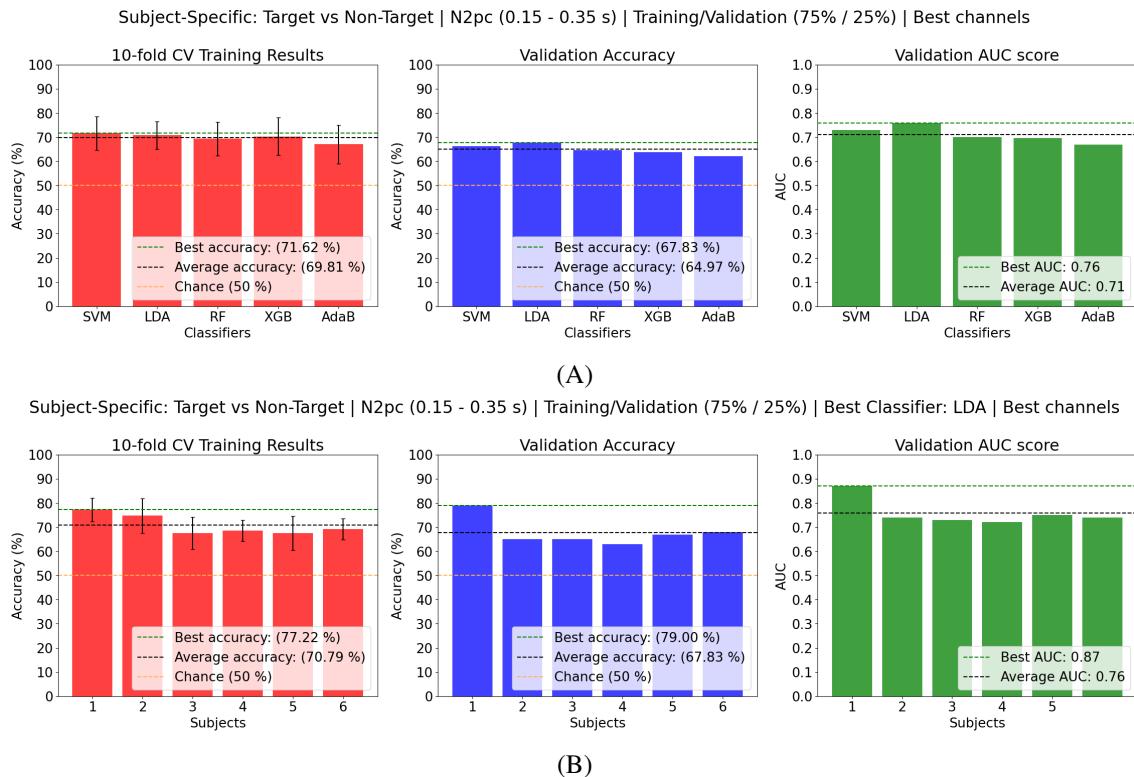


Figure 5.37: The Train/Validation results of the subject-specific N2pc model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found for the 7 best channels was the LDA.

After retraining the optimal subject-specific hyperparameter LDA models with all **Training** session data and computing the performance metrics on **Testing** experiment data, the results indicate a high degree of generalization on unseen EEG data and surpassing the P3 performance,

despite the N2pc generating a much smaller amplitude (Figure 5.38). Notably, even without exacerbating the hemispheric asymmetries by not computing the standard contralateral minus ipsilateral difference and by just feeding the classifier the separated channel pairs of data, this temporal information was more discriminatory than the latter occurring P3, with a subject testing accuracy average of 66.28 % and an AUC of 0.73. More importantly, the subject-specific averages between the train/validation and testing sets were identical.

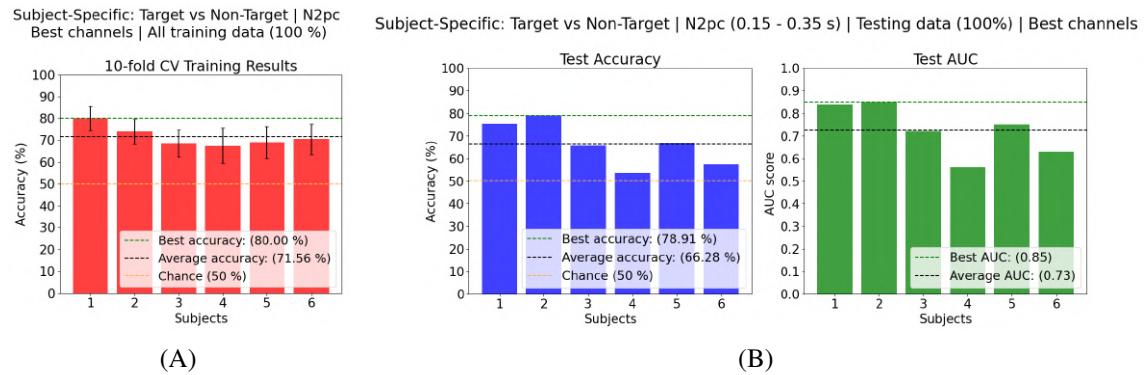


Figure 5.38: The training of the best subject-specific N2pc model with all the **Training** session data (A) and model evaluation with all the **Testing** data session (B). The 7 most discriminative channels were used.

Notably, when looking at the individual testing performances, the highest results were achieved by subject 2 with 78 % of accurately predicted class labels and an AUC of 0.85, which is much greater than the one obtained during the validation set where an accuracy of 65.05 % and a 0.71 AUC was obtained. Closely behind subject 2 is subject 1 whose performance was only tenuously lower on the testing data set than on the validation set, with the accuracy and AUC metrics falling from 79 % and 0.87 to 75.39 % and 0.84. Additionally, subjects 3 and 5 achieved very similar results in both data sets. It is possible to infer from this that the brain, through the N2pc information, was able to generalize the background environment and target visual properties between both sessions. This is especially relevant when considering that these subject-specific models are comparing 60 trials of each class on the validation set with 128 trials per class on the testing set, where there is more opportunity for misclassifications, making subject's 2 performance even more impressive.

However, due to the low amplitude nature of the N2pc elicitation, especially when the channel pair contralateral minus ipsilateral calculation is not performed, the N2pc produces more performance variability between subjects. This refers to subjects 4 and 6 whose performance dropped significantly between both experiments. Nevertheless, subjects 1 and 2 achieved a greater AUC value than the threshold of 0.8 "defined" for a correct BCI control, with subjects 3 and 5 exceeding the acceptable 0.7 AUC threshold.

Figure 5.39 depicts the 8 ordered 8 trial block accuracy performance for the cohort studied. It can be seen that subject 2 displayed a very constant performance throughout the **Testing** experiment through the LDA model. Notably, subject 1 displays a lot of changes in performance, reaching a

maximum accuracy of 85 % in trial block 4 but also exhibiting trial blocks where the model falls below 70 % accuracy. Interestingly, the trend of participants increasing their performance from trial block 1 to trial block 2 is not as present here. This may be attributed to the neural correlates between the N2pc and P3 being different, with the N2pc temporal information mostly capturing some object of interest within the visual field before it is actually categorized as the human target, even on the first trials. In contrast, the P3 is elicited when the human target is identified by its appearance matching its corresponding representation in the working memory. However, in the **Testing** experiment, the human target is different from the **Training** paradigm, which may delay the brain's ability to elicit a discriminatory P3 in the first instances of its appearance.

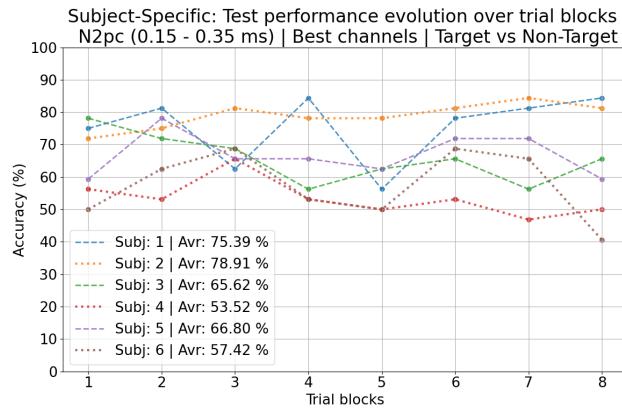


Figure 5.39: Subjects performance for the best 7 N2pc channels in discriminating Target from Non-target stimuli over the 8 trial blocks.

Table 5.5 depicts the sensitivity and specificity results obtained through this model. The subject average specificity is much higher than on the models discussed previously but not the sensitivity, where participants 4 and 6 exhibited chance-like classification performance. However, subjects 1 and 2 lead this model benchmark with 0.71 and 0.86 sensitivity, respectively, and 0.80 and 0.72 specificity, also respectively.

Table 5.5: The sensitivity and specificity of the subject-specific N2pc target vs non-target model for the most discriminative 7 channels. The chance level is 50 %.

	S1	S2	S3	S4	S5	S6	Mean
Sensitivity	0.71	0.86	0.63	0.41	0.60	0.51	0.62
Specificity	0.80	0.72	0.68	0.66	0.73	0.63	0.70

Comparing now the best 7 channel model performance with a doubled feature classifier that included all 14 channel data, the best classifier handling the N2pc data information (LDA) and the inter-subject performance trends were maintained between both classifications for the train/validation set, as shown in Figure 5.40. In fact, the subject averages for the validation accuracy and AUC increased slightly but not significantly. Interestingly, subject 1 maintained the exact same ratio of correctly predicted class labels to incorrectly predicted ones (79 %) but the AUC dropped 0.01 points. Subject 6 on the other, was the only subject that benefited from the increased channel information provided during the model training.

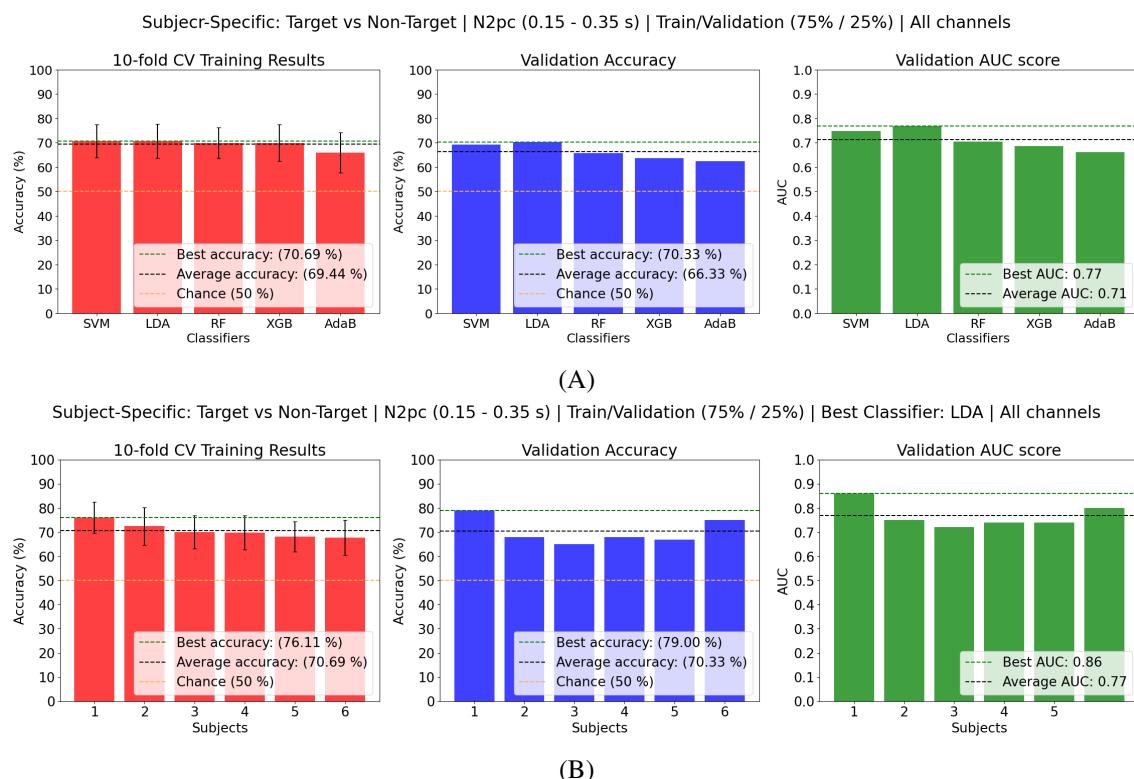


Figure 5.40: The Train/Validation results of the subject-specific N2pc model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the LDA.

With regards to the model evaluation with the **Testing** data session, very similar results were obtained, with the subject averages increasing non-significantly, as depicted in Figure 5.41. Despite the increased performance with all channels for subject 6 in the validation set, this change was not seen in the testing data. Subjects 1 and 2 continued to lead the subject-specific N2pc target vs non-target classification with testing accuracies and AUC above the 70 % and 0.80 range, respectively. Overall, the model combining all 14 channels in the parieto-occipital region performed identically to the model using the 7 most discriminatory channels, motivating the use of the latter in the final deployment phase where less data is advantageous for faster computations.

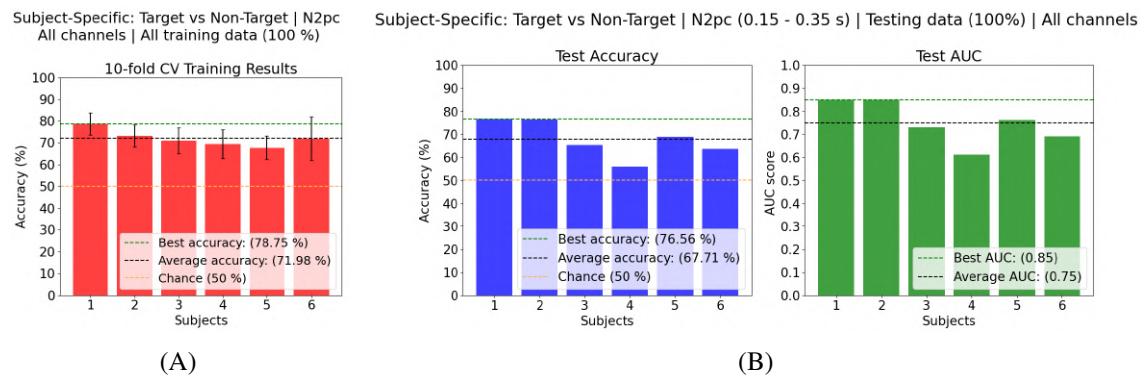


Figure 5.41: The training of the best subject-specific N2pc model with all the **Training** session data (A) and model evaluation with all the **Testing** data session (B).

No studies have been found in the literature solely using the N2pc component information in the *oddball paradigm* to discriminate between target and non-target stimuli. However, Matran-Fernandez and Riccardo Poli demonstrated that by combining data from the N2pc and P3 channels, the average AUC from 11 participants increased from 0.856 when solely using P3 channels to 0.873 [76]. Here, by solely using the N2pc not only better discrimination between the target and non-target stimuli instances was observed relative to the P3 component, but also because the N2pc emerges earlier on the visual processing stream, this classification can be done faster, specifically in the first 350 ms relative to stimuli onset. Despite the lower AUC when compared to the one obtained in the paper [76], it is important to note that here, half of the channels were used for discrimination relative to the study ($n = 28$) and no P3 channel information was included in this model.

When plotting the model performance through the 8 testing trial blocks, similar subject trends were observed as illustrated in Figure 5.42. Subject 1 displays the same ups and downs while reaching its maximum performance over the 4th trial block, while subject 2 exhibited a more constant performance throughout the experiment. Furthermore, subject 6 achieved a below-chance classification on the last trial, probably due to fatigue and lapses of attention. The lowest overall performance was achieved in both models for the 4th subject, having 3 trial blocks where only 50 % of stimuli were correctly classified.

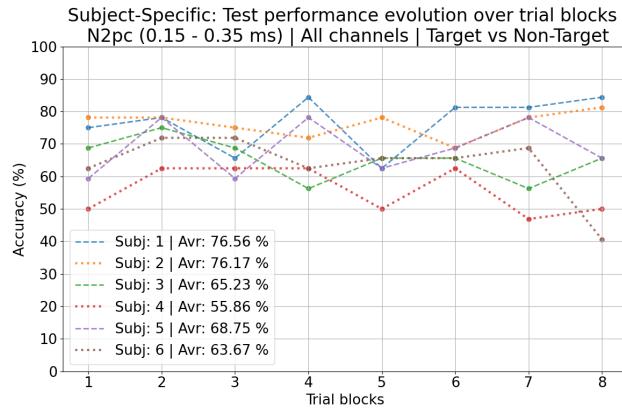


Figure 5.42: Subjects performance for all 14 N2pc channels in discriminating Target from Non-target stimuli over the 8 trial blocks.

In Table 5.6, the sensitivity and specificity achieved with this LDA model are shown. Albeit providing a slight increase in the overall subject average for both metrics relative to the lower feature vector-sized model, the LDA model based on the 7 most discriminative channels provided a more balanced proportion of correctly classified Targets (1) and Non-Targets (0), mainly on subjects 1 and 2, whose performance dominated the cohort. For now, the optimal channel selection model exploiting the capture of the N2pc component displayed the best cost-efficient discrimination of target instances with a more equilibrated sensitivity and specificity on the best performers.

Table 5.6: The sensitivity and specificity of the subject-specific N2pc target vs non-target model for all 14 channels. The chance level is 50 %.

	S1	S2	S3	S4	S5	S6	Mean
Sensitivity	0.69	0.84	0.59	0.49	0.63	0.59	0.64
Specificity	0.84	0.69	0.72	0.63	0.74	0.68	0.72

5.3.1.4 Subject-generalized (N2pc)

Despite the N2pc LDA models providing more subject-specific performance variation than the models handling the P3 potential, a generalized N2pc model was still attempted. The training of the combining subject data is displayed in Figure 5.43 (A) while the test performance from the best model, remaining the LDA architecture for the N2pc data, is displayed in subfigure (B).

As is shown, the model provided a good generalization for subjects 1 and 2, both achieving testing accuracies of 70.70 %. Furthermore, both achieved an AUC above the 0.7 threshold, with 0.79 for subject 1 and 0.78 for subject 2. Subjects 4, 5, and 6 also exhibited a similar ratio of correctly predicted trials to incorrectly predicted ones as the subject-specific models, despite the former (S4) always displaying an accuracy below the acceptable range for all N2pc models discussed. Moreover, subjects 5 and 6 were able to surpass the 0.70 AUC threshold, albeit being much closer to this range than subjects 1 and 2. Interestingly, subject 6 did not exhibit an AUC above the acceptable range in neither of the subject-specific N2pc models but did here.

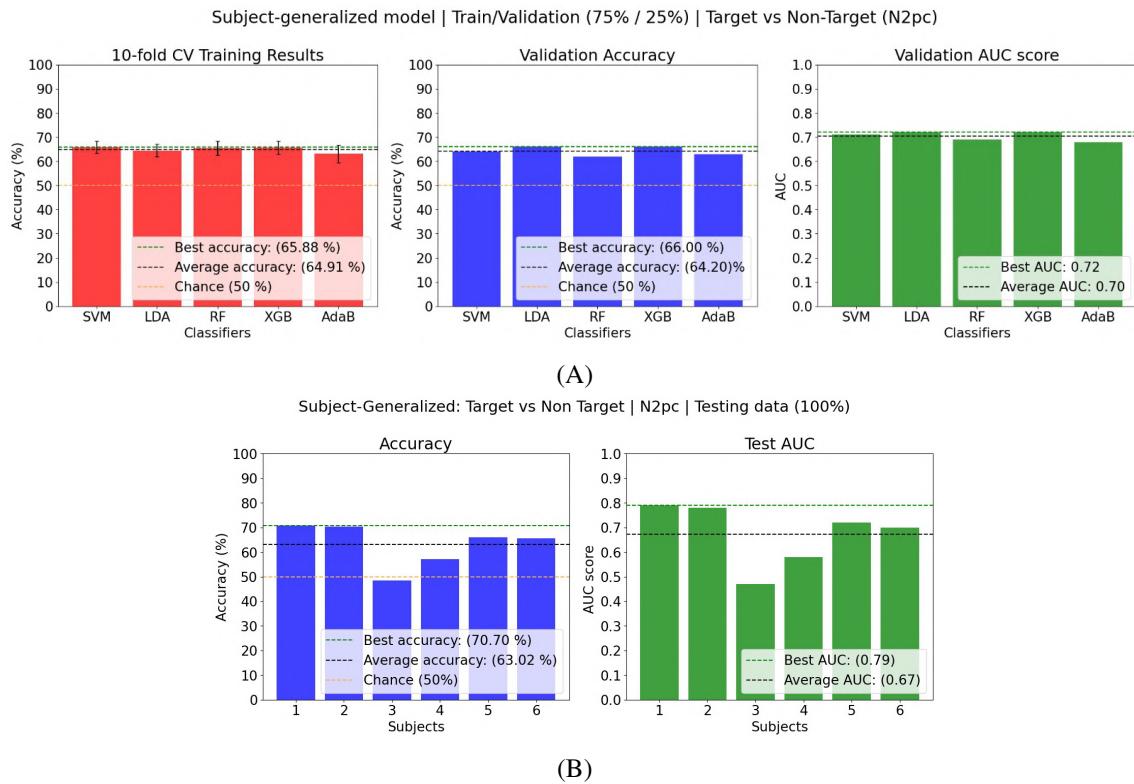


Figure 5.43: The Train/Validation results of the subject-generalized N2pc model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the LDA architecture and the model retrained with all **Training** data session achieved 65.83 % CV accuracy with 2.24 % std.

Despite an average testing accuracy of 63.03 % not being much lower than the subject-specific model average, the average AUC decreased significantly to 0.67. This drop in average performance is attributed to subject 4 but mostly to subject 3, whose performance was much lower with the generalized model than on both subject-specific LDA channel model variations. This indicates that the N2pc waveform generated from this subject has substantially different properties than the N2pc of other subjects.

Looking at Table 5.7 where the sensitivity and specificity achieved through this model are summarized for all subjects, the subject's 1 ratio of correctly predicted target stimuli was the highest among its dedicated subject-specific models (0.80). However, its specificity dropped significantly to 0.69, being the lowest compared to its dedicated subject-specific model. Subject 2 also displayed a good rate of correctly predicted target stimuli but only accurately predicted the non-target class 61 % of the time, lower than its subject-specific models. Furthermore, from this table it can be seen that the high decrease in performance through this subject-generalized model for subject 3 arises from its extremely low sensitivity, achieving a value of 0.36. Finally, subject 6 also displayed the highest amount of correctly predicted non-target stimuli with this model when compared to models trained specifically with this subject data but sensitivity decreased, whose highest value for now coming from the subject-specific 14 channel LDA model, as seen in Table

5.6. Overall, this subject-generalized model was able to produce similar sensitivities on average but suffers from producing lower specificity values than the subject-specific variations.

Table 5.7: The sensitivity and specificity of the subject-generalized N2pc target vs non-target model for all 14 channels. The chance level is 50 %.

	S1	S2	S3	S4	S5	S6	Mean
Sensitivity	0.80	0.79	0.36	0.55	0.66	0.55	0.62
Specificity	0.61	0.61	0.61	0.59	0.66	0.77	0.62

5.3.1.5 Subject-specific (N2pc and P3)

In an attempt to combine the best features of the N2pc and P3 potentials, a model was designed to select the 9 most discriminative channels out of the 19 provided by both ERPs. Table 5.8 summarizes the output of the MBI channel selection for this discriminative task.

Table 5.8: The 9 most discriminative channels in classifying target vs non-target stimuli for the combined N2pc and P3 ERPs.

	S1	S2	S3	S4	S5	S6
N2pc - Left hemisphere	PO7, PO5	PO7, PO5, P5	PO5	PO5, P5	PO7, PO5, P5	PO7, PO5, P5
N2pc - Right hemisphere	PO8, P4	P8, PO4, P4	P8, PO8, PO6	PO8, PO4, O2	P8, PO6, P4	P8, PO8, PO6, P4
P3	Cz, Cpz, Pz, CP1, CP2	CPz, CP1, CP2	Cz, CPz, Pz, CP1, CP2	Cz, CPz, CP1, CP2	Pz, CP1, CP2	Cz, CPz

As is shown, out of all channels, the PO8 and PO7 remain some of the most informative N2pc channels selected for the target vs non-target classification problem across the cohort. Unexpectedly, however, more P3 channels hold the discriminative ability than the N2pc ones to discriminate both classes, with the exception of subject 6. Perhaps, this can be attributed to the period of 200 ms where these ERPs were segmented in the temporal dimension. Even after normalizing the feature vectors, because of its overall size and temporal width in comparison to the N2pc, the P3 has an increased influence over this 200 ms time window than the N2pc. However, it is important to note that this period was selected to encapsulate both the N2pc and P3 single-trial latency variation occurring due to the different cognitive loads associated with the specific distance from the image center where each human target appears, a characteristic invisible on the grand average plots.

When comparing all model architectures, the SVM provided the best subject-average accuracy (76.67 %) and AUC (0.84) on the validation set out of all classifiers tested so far, as depicted in Figure 5.44 (A). Looking at the subject-specific performance (B), subject 1 achieved an accuracy of 87 % and an AUC 0.94 with all participants being able to obtain an AUC equal or above the BCI AUC requirement (0.8). Furthermore, these results are comparable to some of the best AUC accomplished in the literature when combining both ERPs during an RSVP *oddball paradigm*. As already pointed out, an AUC of 0.876 has been achieved in another study [76]. An important consideration when comparing these is that 28 channels were used to perform the classification on the paper whereas only 9 were used here.

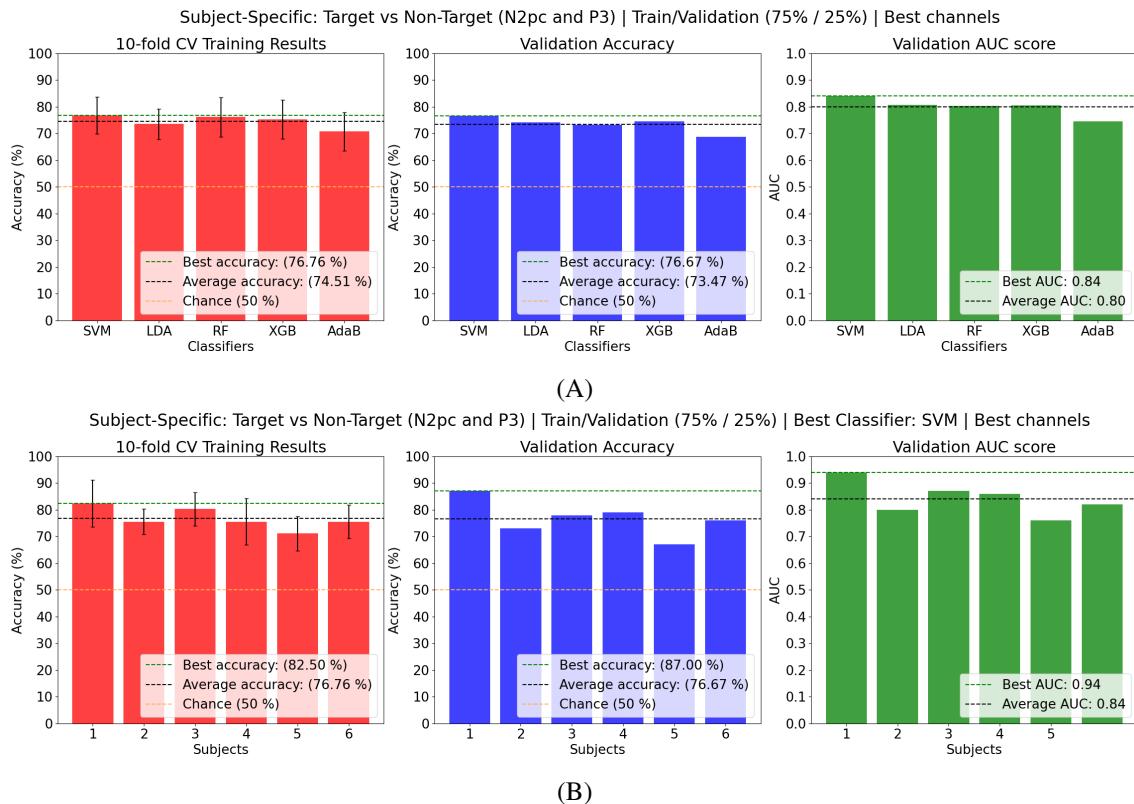


Figure 5.44: The Train/Validation results of the subject-specific N2pc-P3 combined model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found for the 9 best channels was the SVM.

However promising, these models were able to generalize as well on the unseen data, as shown in Figure 5.45. The subject average testing accuracy achieved was 67.71 %, roughly the same as the one provided by only using the N2pc component with all 14 channels. Subject 1 was the one who suffered the most drop in performance in the testing data set when compared to the validation set decreasing from 87 % accuracy to 66.41 %. However, this participant's AUC reached the acceptable range of 0.78. The subjects who dominated this model were individuals 2 and 3, accurately predicting 73.44 % and 74.44 % of the classes. Furthermore, both reached an AUC value of 0.83 and 0.82, respectively. As for the rest of the individuals, only subject 4 achieved an AUC below the acceptable threshold.

When looking at the performance over the 8 trial blocks displayed in the **Testing** experiment in Figure 5.46, similar performance trends are observed as the ones described previously for other models. Most individuals share big variations in performance throughout the experiment. In fact, because the mean is much more sensible to outliers than the median, even the best subjects' performance, *i.e.*, 2 and 3, appear low due to 2 trial blocks where their performance drops significantly. This is the limitation of displaying the average performance for such a large number of trials ($n = 128$ for each class), occurring for all models discussed.

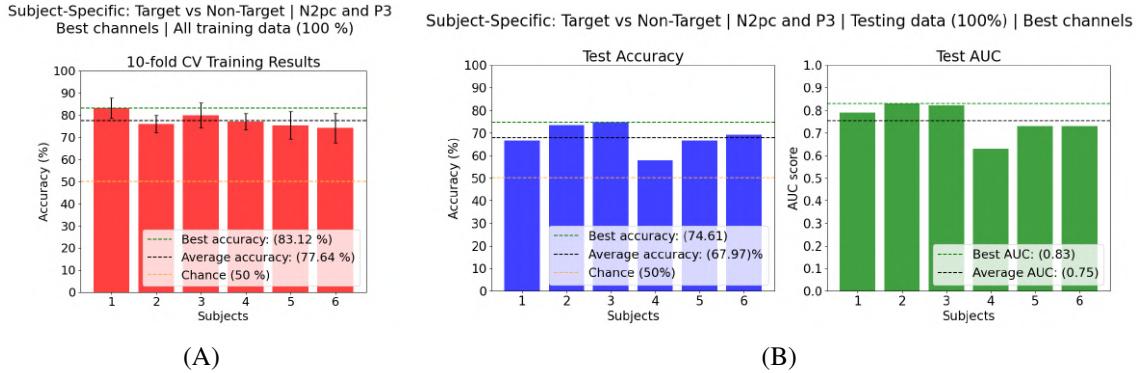


Figure 5.45: The training of the best subject-specific N2pc-P3 model with all the **Training** session data (A) and model evaluation with all the **Testing** data session (B). The best 9 channels were used for each participant.

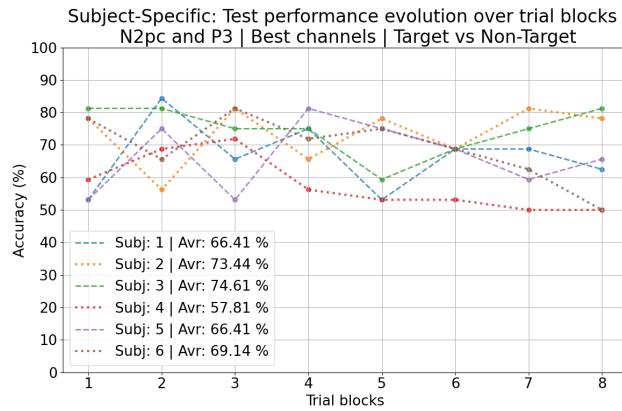


Figure 5.46: Subjects performance for the best 9 N2pc-P3 channels in discriminating Target from Non-target stimuli over the 8 trial blocks.

Despite the low drop in performance from the validation data set to the testing one, combining both ERP information to distinguish targets from non-targets provided the best sensitivity results out of all models presented so far, as summarized in Table 5.9. However, in comparison to the N2pc models, the specificity achieved was lower (0.65).

Table 5.9: The sensitivity and specificity of the subject-specific N2pc-P3 target vs non-target model for the best 9 channels. The chance level is 50 %.

	S1	S2	S3	S4	S5	S6	Mean
Sensitivity	0.62	0.74	0.83	0.68	0.71	0.70	0.71
Specificity	0.71	0.73	0.66	0.48	0.62	0.68	0.65

Comparing now the model combining both ERP data but with all 19 channels, the SVM still was the dominant architecture, as shown in Figure 5.47 (A). However, using all channels relevant for both ERP analyses lowered the train/validation set results where a subject average accuracy of 75 % and AUC of 0.80 was achieved.

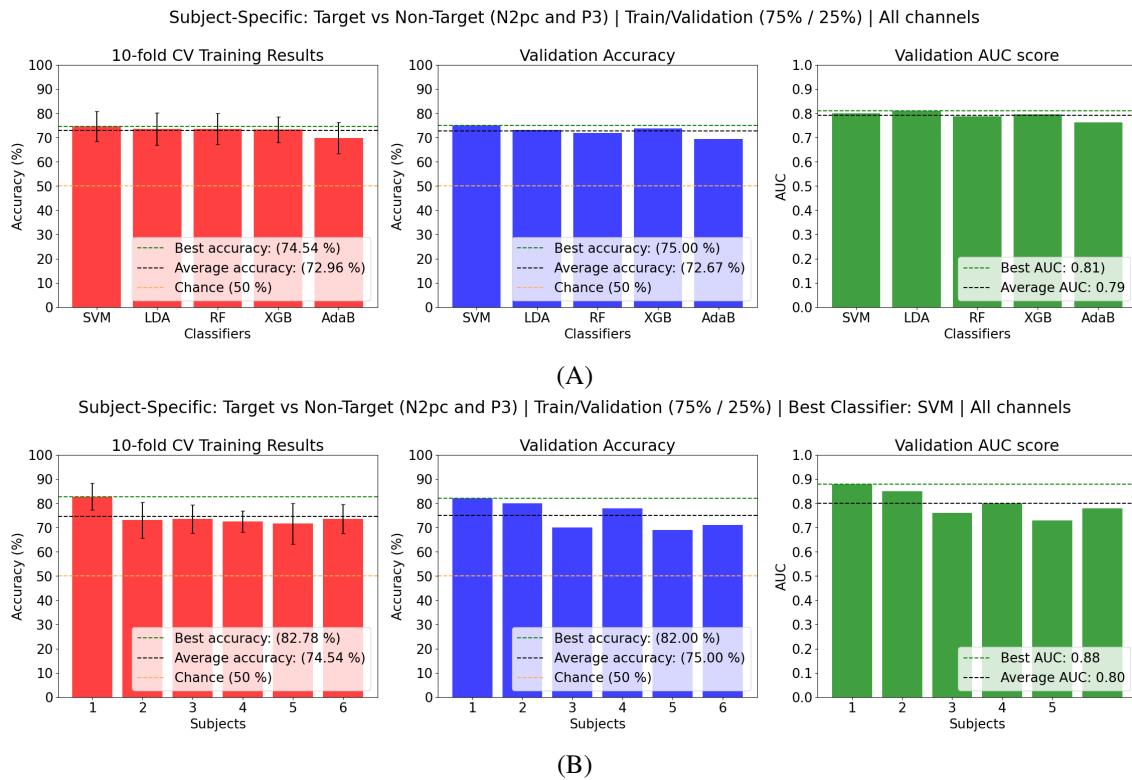


Figure 5.47: The Train/Validation results of the subject-specific N2pc-P3 combined model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found for all 19 channels was the SVM.

The performance obtained for the testing set yielded a similar trend with the overall performance dropping slightly from the models trained and tested with the best 9 channels (Figure 5.48). This reveals that adding all channels relevant to both ERPs is not necessary. Subjects 2 and 3 still dominated the performance obtained, both achieving an accuracy and AUC above the 70 % and 0.8 thresholds, respectively.

Despite not generalizing as well to unseen images by the brain when compared to the N2pc-based models, the good results achieved on the validation set for the merged N2pc-P3 models indicate that training subjects on similar backgrounds and target characteristics works very well and accomplishes results comparable to the literature. Effectively, the images presented throughout each trial block during the **Training** session are the same, although being randomly ordered each time, but the EEG trial data fed to the classifiers in the validation set is not the same as the EEG data used to train the models. This means that the models are validated on unseen EEG data in this phase too, as is done in the literature [20, 76]. On the testing set, an extra step is walked and the models also attempt to generalize not only unseen images by the brain but also images with different qualities, and subsequently slightly different ERP waveforms, offering another level of difficulty. However, because no study approaches the brain generalization to different visual stimuli properties in the way attempted here, no direct comparison for the **Testing** phase can be performed, especially when considering the high amount of trials in which these models are tested,

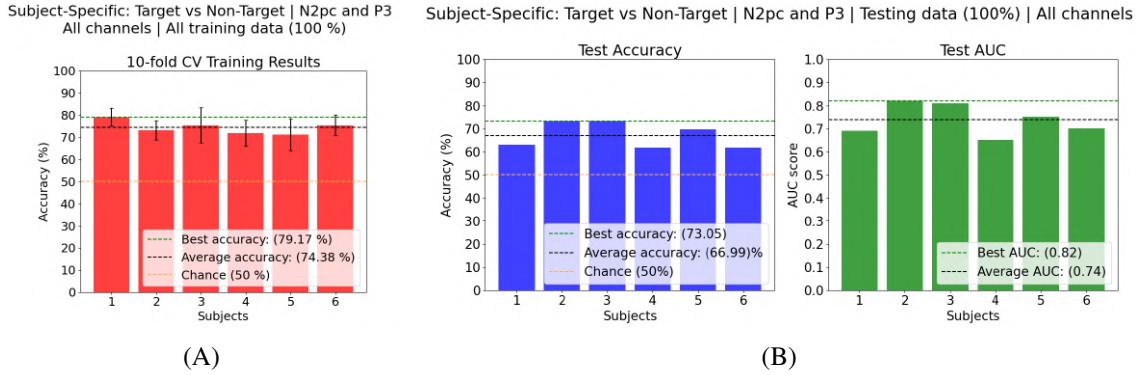


Figure 5.48: The training of the best subject-specific N2pc-P3 model with all the **Training** session data (A) and model evaluation with all the **Testing** data session (B).

thus offering novel research in this area.

Figure 5.49 reveals the cohort performance over the 8 trial blocks presented during the **Testing** session with the N2pc-P3 model using all 19 channels. The performance graph is yet again very similar to others already discussed. However, it needs to be pointed out that subject 2 achieves a 90 % accuracy in discerning both target and non-target classes during trial block 3 with this model.

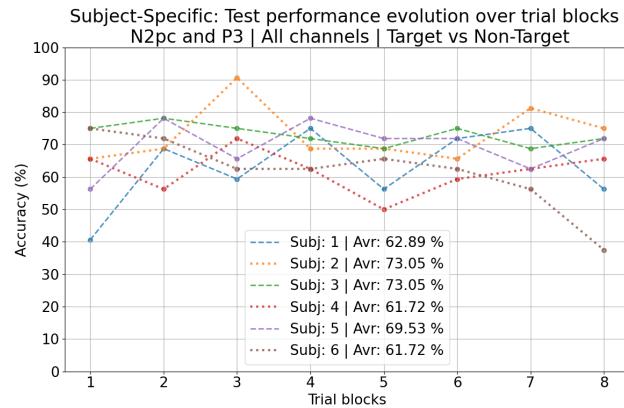


Figure 5.49: Subjects performance for the N2pc-P3 model combining all 19 channels in discriminating Target from Non-target stimuli over the 8 trial blocks.

Table 5.10 summarizes all sensitivity and specificity results for the N2pc-P3 model. Using all 19 channels improves the already topped sensitivity results provided by the best 9 channel selection model (0.71), here achieving an average value of 0.73. However, the lower overall performance exhibited by this model arises from a drop in the specificity metric, suggesting that this model is not able to accurately predict the non-target class in general.

Table 5.10: The sensitivity and specificity of the subject-specific N2pc-P3 target vs non-target model for all 19 channels. The chance level is 50 %.

	S1	S2	S3	S4	S5	S6	Mean
Sensitivity	0.60	0.84	0.80	0.73	0.75	0.64	0.73
Specificity	0.66	0.62	0.66	0.51	0.64	0.59	0.61

5.3.1.6 Subject-generalized (N2pc and P3)

As expected from observing the subject-specific Train/Validation set results, the subject generalized model combining both ERPs yielded the best results out of all models in **Training** experiment. As illustrated in Figure 5.50 (A), the best classifier architecture, the SVM, was able to correctly predict 79 % of all trials in the validation set while reaching an AUC of 0.88, on average for all 6 subjects.

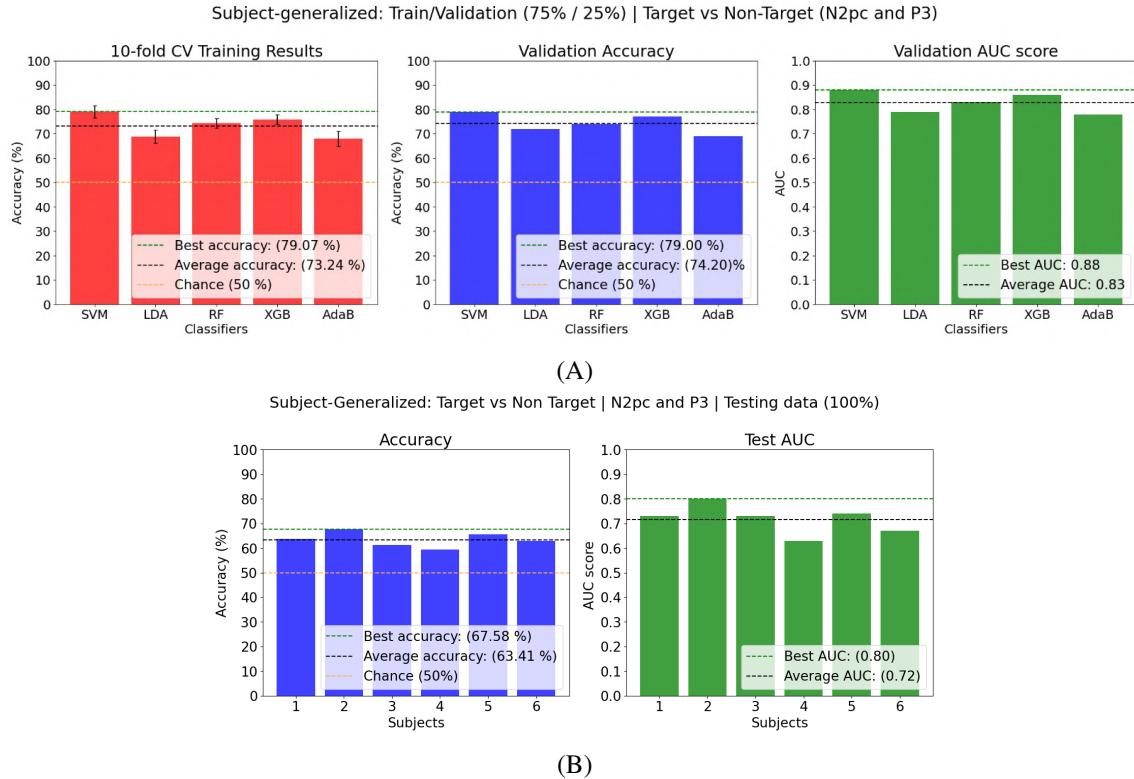


Figure 5.50: The Train/Validation results of the subject-generalized N2pc-P3 model to discriminate Target vs Non-Target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the SVM architecture and the model retrained with all **Training** data sessions achieved 88.41 % CV accuracy with 2.28 % std.

In the **Testing** session data, the performance drop exhibited on the subject-specific N2pc-P3 models was also similar (B). However, the inter-subject performance was much more balanced than on the N2pc generalized model, putatively attributed to the presence of the P3 channel data. The testing accuracy and AUC results were not impressive, with a subject average of 67.58 % and 0.72, respectively. Nevertheless, an AUC of 0.8 was achieved for subject 2 through this model. Additionally, three other subjects, namely participants 1, 3, and 5 were able to surpass the 0.7 AUC threshold.

The low accuracy performance achieved with this SVM model resulted from the very low specificity of 0.48 on average, as represented in Table 5.11. However, the sensitivity here obtained yielded the best results so far, with a subject average of 0.79. Furthermore, through this model, subject 3 target trial data was correctly predicted 89 % of the time.

Table 5.11: The sensitivity and specificity of the subject-generalized N2pc-P3 target vs non-target model for all 19 channels. The chance level is 50 %.

	S1	S2	S3	S4	S5	S6	Mean
Sensitivity	0.79	0.81	0.89	0.72	0.76	0.76	0.79
Specificity	0.48	0.54	0.34	0.47	0.55	0.5	0.48

5.3.2 Sensitivity and Specific: Optimal model combination for target and non-target discrimination

It was shown that the Human Pose Estimation model yielded poor results in comparison to the BCI proposed here. However, the latter approach can actually be improved. All these target vs non-target model variations presented above can actually be used in parallel to fill each other gaps and merge their strengths. For instance, the subject-generalized N2pc-P3 model turned out to be specifically optimized to accurately predict a target appearance due to its high sensitivity. Despite suffering from low specificity across subjects and misclassifying the non-target class more regularly, other models were able to achieve good specificity but not sensitivity and these can, therefore, be used to more accurately predict the non-target class while the subject-generalized N2pc-P3 model handles the target stimuli identification.

For instance, in subject 1, the subject-generalized N2pc model can be used to accurately identify the presence of a target on unseen images since a sensitivity of 0.8 was achieved. To fill the specificity gap of 0.61 when using this model with this subject, its subject-specific N2pc model with all 14 channels can be used, with which 84 % of all non-target stimuli were correctly identified on the unseen environment background. In other words, this forms a parallel combination of 0.80 sensitivity and 0.84 specificity for subject 1. Subject 4 was the participant who consistently achieved the worst BCI results, often below the acceptable performance range. However, when conjugating the specificity of 0.74 achieved with the subject-generalized P3 model for this subject with the sensitivity achieved through the subject-specific N2pc-P3 model (0.73), this individual's performance on an unseen background and human target characteristics improves significantly. This optimal model combination for each individual can indeed pose an interesting application to tremendously improve the results for target identification of unseen images by the brain.

Table 5.12 summarizes the optimal combination of models for each specific participant in order to improve the target vs non-target discrimination. It can be seen that the overall sensitivity and specificity handled this way substantially improves each individual performance to correctly discriminate a target from non-target stimuli appearances on unseen images. A subject average of 0.79 and 0.76 are achieved for the sensitivity and specificity, respectively.

The BCI field allows this type of model selection. In contrast to some medical applications where models are trained on a subset of data from patients with a specific disease and must generalize to other people to possibly output a diagnosis on that same disease, in BCIs, models are specifically designed for certain individual needs, and as a result, are optimized in such a way to achieve the best performance possible for a certain application where only one individual is expected to perform in the future, without the requirement of generalizing for other subjects. This

Table 5.12: The best model combination fill to improve each individual specificity and sensitivity. SG refers to subject-generalized models while SS represents a subject-specific model. BC acronym represents the best channel model variation, whereas AC depicts the all-channel model variation.

	S1	S2	S3	S4	S5	S6	Mean
Sensitivity (S_s)	0.80	0.86	0.83	0.73	0.76	0.76	0.79
Specificity (S_p)	0.84	0.73	0.72	0.74	0.74	0.77	0.76
Model (S_s/S_p)	SG-N2pc/ SS-N2pc (AC)	SS-N2pc (BC)/ SS-N2pc-P3 (BC)	SS-N2pc-P3 (BC)/ SS-N2pc (AC)	SS-N2pc-P3 (AC)/ SG-P3	SG-N2pc-P3/ SS-N2pc-P3 (AC)	SG-N2pc-P3/ SG-N2pc	

is because in BCI actuation, the brain is expected to also follow a learning curve within the application context, not just the machine learning algorithm. In the context of this project, participants are all trained for the visual search of human targets in aerial images at 5 Hz. Not only is the best model for each participant built but also the best participant whose brain allows better target identification would also be selected from the participant pool. In this case, subjects 1 and 2 provide the best trade-off between correctly identified targets and correctly identified non-targets on unseen images. These subjects, with the aid of the models achieving the best results on their data, would then be used in the real-time scenarios of visual search to better carry out this assignment.

5.3.3 Right Visual Field vs Left Visual Field classification (N2pc)

Reaching now the discrimination analysis of the N2pc to point in which lateral region of the image the target was displayed, both an optimal 7 channel selection and all 14 channel model variations were tested. Starting with the lower channel number model, the most prominent activity for the RVF vs LVF distinction using the N2pc is described in Table 5.13. It can be seen that most of the channels whose activity was more discriminative for the target vs non-target classification when using the N2pc are also present here. However, contrastingly, here a preference for the left hemisphere electrodes was given by the MBI algorithm. With this optimal channel selection, the CV loop classification performance is given in Figure 5.51.

Table 5.13: The 7 most discriminative channels to classify RVF and LVF stimuli through the N2pc ERP.

	S1	S2	S3	S4	S5	S6
Left hemisphere	PO7, PO5, P3	P7, PO7, PO5, P3	P7, PO7, PO3, PO5, P3	P7, PO7, PO3, P5	P7, PO7, PO3	PO7, PO3, PO5, P3
Right hemisphere	PO8, PO6, P6, O2	P8, PO4, P4	P8, PO6	PO4, P6 O2	PO8, PO6, O2	P8, PO8, PO6

The choice of the best model architecture for this discrimination task proved to be more ambiguous than the models discussed previously. Focusing on the CV accuracy, the Random Forest achieved the greatest subject average accuracy for the 10-fold classification (71.44 %), followed by the XGBoost and then the SVM. However, in the validation accuracy and AUC, the leading architectures were the SVM and LDA, both achieving 64.33 % and 0.70 accuracy and AUC on average. Additionally, the Random Forest actually was the second worst architecture in terms of

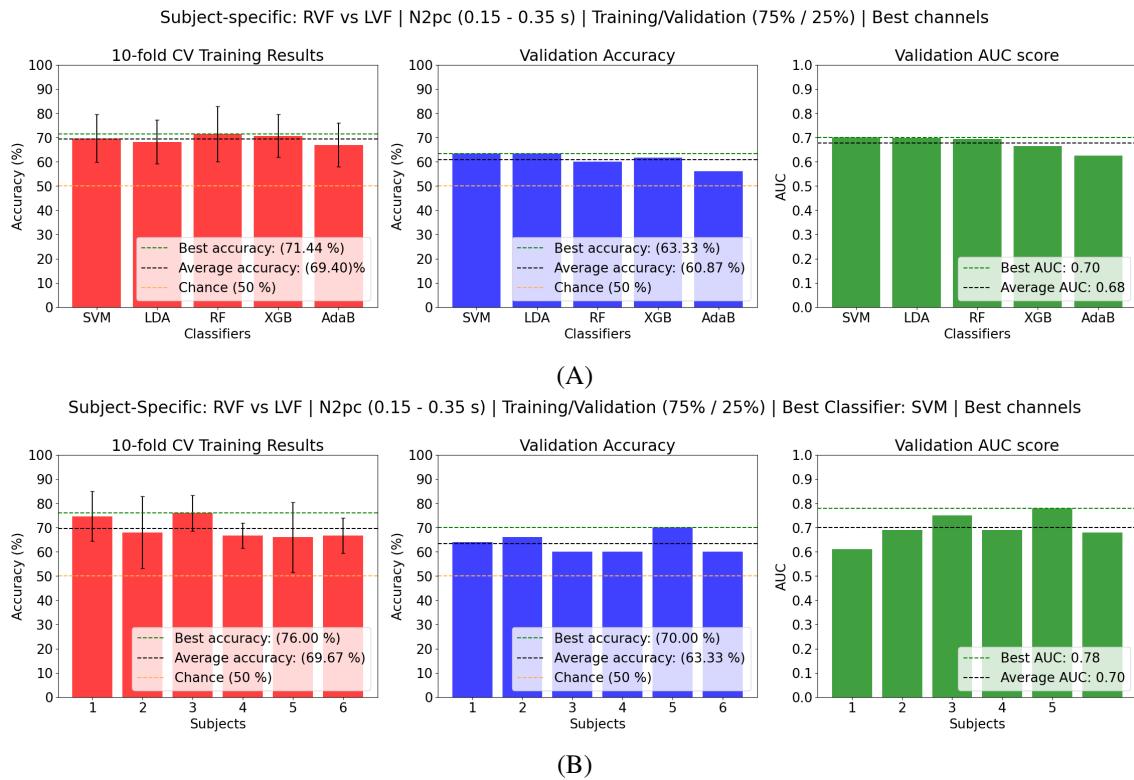


Figure 5.51: The Train/Validation results of the subject-specific N2pc model to discriminate RVF vs LVF target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found for the 7 best channels was the SVM.

validation accuracy, only losing to the AdaBoost. These three architectures were compared on the testing in order to break the tie, where the SVM proved to provide more consistent performance across subjects and was able to more appropriately generalize on the **Testing** data session.

Before looking at the performance achieved during the testing data set, a look at the individual performance for each subject with the validation set (Figure 5.51 (B)) indicates that this discrimination was not as successful as most of the target vs non-target classification problem. The subject-specific model only achieved a subject average accuracy of 63.33 % and an average AUC of 0.7. Moreover, only subject 5 was able to reach the 70 % accuracy mark, also displaying the best AUC of 0.78. With the exception of subject 5, only subject 3 surpassed the 0.70 AUC threshold.

These results stem from not performing the standard contralateral minus ipsilateral computation, commonly done in the literature, even under the BCI application context [19]. However, doing this introduces a bias that is not reproducible in online situations. For this reason, it was not adopted here in order to estimate a true blind RVF vs LVF discrimination where both feature vectors are treated equally.

Looking at the testing results represented in Figure 5.52, a similar trend can be observed. Remarkably, despite the results still being underwhelming for most subjects, the overall performance increased in participants in response to not only the unseen EEG data by the trained models but

also the unseen images by the brain. The average cohort testing accuracy increased from 63.33 % to 67.41 %, whereas the average AUC went from 0.70 to 0.73. The most noticeable performance increase is seen for subjects 2 and 5, with an accuracy increase from 65 % and 70 % to 74.11 % and 78.57 %. These also reach the minimum AUC requirements for correct BCI control in the **Testing** session. Overall, all subjects actually increase their RVF vs LVF discrimination in the latter experiment. This yet again proves that the brain, through the N2pc ERP, can generalize the RSVP paradigm in terms of the background environment presented and the human target searched for, concomitantly. When the number of trials used is contextualized for both sessions this drives even further the good generalization achieved, with the models being validated with 25 trials for each lateral region and tested with 56 trials for each class.

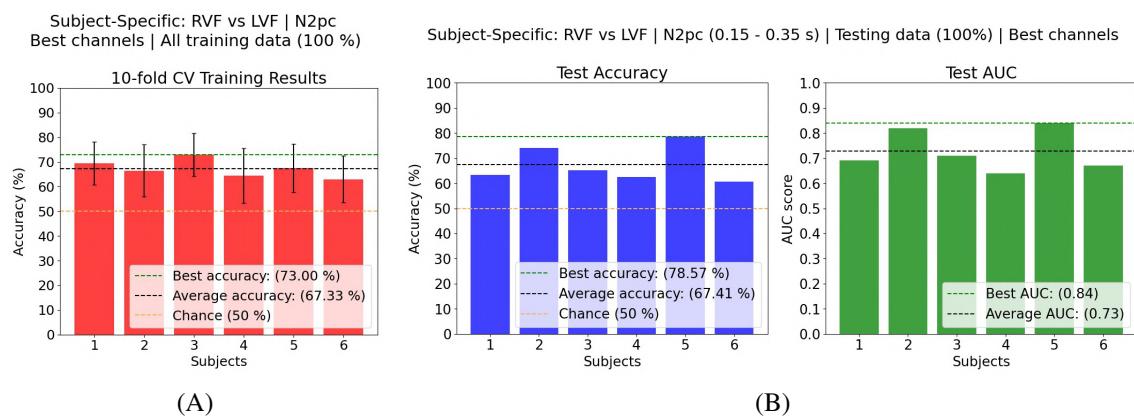


Figure 5.52: The training of the best subject-specific N2pc model with all the **Training** session data (A) and model evaluation with all the **Testing** data session (B) for the RVF vs LVF discrimination with the 7 best channels.

In addition to hinting at a good visual stimuli generalization, this also poses an interesting difference empirically observed between both potentials. The differences in performance between the **Training** and **Testing** data seem to be more negatively impacted when the P3 is present while models based on the N2pc are able to occasionally provide even better performance on the later EEG recording. This may indicate that N2pc is more robust to the mental fatigue accumulated during the first 2400 images seen but also more constant to different visual stimuli. Exploring further this idea, when the N2pc grand averages were discussed, it was stated that the RVF response during **Testing** experiment was more discernible than during the initial EEG recording. This indeed can explain the increase in performance and poses an interesting question to investigate more in the future.

When visualizing the performance over trial blocks displayed in Figure 5.53, a high variability in performance is again evident with the N2pc, even more so here than during the target vs non-target discrimination. For instance, subject 1 starts the **Testing** session with an accuracy of around 50 % but finishes the experiment with an accuracy of 85 %. On the other hand, subject 2 reaches a minimum performance of 57 % of correctly predicted labels with this model but is able to correctly predict all 14 lateral targets on the last trial block. Subject 5 also is able to reach a great

performance (92.16 %) over the 5th trial block but not always.

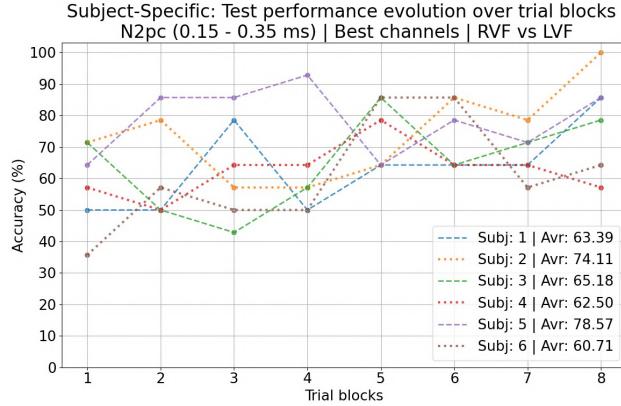


Figure 5.53: Subjects performance for the N2pc model combining the best 7 channels in discriminating RVF from LVF stimuli over the 8 trial blocks.

With the RVF stimuli representing the positive (1) and LVF the negative (0) classes, the model's sensitivity and specificity are displayed in Table 5.14. The subject average of correctly predicted RVF (0.65) and correctly predicted LVF (0.66) was fairly balanced. However, some subjects exhibited some clear dominance over one visual field with this model. These refer to subject 2, whose dominant visual field was the left with a specificity of 0.77 but a sensitivity of 0.68. Similarly, subject 4 exhibited a specificity of 0.68 but a sensitivity of 0.5. Contrastingly, subject 5 displayed an equal number of correctly predicted LVF and RVF stimuli.

Table 5.14: The sensitivity and specificity of the subject-specific N2pc RVF vs LVF model for the 7 best channels. The chance level is 50 %.

	S1	S2	S3	S4	S5	S6	Mean
Sensitivity	0.66	0.68	0.68	0.50	0.71	0.66	0.65
Specificity	0.57	0.77	0.63	0.68	0.71	0.61	0.66

Shifting now to the discussion of the 14 channel RVF vs LVF model, the subject average performance increased using the same SVM architecture, with the average validation accuracy going from 63.33 % to 67 %, as depicted in figure 5.54. This increase was produced by a rise in validation accuracy for subjects 1, 2, and 6. The best participant in the RVF vs LVF discrimination, *i.e.*, subject 5, interestingly did not benefit from doubling the channel information in this classification task, maintaining a validation accuracy of 70 %. Regarding the validation AUC metric, the subject average also increased with the 14 channel model variation relative to the 7 channel models, specifically from 0.7 to 0.75, with subject 5 maintaining the same score (0.78) and subjects 1, 2, and 6 now surpassing the 0.7 acceptable AUC range, which did not occur before.

When attempting to generalize the predictions to the testing data set, this N2pc model achieved a better subject-average performance than on the validation set, with an average accuracy of 68.75 %, as shown in Figure 5.55, only 1 % better than using half of the channel information. Furthermore, subjects 2, 3, and 5 achieved a greater accuracy on the testing set relative to the validation data. Interestingly, subject 5 increased from a validation accuracy of 70 % to 78.57 %, exactly the

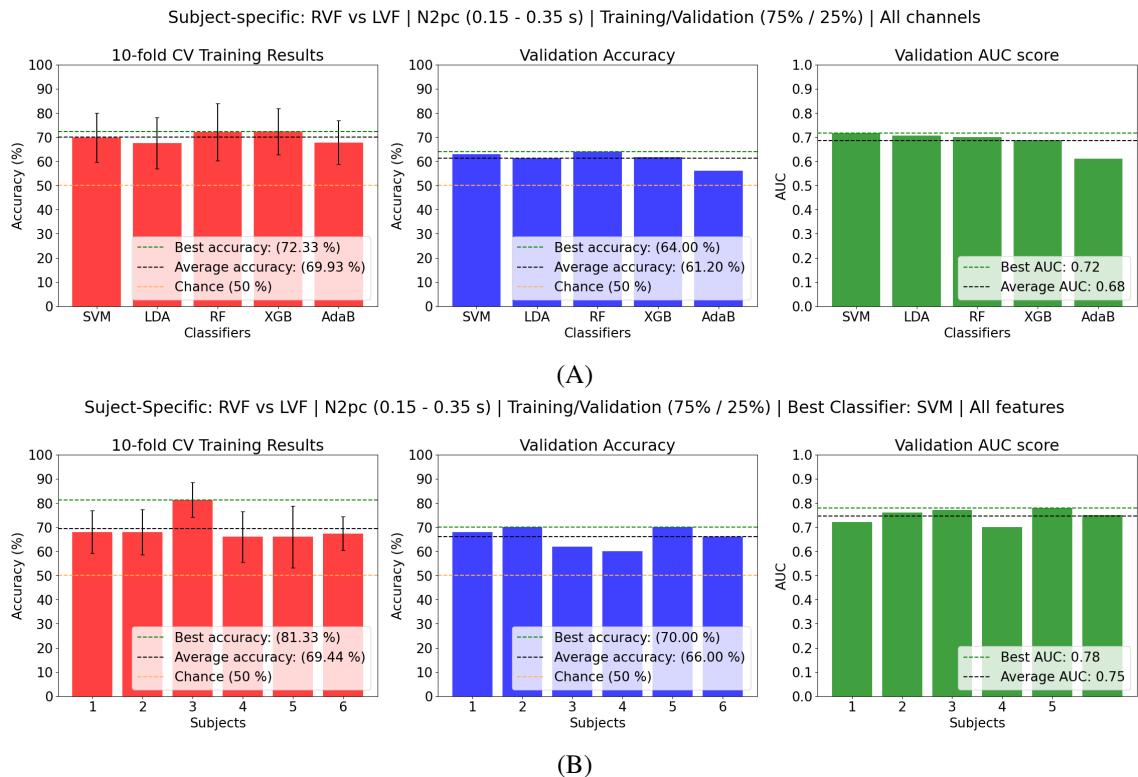


Figure 5.54: The Train/Validation results of the subject-specific N2pc model to discriminate RVF vs LVF target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found for the 9 best channels was the SVM.

same performance difference seen between the validation and testing sets on the 7 best channel models for this subject. However, with 14 channels, subjects 1 and 6 had a slight performance drop in the testing set relative to the validation set. Nevertheless, for these subjects, the testing performance was slightly greater with more channel data included in the classification.

Notably, the AUC displayed on the testing set was identical (0.75) to the validation set when using 14 channels but superior to the testing AUC when using only 7 channels (0.73) to discriminate RVF vs LVF. With the additional channel information, subject 5 surpassed the 0.8 AUC threshold for correct BCI operation within the RVF vs LVF classification, achieving 0.87 on the AUC score, even better than on the 7 channel model (0.84). Additionally, subject 2 also displayed an AUC superior to 0.8, much greater than on the validation set. However, this participant did not benefit from more channel data in this discrimination task. Finally, subjects 2 and 6 achieved exactly 0.7 AUC during the **Testing** session data, touching the acceptable AUC threshold. However, subjects 1 and 4 suffered a tenuous decrease in AUC when compared to the validation set, falling slightly behind the 0.7 AUC range with the former displaying an AUC of 0.69 and the latter 0.67, only 0.01 units lower between both data sets.

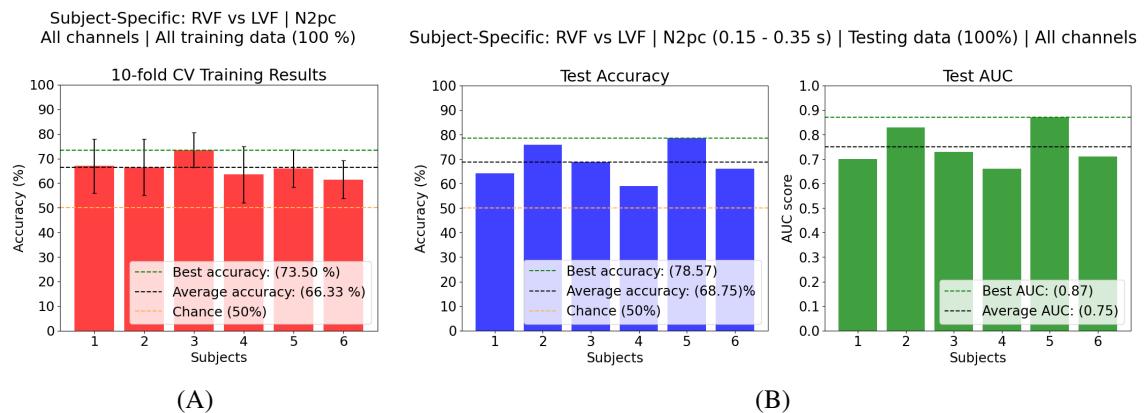


Figure 5.55: The training of the best subject-specific N2pc model with all the **Training** session data (A) and model evaluation with all the **Testing** data session (B) for the RVF vs LVF discrimination.

These results validate once again the brain's ability to generalize the background environment and target property differences. Such results are comparable to the ones observed in the literature where a median AUC of 0.76 was achieved across an 11-participant cohort in classifying RVF vs LVF targets during a 5 Hz RSVP *oddball paradigm* [76]. In a similar paradigm and cohort size, a median of 0.77 was also obtained [19]. However, it has to be emphasized that the subject average test AUC of 0.75 reached here was not obtained by biasing the N2pc contralateral minus ipsilateral computation depending on the lateral position of the human target, whereas in both papers referred, this is the case [19, 76]. This means that virtually the same performance was obtained the methodology used here can actually be extrapolated to a real-time visual search scenario without the need to assume the target lateral position priorly, essentially representing a true RVF vs LVF blind discrimination.

When looking at the performance over the 8 trial blocks (Figure 5.56), the inter-trial block variability was once again very inconsistent. In subject 1, the 7 channel model achieved its worst accuracy of 50 % in trial block 2. However, by adding another 7 N2pc-relevant channels to the discrimination, the performance achieved was even lower within this trial block, with 35 % of correctly predicted labels, well below the chance accuracy for a binary classification. Despite this possible performance outlier, the average accuracy for this subject across all 8 trial blocks was still 1 % superior with this model relative to the 7 channel one. In the last trial, subject 2 maintained its accuracy of 100 % between both model variations. Notably, in the best trial block for subject 3, specifically trial block 5, the 14 channel model was able to correctly discriminate 93.12 % of all lateral targets, while with the 7 channel model, an accuracy of 86 % was achieved. Furthermore, despite the average accuracy being exactly the same between both model variations for subject 5 (78.57 %), with the extra channel information the overall performance was much more consistent throughout the 8 trial blocks.

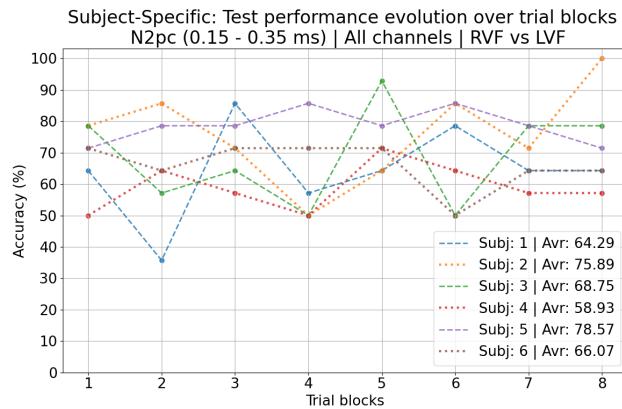


Figure 5.56: Subjects performance for the N2pc model combining all 14 channels in discriminating RVF from LVF stimuli over the 8 trial blocks.

When comparing the sensitivity and specificity results between both model variations (Table 5.15), the subject averages for these metrics increased slightly with the extra channel information provided, mainly due to subject 5, whose performance is carrying this discrimination task. Here, the ratio of correctly predicted RVF targets was 0.68 and the ratio of correctly predicted LVF targets was 0.69, essentially increasing by 4 % both metrics relative to the 7 channel N2pc model. Subject 5 achieved a great discrimination between RVF and LVF stimuli, with a better tendency to the right visual field where a sensitivity of 0.82 against 0.75 specificity was obtained. Already confirmed by the other metrics discussed, subject 2 also displayed good discrimination, exhibiting a sensitivity of 0.78 and specificity of 0.73.

Table 5.15: The sensitivity and specificity of the subject-specific N2pc RVF vs LVF model for the 14 N2pc channels. The chance level is 50 %.

	S1	S2	S3	S4	S5	S6	Mean
Sensitivity	0.66	0.78	0.72	0.50	0.82	0.61	0.68
Specificity	0.63	0.73	0.64	0.68	0.75	0.71	0.69

5.3.4 Top Visual Field vs Bottom Visual Field classification (N2pcb)

With the lateral regions covered by the N2pc, an attempt to analogously do the same for the vertical region using the N2pcb was made. As previously stated when discussing the grand-average N2pcb waveform, by observing the channel data for this ERP it was stated that a classifier could have trouble in separating both TVF and BVF classes. As it turns out, this intuition was right, as confirmed by Figure 5.57.

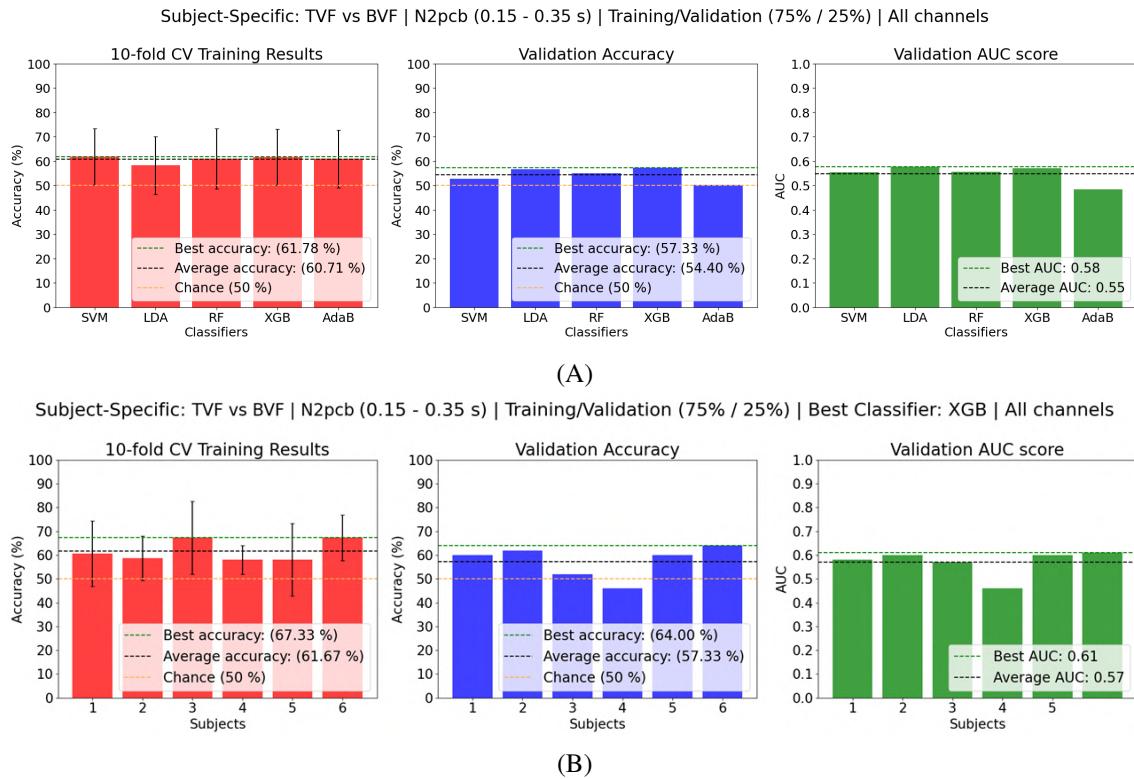


Figure 5.57: The Train/Validation results of the subject-specific N2pcb model to discriminate TVF vs BVF target stimuli across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was the XGBoost.

As can be visualized, the best architecture found for this problem was the SV. However, the results from this classification were not satisfactory. Because of this only, the 14 channel model variation was presented since performance for the optimal channel selection was even worse on average (54.40 % validation accuracy), despite a possible outlier with subject 5 displaying an accuracy of 70 % in the validation set.

With all channel data, a subject average of 57.33 % accuracy was obtained, with subject 6 exhibiting the best accuracy out of the cohort (64 %). The AUC results were also chance-like and well below the acceptable AUC threshold. Due to not reaching satisfactory results within the validation set, this potential was not explored further for the **Testing** session data.

5.3.5 Four quadrant classification (N2pc)

Similarly to the N2pcb results in discriminating the TVF and BVF target stimuli, attempting a 4 quadrant discrimination using the N2pc did not achieve satisfactory results, as indicated by Figure 5.58, where the multiclass model performance is displayed for the 14 channel feature vectors.

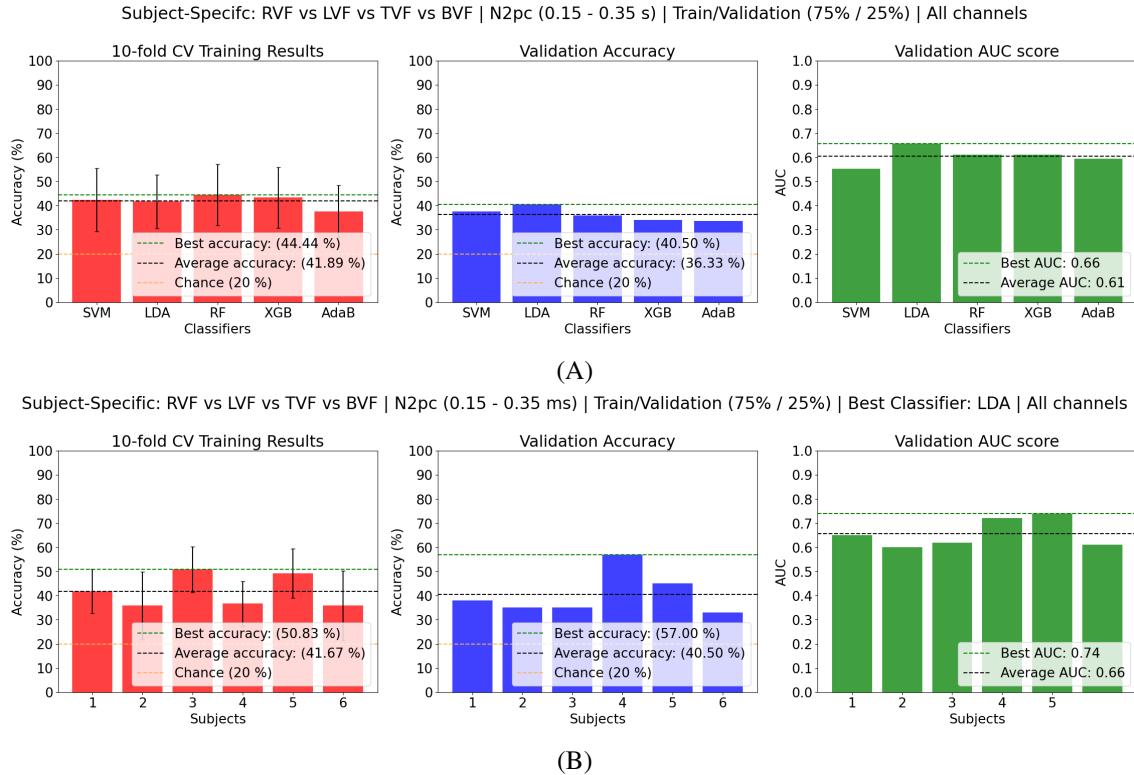


Figure 5.58: The Train/Validation results of the subject-specific N2pc model to discriminate the 4 quadrants (RVF-TVF, RVF-BVF, LVF-TVF, LVF-BVF) of target appearance across all the classifiers tested (A) and the best classifier performance across subjects (B). The best architecture found was LDA.

The best performance was achieved by the LDA architecture. However, this loses relevance since a subject average accuracy of 40.50 % was achieved, with subject 4 providing the best accuracy results, *i.e.*, 57 %. Despite being above the chance level for a 4-class discrimination problem (25 %), these performances are still below the satisfactory level. Despite the low accuracy results, a high AUC relative to the accuracy metric was obtained, with an average value of 0.74 (Figure 5.58 (B)). However, this stems from the fact that the AUC was calculated with the One Versus The Rest method, where a class is placed against all others combined, essentially working as 4 binary models. Because of the low performance, this model was also not explored further.

5.3.6 Horizontal centroid estimation (N2pc)

An Artificial Neural Network architecture was used in order to estimate the x-axis position, represented by a pixel point within the lateral image dimension where a target centroid appeared. The validation results in the form of the average Pearson Correlation Coefficient between the predicted pixel values and the ground truth within each target appearing trial are summarized for each subject in Table 5.16. The statistical significance of this correlation is also provided by the *p-value*.

Table 5.16: The Pearson Correlation Coefficient (PCC) and corresponding correlation statistical significance (*p*-value) between the predicted and real x-axis centroids for each subject in the validation set.

	S1	S2	S3	S4	S5	S6	Mean
PCC	0.77	0.57	0.55	0.19	0.56	0.32	0.49
p-value	< 0.001	< 0.001	< 0.001	0.195	< 0.001	0.026	-

It can be seen that with the validation set, a subject average PCC of 0.49 was obtained. Despite not being impressive, the results obtained here are much greater than the ones obtained in the literature attempting to achieve the same lateral centroid estimation within images. During two 5 Hz RSVP *oddball paradigm*, in a cohort of 11 subjects, an average PCC of 0.19 [19] and 0.405 [76] were obtained. In contrast to these papers, which used an 8-neuron hidden layer, a 32-neuron hidden layer was used here to accommodate the higher dimensionality feature vectors ($n = 196$ features) resulting from using the N2pc channel pair information separated.

The authors who achieved a PCC of 0.19 for a single-user BCI, also attempted to perform a combined BCI decision for a specific centroid using groups composed of 2 up to 9 subjects with similar BCI performance. In groups of 2 and 3 subjects visualizing the same target position concomitantly, the average PCC was 0.4 and 0.46, respectively. The single-user BCI attempted here for the same problem was, on average, superior to using the combined information of multiple users. Only when groups of 4 participants were formed, the average PCC surpassed the value achieved here (0.51).

Not only was the average correlation satisfactory but also some subjects were able to perform well above the subject average. Subject 1 presented a PCC of 0.77, closely followed by subjects 2 (0.56), 5 (0.56), and 3 (0.55). Additionally, the correlation between the predicted lateral centroid outputs and the ground truths for the validation trials was statistically significant for these subjects, as evidenced by the *p-value*. More importantly, subject 1 achieved a greater performance than all subject combinations done in the paper. For context, when conjugating 9 participants, achieving the best performance out of all combinations attempted, the average PCC was 0.62 [19].

Looking now at the results obtained by predicting the target centroid with the testing data (Table 5.17), the good results were maintained and the subject's average PCC increased from 0.49 to 0.52, now surpassing the 4 participant combination on the same single-trial target, where an average of 0.51 PCC was obtained [19]. Now, only subject 6 did not achieve a statistically significant correlation, where a *p-value* of 0.184 was obtained. Furthermore, subject 2 was the

one who achieved the best performance on the testing data set, with a PCC increase from 0.57 to 0.72. Subject 4 was the one who exhibited a steeper increase between both experiments, where his PPC went from 0.19 on the validation set to 0.48 on the testing set. Additionally, subjects 1 and 5 are right behind subject 2 with the best PCC value, both achieving a correlation of 0.65 and 0.64, respectively.

Table 5.17: The Pearson Correlation Coefficient (PCC) and corresponding correlation statistical significance (p-value) between the predicted and real x-axis centroids for each subject in the testing set.

	S1	S2	S3	S4	S5	S6	Mean
PCC	0.65	0.72	0.52	0.48	0.64	0.13	0.52
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.184	-

Out of the 6 participants, three increased their PCC in the testing set relative to the validation set. This indicates that the lateral centroid estimation is also generalized to first-time unseen environments and human targets through the N2pc segmented between 150 to 350 ms after stimuli onset. Notably, it is important to note that none of the target centroids used to train the models are coincident with the target positions presented during the **Testing** session, driving further the generalization achieved through the Neural Network models.

To provide a better visualization of the correlation between the predicted x-axis target centroid relative to the ground truth, subject 1 performance is illustrated in Figure 5.59. In this image, the predicted value is plotted against the real centroid value for that specific target along the horizontal image dimension. The LVF and RVF ground truths are plotted with blue and red lines, respectively. Furthermore, the target size in pixels within the image has relevance in the context of predicting its position within the image. This is because the retinotopic map in the visual cortex does not pick a single point but rather the whole target. Because of this, it is acceptable for the prediction to deviate slightly from the ground truth, since the target representation of the retinotopic map is not as precise, even less so the activity picked by electrodes on the scalp resulted from this representation. The human target in the image presentation displayed has a maximum width of approximately 120 pixels. The green rectangle presented in the Figure illustrates this acceptable margin between the predicted point and the actual centroid.

It can be seen that for subject one, the deviation of the predicted lateral target centroid from the ground truth is minimal, especially when considering the acceptable green margin. Out of 50 points, only 11 were not encapsulated by this margin. However, only 8 were actually predicted to appear on the opposite visual angle side. Additionally, a clear preference for the right visual field can be seen for subject 1, already hinted at from the sensitivity and specificity metrics in the RVF vs LVF discrimination models.

A similar trend can be seen for participant 2, as depicted in Figure 5.60, where the right visual appears to dominate the correct predictions. It is important to note that, despite the LVFA discussed previously may not have a direct influence on the actual model's performance. It refers to the fact that LVF targets generate a faster response in the visual cortex than RVF targets. However, these

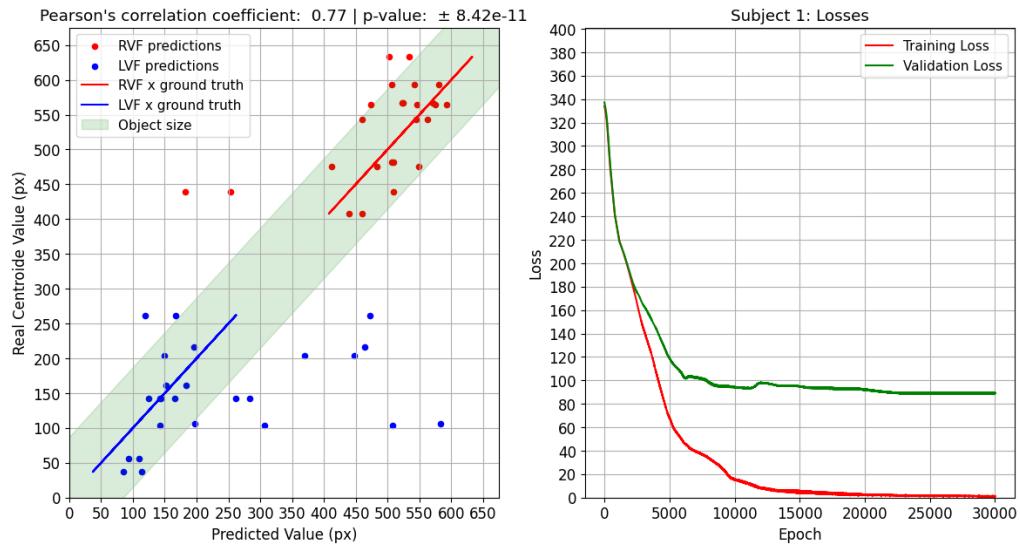


Figure 5.59: Subject 1 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model’s train and validation loss (right).

models appear to be able to better discern the N2pc modulation for RVF targets. This may be indicative of the dominant eye effect.

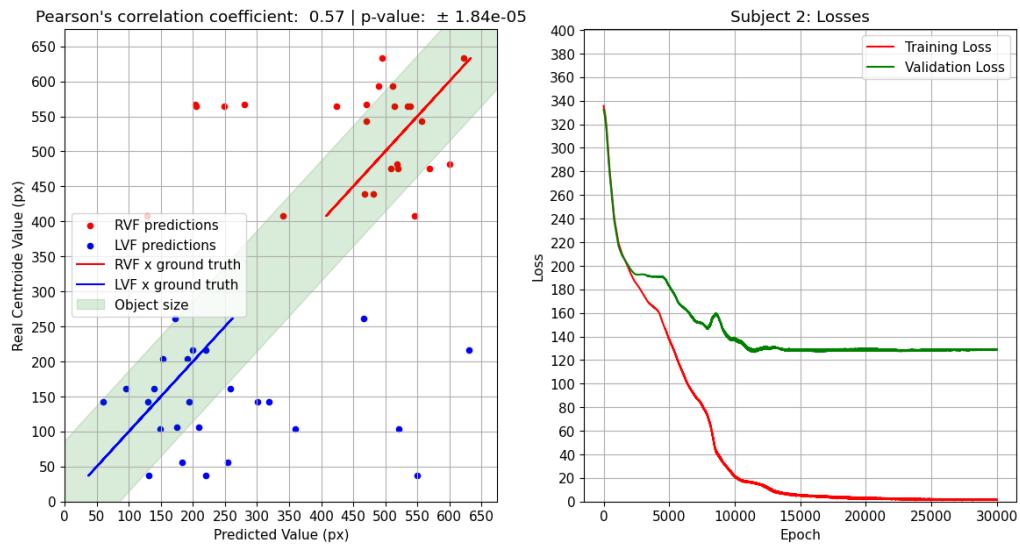


Figure 5.60: Subject 2 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model’s train and validation loss (right).

Looking now at the subject performances but for the **Testing** data session, represented in Figure 5.61, it can be seen that the increased number of trials produced a higher number of noisy centroid estimations but the overall PPC values were still high, meaning the majority of predictions was still focused around the ground truth. Furthermore, in both subjects 1 (A) and 2 (B), the same trends of an RVF preference can be seen. Despite the reduction in performance between both experiments for subject 1, a cluster of predicted points inside the target centroid size is still very

discernible. However, the lower performance within the LVF is much more noticeable here than on the validation set, meaning that these N2pc modulations were not captured as well.

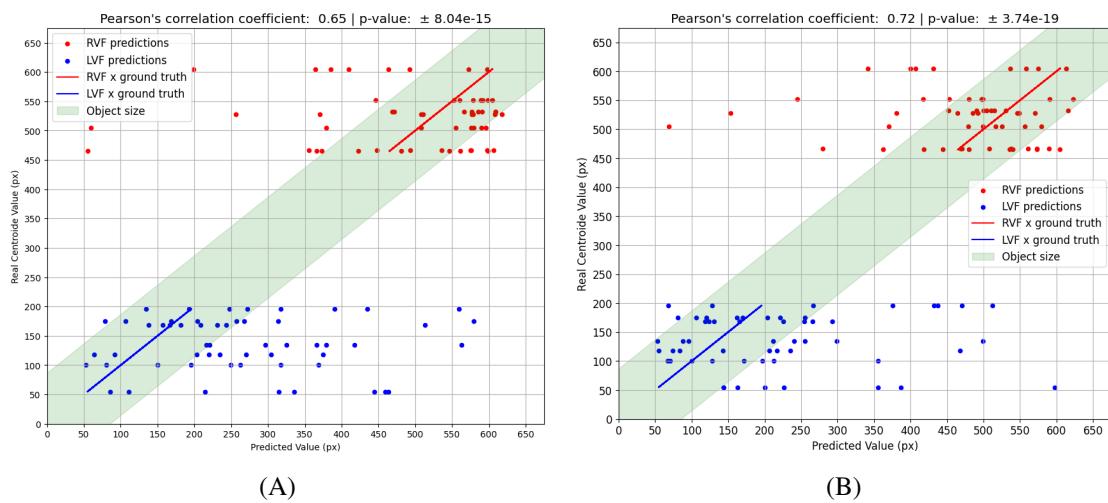


Figure 5.61: Subject 1 (A) and 2 (B) performance in predicting the correct pixel where a lateral target appears within the image for the single-trial testing data.

Nevertheless, the subject's 2 model was able to center these predictions much more closely within the margins in the testing set as indicated by the significant increase in PCC from 0.57 to 0.72, emphasizing again the good brain and model generalization achieved with this project.

The subject 3, 4, 5, and 6 train/validation set x-centroid estimation performance are displayed in Figures A.15, A.16, A.17, and A.18 of Appendix A. Finally, the corresponding test performance can also be seen for subjects 3 and 4 in Figure A.19 and for subjects 5 and 6 in Figure A.20 of Appendix, where similar observations can be made.

5.4 Chapter 5 summary

This Chapter entails all the results achieved with this project. Firstly, the LDR circuit developed and ordered from *Eurocircuits* was able to successfully synchronize the continuous EEG data in all data recordings. Furthermore, the encapsulation design for the circuit and the piece built to hold the circuit to the presentation monitor provided good fixation throughout all experiments.

Regarding the EEG, the signal quality analysis yielded good results, with only one subject (S2) suffering from significant smeared channels. Furthermore, the SSVEP component Signal-To-Noise ratio estimation demonstrated that this component, despite provoking an increase in this metric at stimulation frequency (5Hz) and its main harmonics (10Hz and 15Hz) when compared to other frequencies, was not significant in the occipital region in comparison to the rest of the scalp. Additionally, it was possible to untangle this component from the targeted EEG responses with ICA. Because of this, additional steps to remove this component were not explored. The ICA algorithm was also deemed unnecessary to remove other artifacts, with the low-pass filtering used

(28 Hz) being enough to handle them, even muscle artifacts. By not using ICA, the pipeline is also much faster.

In the classification department, all Target vs Non-Target BCI models were able to produce much better results than the classic Human Pose Estimation model used here, whose performance was around the chance-like threshold. The different ERPs used for this discrimination produced similar results, with all generalized models performing lower results than the subject-specific variation. However, it was noted that some worked well to discriminate one class but not the other in particular subjects. By selecting the optimal model combination to discriminate each class (Target vs Non-Target), a subject's average sensitivity and specificity of 0.79 and 0.76 to discriminate Target and Non-target images, respectively, was obtained, yielding very competitive results. More importantly, from this combination subject 1 was able to achieve the best results with two N2pc-based models, providing a much faster decision output when compared to other subjects where at least one of the models required the P3.

In all Target vs Non-Target BCI models, it was noticed that similar accuracy and AUC results were obtained between both experiment responses. This indicates a high degree of generalization to novel images comprising never-seen environments and human targets. This motivates that through these ERP and methodology proposed here, a good performance could be maintained in a real-time BCI applied to search and rescue missions.

Similar observations were made for the lateral target discrimination using the N2pc. In the **Training** experiment, an average accuracy and AUC of 66% and 0.75 across all subjects was obtained. In the **Testing** paradigm, an even better average accuracy of 68.75% was achieved, while the average AUC was maintained at 0.75. These results are on the same level as the ones found in the literature. Notably, some subjects performed substantially better in the **Testing** paradigm in comparison to the **Training** session.

As for the centroid estimation regressor built, the performance obtained here was even better than in the literature. A subject-average correlation between the predicted centroid and the real position of 0.49 was achieved in the **Training** experiment and 0.52 in the **Testing** session, once again validating the good generalization achieved.

Despite the good lateral target results, the vertical dimension was not captured as well with both the N2pcb and the N2pc targeting the 4 quadrants of target centroids.

Chapter 6

Conclusion

The goal of this project was based on the necessary improvements required for the visual search of aerial images. This was done by introducing a brain-in-the-loop in the whole system to handle the fast object of interest identification within the images. For this, a BCI targeting the visual N2pc and P3 ERP potentials was designed during a Rapid-Serial Visual Presentation *oddball paradigm*, where aerial images taken by a drone and containing a human target were displayed at 5Hz speed. The 6-subject cohort was asked to mentally count each instance when a human target appeared lying down in 10 % of the images presented.

However, before handling the EEG acquisition, a circuit capable of synchronizing each class of images presented with EEG data was designed and fixed to the monitor where the image presentation took place. This circuit was able to deliver the precise temporal synchronization necessary to segment the biosignal.

Regarding the EEG acquisition, in order to test the brain's ability to generalize the identification of a different human target in different background environments, two visual paradigms were designed with different properties. Both depict the same human target visual search of aerial images but with independent sets of background environments and different human targets. In addition, in contrast to the first picture dataset, the images composing the second dataset were never seen by the subjects, being only observed during the acquisition period.

By testing several subject-specific and subject-generalized models based on the single-trial N2pc and P3 potentials and by combining different sets of channels where these ERPs are elicited, this study concluded that their temporal emergence in the visual processing stream can be successfully exploited at fast presentation rates to identify a human target within the images. More importantly, the cohort achieved similar results in both experiments with some subjects even surpassing their performance for novel images containing different visual information. Moreover, by combining the best target identification models for each subject, the results achieved were on par with the literature. More importantly, the Deep Learning model tuned to identify humans used here achieved significantly worse results than the BCI used here.

Furthermore, a model exploiting the single-trial N2pc potential was used to discriminate the lateral visual field in which the human target appears. Albeit not achieving results as consistent

for all subjects, some participants also displayed a great performance in discriminating targets on the right and left visual fields. Notably, the results also hint at a good generalization to novel images and are on the same level as studies attempting the same in the literature. This idea was explored even further and the same potential was successfully used to estimate the human target centroid along the horizontal dimension of the images, substantially surpassing the results from the literature, even with the unseen images. However, models attempting to cover the vertical target positions did not achieve satisfactory results.

These results demonstrate that the methodology used here is transversal to new images containing different visual information and can be exploited to vastly improve the efficiency of visual search missions. With the generalization seen through the visual N2pc and P3, this methodology can even perhaps be extrapolated to medical-related applications such as providing gaze-independent control options for people who lose motor abilities, including ocular movements.

Chapter 7

Future work

From some good results coming out of this project, some holes regarding the ERPs and BCI performance must be explored in the future. Firstly, the presence of the N170 face recognition potential within the N2pc target response must be specifically targeted, which may be behind the RVF and LVF amplitude differences observed. This needs to be unraveled in order to assess if this BCI can be used irrespective of a human target and if it can be generalized to other types of targets. Furthermore, although the RVF and LVF lateral target discrimination worked well for some subjects, the vertical position was not captured correctly with either the N2pc or N2pcb. This is required to cover all quadrants of the target appearance.

Additionally, the main idea to explore in the future is the real-time performance assessment of this BCI. One of the initial proposals for this project was to evaluate the real-time performance of the BCI developed here within a Virtual Reality (VR) environment, where the system could be safely tested without damaging the drone and where all the visual variables are more tightly controlled. This would facilitate the image dataset formation since both the visual stimuli size, distance from the drone, and visual angles would be more easily planned. Furthermore, the images would have been presented through an *Occulus Rift* since immersiveness has been shown to improve BCI performance across several paradigms [124].

A replica from the Camp Roberts Military Base was available in a VR environment developed in *Unreal Engine 4.0* and integrated with *AirSim*, a simulation platform for AI research and experimentation developed by *MicroSoft* where virtual drones, including all its physics and operation setup, are already available to use. However, this environment faced many problems. Firstly, the replica was very unrealistic and clumsy, which directly affects the immersiveness required. Secondly, the area recreated was very small and not suited for drone testing. Because of this, a more realistic VR environment covering a larger area of Camp Roberts was built from scratch in *Unreal Engine 5.0*. Figure 7.1 showcases the environment available (A) and the one created during this thesis (B).

It can be seen that the environment created covers the whole area of Camp Roberts and is much more realistic despite some texture popups and resolution issues (B) due to the camera distance when taking the screenshot, which was unnoticeable when traversing the environment in "player



(A)



(B)

Figure 7.1: Comparison between the VR environment of Camp Roberts already built (A) and the one created during this project (B).

mode". However, the *AirSim* project was discontinued before *Unreal Engine 5.0* emerged. Despite 3rd parties developing a version compatible with version 5.0 of *Unreal Engine*, many challenges were faced during its setup due to requiring specific tools to build the source code in *C++*. The time spent to overcome the complex compiling errors encountered was piling up and this idea was not tested further in order to shift the focus to the EEG acquisition and analysis. Nevertheless, the environment was at a stage to be used if it were not for the *AirSim* problems faced. In the future, BCI systems, not exclusive to RSVP *oddball paradigms*, could be easily and more safely tested within this simulation.

References

- [1] J. C.R. Licklider. Man-Computer Symbiosis. *IRE Transactions on Human Factors in Electronics*, HFE-1(1):4–11, 1960. doi:10.1109/THFE2.1960.4503259.
- [2] Gerwin Schalk. Brain-computer symbiosis. *Journal of neural engineering*, 5(1), mar 2008. URL: <https://pubmed.ncbi.nlm.nih.gov/18310804/>, doi:10.1088/1741-2560/5/1/P01.
- [3] 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. doi:10.1109/JPROC.2015.2419736.
- [4] WHO. World report on disability. *World Health Organization 2011*, 2011. Last accessed on 2024-01-19. URL: <https://www.who.int/publications/i/item/9789241564182>.
- [5] Jin Park, Jee Eun Kim, and Tae Jin Song. The Global Burden of Motor Neuron Disease: An Analysis of the 2019 Global Burden of Disease Study. *Frontiers in Neurology*, 13:672, apr 2022. doi:10.3389/FNEUR.2022.864339/BIBTEX.
- [6] Aya Rezeika, Mihaly Benda, Piotr Stawicki, Felix Gembler, Abdul Saboor, and Ivan Volosyak. 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>, doi:10.3390/BRAINSCI8040057.
- [7] T.D. Sunny, A. Aparna, J. Venkateswaran, V. Vishnupriya, and P.S. Vyas. Robotic Arm with Brain – Computer Interfacing. *Procedia Technology*, 24:1089–1096, jan 2016. doi:10.1016/J.PROTCY.2016.05.241.
- [8] Á Fernández-Rodríguez, F Velasco-Álvarez, and R Ron-Angevin. Review of real brain-controlled wheelchairs. *Journal of neural engineering*, 13(6), oct 2016. URL: <https://pubmed.ncbi.nlm.nih.gov/27739401/>, doi:10.1088/1741-2560/13/6/061001.
- [9] Rupert Otner, Jitka Annen, Tim von Oertzen, Arnau Espinosa, Javi Rodriguez, Brendan Z. Allison, Gunter Edlinger, Steven Laureys, Martin Hmaberger, Andrea Kammerhofer, Florian Guttmann, and Christoph Guger. BCIs for DOC Patients: Assessment, Communication, and New Directions. *International Conference on Universal Access in Human-Computer Interaction*, pages 62–71, 2016. URL: https://link.springer.com/chapter/10.1007/978-3-319-40244-4_7.

- [10] Caroline Schnakers, Audrey Vanhaudenhuyse, Joseph Giacino, Manfredi Ventura, Melanie Boly, Steve Majerus, Gustave Moonen, and Steven Laureys. Diagnostic accuracy of the vegetative and minimally conscious state: Clinical consensus versus standardized neurobehavioral assessment. *BMC Neurology*, 9(1):1–5, jul 2009. URL: <https://bmcneurol.biomedcentral.com/articles/10.1186/1471-2377-9-35>, doi:10.1186/1471-2377-9-35/TABLES/2.
- [11] María A. Cervera, Surjo R. Soekadar, Junichi Ushiba, José del R. Millán, Meigen Liu, Niels Birbaumer, and Gangadhar Garipelli. Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis. *Annals of Clinical and Translational Neurology*, 5(5):651–663, may 2018. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5945970/>, doi:10.1002/ACN3.544.
- [12] Rahul Kumar Chaurasiya, Narendra D. Londhe, and Subhojit Ghosh. An efficient P300 speller system for Brain-Computer Interface. *Proceedings of 2015 International Conference on Signal Processing, Computing and Control, ISPCC 2015*, pages 57–62, jan 2016. doi:10.1109/ISPCC.2015.7374998.
- [13] Christoph Reichert, Stefan Durschmid, Mandy V. Bartsch, Jens-Max Hopf, Hans Jochen Heinze, and Hermann Hinrichs. Decoding the covert shift of spatial attention from electroencephalographic signals permits reliable control of a brain-computer interface. *Journal of Neural Engineering*, 17(5):056012, oct 2020. URL: <https://iopscience.iop.org/article/10.1088/1741-2552/abb692>, doi:10.1088/1741-2552/ABB692.
- [14] 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. doi:10.1088/1741-2560/8/2/025005.
- [15] Soram Kim, Seungyum Lee, Hyunsuk Kang, Sion Kim, and Minkyu Ahn. P300 Brain-Computer Interface-Based Drone Control in Virtual and Augmented Reality. *Sensors (Basel, Switzerland)*, 21(17), sep 2021. URL: [/pmc/articles/PMC8434009//pmc/articles/PMC8434009/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC8434009/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8434009/), doi:10.3390/S21175765.
- [16] Yufeng Zhang, Hongxin Zhang, Xiaorong Gao, Shangen Zhang, and Chen Yang. UAV Target Detection for IoT via Enhancing ERP Component by Brain Computer Interface System. *IEEE Internet of Things Journal*, oct 2023. doi:10.1109/JIOT.2023.3273163.
- [17] Josef Faller, Jennifer Cummings, Sameer Saproo, and Paul Sajda. Regulation of arousal via online neurofeedback improves human performance in a demanding sensory-motor task. *Proceedings of the National Academy of Sciences of the United States of America*, 116(13):6482–6490, mar 2019. URL: <https://www.pnas.org/doi/abs/10.1073/pnas.1817207116>, doi:10.1073/PNAS.1817207116/SUPPL_FILE/PNAS.1817207116.SAPP.PDF.
- [18] Allison Ryan and J. Karl Hedrick. A mode-switching path planner for UAV-assisted search and rescue. *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference, CDC-ECC '05*, 2005:1471–1476, 2005. doi:10.1109/CDC.2005.1582366.

- [19] 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. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0178498>, doi:10.1371/JOURNAL.PONE.0178498.
- [20] Ana Matran-Fernandez and Riccardo Poli. Collaborative brain-computer interfaces for target localisation in rapid serial visual presentation. *2014 6th Computer Science and Electronic Engineering Conference, CEEC 2014 - Conference Proceedings*, pages 127–132, nov 2014. doi:10.1109/CEEC.2014.6958567.
- [21] 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. URL: https://link.springer.com/chapter/10.1007/978-1-84996-272-8_9, doi:10.1007/978-1-84996-272-8_9.
- [22] Johannes Jacobus Fahrenfort, Anna Grubert, Christian N.L. Olivers, and Martin Eimer. Multivariate EEG analyses support high-resolution tracking of feature-based attentional selection. *Scientific Reports* 2017 7:1, 7(1):1–15, may 2017. URL: <https://www.nature.com/articles/s41598-017-01911-0>, doi:10.1038/s41598-017-01911-0.
- [23] Hans Berger. Über das Elektrenkephalogramm des Menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87(1):527–570, dec 1929. doi:10.1007/BF01797193.
- [24] 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>.
- [25] L. A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and clinical neurophysiology*, 70(6):510–523, 1988. URL: <https://pubmed.ncbi.nlm.nih.gov/2461285/>, doi:10.1016/0013-4694(88)90149-6.
- [26] Jonathan R. Wolpaw, Dennis J. McFarland, Gregory W. Neat, and Catherine A. Forneris. An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology*, 78(3):252–259, 1991. URL: <https://www.sciencedirect.com/science/article/abs/pii/001346949190040B>, doi:10.1016/0013-4694(91)90040-B.
- [27] 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>, doi:10.1038/s41928-023-00929-9.
- [28] Simanto Saha, Khondaker A. Mamun, Khawza Ahmed, Raqibul Mostafa, Ganesh R. Naik, Sam Darvishi, Ahsan H. Khandoker, and Mathias Baumert. Progress in Brain Computer Interface: Challenges and Opportunities. *Frontiers in Systems Neuroscience*, 15:578875, feb 2021. URL: [/pmc/articles/PMC7947348/](https://PMC7947348/)

- PMC7947348/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7947348/, doi:10.3389/FNSYS.2021.578875.
- [29] Bojan Kerous, Filip Skola, and Fotis Liarokapis. EEG-based BCI and video games: a progress report. *Virtual Reality*, 22(2):119–135, 2018. URL: <https://link.springer.com/article/10.1007/s10055-017-0328-x>, doi:10.1007/s10055-017-0328-x.
 - [30] 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/>, doi:10.3389/fnbot.2020.00025.
 - [31] Sylvain Chevallier, Emmanuel K. Kalunga, Quentin Barthélémy, and Florian Yger. Riemannian Classification for SSVEP-Based BCI. In Chang S. Nam, Anton Nijholt, and Fabien Lotte, editors, *Brain–Computer Interfaces Handbook*, chapter 19, pages 371–396. CRC Press, Boca Raton, 1st edition, mar 2019. URL: <https://www.taylorfrancis.com/chapters/edit/10.1201/9781351231954-19>, doi:10.1201/9781351231954-19.
 - [32] Jiahui Pan and Yuanqing Li. Hybrid Brain–Computer Interfaces and Their Applications. In Chang S. Nam, Anton Nijholt, and Fabien Lotte, editors, *Brain–Computer Interfaces Handbook*, chapter 27, pages 525–548. CRC Press, Boca Raton, 1st edition, mar 2019. URL: <https://www.taylorfrancis.com/chapters/edit/10.1201/9781351231954-27>, doi:10.1201/9781351231954-27/HYBRID-BRAIN.
 - [33] 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>, doi:10.1088/1741-2552/AA9817.
 - [34] 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. URL: <https://www.taylorfrancis.com/chapters/edit/10.1201/9781351231954-1>, doi:10.1201/9781351231954-1.
 - [35] Han Jeong Hwang, Soyoun Kim, Sooboom Choi, and Chang Hwan Im. EEG-based brain-computer interfaces: A thorough literature survey. *International Journal of Human-Computer Interaction*, 29(12):814–826, 2013. URL: <https://www.tandfonline.com/doi/abs/10.1080/10447318.2013.780869>, doi:10.1080/10447318.2013.780869/SUPPL_FILE/HIHC_A_780869_SM5293.PDF.
 - [36] Stewart Hendry. Membrane potentials. *Neuroscience in the 21st Century: From Basic to Clinical*, pages 109–133, nov 2013. URL: https://link.springer.com/referenceworkentry/10.1007/978-1-4614-1997-6_6, doi:10.1007/978-1-4614-1997-6_6/COVER.

- [37] M. Teplan. Fundamentals of EEG measurements. *Measurement Science Review*, 2(Section 2):1–11, 2002. URL: <http://www.edumed.org.br/cursos/neurociencia/MethodsEEGMeasurement.pdf>.
- [38] IMotions. EEG (Electroencephalography): The Complete Pocket Guide, 2016. Last accessed on 2024-01-19. URL: <https://imotions.com/blog/eeg/>.
- [39] 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. URL: https://link.springer.com/chapter/10.1007/978-3-030-91885-9_35, doi:10.1007/978-3-030-91885-9_35/FIGURES/11.
- [40] Felipe Lage Teixeira, Miguel Rocha e. Costa, José Pio Abreu, Manuel Cabral, Salviano Pinto Soares, and João Paulo Teixeira. A Narrative Review of Speech and EEG Features for Schizophrenia Detection: Progress and Challenges. *Bioengineering*, 10(4), apr 2023. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10135748/>, doi:10.3390/BIOENGINEERING10040493.
- [41] 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>, doi:10.1016/j.clinph.2017.06.254.
- [42] Jayant N. Acharya and Vinita J. Acharya. Overview of EEG Montages and Principles of Localization. *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, 36(5):325–329, 2019. URL: https://journals.lww.com/clinicalneurophys/Abstract/2019/09000/Overview_of_EEG_Montages_and_Principles_of_2.aspx, doi:10.1097/WNP.0000000000000538.
- [43] Włodzimierz Klonowski. Everything you wanted to ask about EEG but were afraid to get the right answer. *Nonlinear Biomedical Physics*, 3:1–5, 2009. URL: <https://pubmed.ncbi.nlm.nih.gov/19470156/>, doi:10.1186/1753-4631-3-2.
- [44] Humaira Nisar and Kim Ho Yeap. Introduction. In *EEG/ERP Analysis*, pages 16–35. CRC Press, oct 2014. URL: <https://www.taylorfrancis.com/chapters/edit/10.1201/b17605-5/introduction-humaira-nisar-kim-ho-yeap>, doi:10.1201/B17605-5.
- [45] J E Hall and M E Hall. *Guyton and Hall Textbook of Medical Physiology E-Book*. Guyton Physiology. Elsevier Health Sciences, 2020. URL: <https://books.google.co.id/books?id=H1rrDwAAQBAJ>.
- [46] Peter J. Uhlhaas and Wolf Singer. Abnormal neural oscillations and synchrony in schizophrenia. *Nature Reviews Neuroscience* 2010 11:2, 11(2):100–113, feb 2010. URL: <https://www.nature.com/articles/nrn2774>, doi:10.1038/nrn2774.
- [47] H Blumenfeld. *Neuroanatomy overview and basic definitions*. Oxford Press University, 2010. URL: <https://www.oup.com.au/books/higher-education/science/9781605359625-neuroanatomy-through-clinical-cases>.

- [48] Robin Laycock, Sheila G. Crewther, and Philippe A. Chouinard. Blink and You Will Miss It: a Core Role for Fast and Dynamic Visual Processing in Social Impairments in Autism Spectrum Disorder. *Current Developmental Disorders Reports*, 7(4):237–248, dec 2020. URL: <https://link.springer.com/article/10.1007/s40474-020-00220-y>, doi:10.1007/S40474-020-00220-Y/FIGURES/3.
- [49] Eris Chinellato and Angel P. del Pobil. The neuroscience of action and perception. *Cognitive Systems Monographs*, 28:7–38, 2016. URL: https://link.springer.com/chapter/10.1007/978-3-319-20303-4_2, doi:10.1007/978-3-319-20303-4_2/FIGURES/3.
- [50] Janir Ramos da Cruz, Vitaly Chicherov, Michael H. Herzog, and Patrícia Figueiredo. An automatic pre-processing pipeline for EEG analysis (APP) based on robust statistics. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 129(7):1427–1437, jul 2018. URL: <https://pubmed.ncbi.nlm.nih.gov/29730542/>, doi:10.1016/J.CLINPH.2018.04.600.
- [51] Teresa Bucho, Gina Caetano, Athanasios Vourvopoulos, Floriana Accoto, Ines Esteves, Sergi Bermudez I Badia, Agostinho Rosa, and Patricia Figueiredo. Comparison of Visual and Auditory Modalities for Upper-Alpha EEG-Neurofeedback. *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019:5960–5966, jul 2019. URL: <https://pubmed.ncbi.nlm.nih.gov/31947205/>.
- [52] Qing Zhou, Ruidong Cheng, Lin Yao, Xiangming Ye, and Kedi Xu. Neurofeedback Training of Alpha Relative Power Improves the Performance of Motor Imagery Brain-Computer Interface. *Frontiers in Human Neuroscience*, 16:199, apr 2022. doi:10.3389/FNHUM.2022.831995/BIBTEX.
- [53] Min Ho Lee, O. Yeon Kwon, Yong Jeong Kim, Hong Kyung Kim, Young Eun Lee, John Williamson, Siamac Fazli, and Seong Whan Lee. EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy. *GigaScience*, 8(5), may 2019. URL: <https://pubmed.ncbi.nlm.nih.gov/30698704/>, doi:10.1093/GIGASCIENCE/GIZ002.
- [54] Minkyu Ahn, Hohyun Cho, Sangtae Ahn, and Sung Chan Jun. High Theta and Low Alpha Powers May Be Indicative of BCI-Illiteracy in Motor Imagery. *PLOS ONE*, 8(11):e80886, nov 2013. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0080886>, doi:10.1371/JOURNAL.PONE.0080886.
- [55] Femke Nijboer, Niels Birbaumer, and Andrea Kübler. The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis - a longitudinal study. *Frontiers in neuroscience*, 4(JUL), 2010. URL: <https://pubmed.ncbi.nlm.nih.gov/20700521/>, doi:10.3389/FNINS.2010.00055.
- [56] Eva Maria Hammer, Sebastian Halder, Benjamin Blankertz, Claudia Sannelli, Thorsten Dickhaus, Sonja Kleih, Klaus Robert Müller, and Andrea Kübler. Psychological predictors of SMR-BCI performance. *Biological Psychology*, 89(1):80–86, jan 2012. doi:10.1016/J.BIOPSYCHO.2011.09.006.

- [57] 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/>, doi:10.1007/S10548-013-0327-5.
- [58] 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>, doi:10.1093/OXFORDHB/9780195374148.001.0001.
- [59] Thomas Hummel, Alexandra Genow, and Basile Nicolas Landis. Clinical assessment of human gustatory function using event related potentials. *Journal of Neurology, Neurosurgery & Psychiatry*, 81(4):459–464, apr 2010. URL: <https://jnnp.bmjjournals.org/content/81/4/459.abstract>, doi:10.1136/JNNP.2009.183699.
- [60] Ahmad Mayeli, Obada Al Zoubi, Evan J. White, Sheridan Chappelle, Rayus Kuplicki, Alexa Morton, Jaimee Bruce, Ryan Smith, Justin S. Feinstein, Jerzy Bodurka, Martin P. Paulus, and Sahbib S. Khalsa. Parieto-occipital ERP indicators of gut mechanosensation in humans. *Nature Communications* 2023 14:1, 14(1):1–17, jun 2023. URL: <https://www.nature.com/articles/s41467-023-39058-4>, doi:10.1038/s41467-023-39058-4.
- [61] Shravani Sur and VK Sinha. Event-related potential: An overview. *Industrial Psychiatry Journal*, 18(1):70–73, jan 2009. URL: <https://europepmc.org/article/MED/21234168>, doi:10.4103/0972-6748.57865.
- [62] Martin Eimer. The Face-Sensitive N170 Component of the Event-Related Brain Potential. *Oxford Handbook of Face Perception*, nov 2011. URL: <https://academic.oup.com/edited-volume/28040/chapter/211940504>, doi:10.1093/OXFORDHB/9780199559053.013.0017.
- [63] Gregory A. Light, Lisa E. Williams, Falk Minow, Joyce Srock, Anthony Rissling, Richard Sharp, Neal R. Swerdlow, and David L. Braff. Electroencephalography (EEG) and Event-Related Potentials (ERP’s) with Human Participants. *Current protocols in neuroscience / editorial board, Jacqueline N. Crawley ... [et al.]*, CHAPTER(SUPPL. 52):Unit, jul 2010. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2909037/>, doi:10.1002/0471142301.NS0625S52.
- [64] Jonathan R. Folstein and Cyma Van Petten. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*, 45(1):152–170, 2008. URL: <https://pubmed.ncbi.nlm.nih.gov/17850238/>, doi:10.1111/J.1469-8986.2007.00602.X.
- [65] F. Nijboer, E. W. Sellers, J. Mellinger, M. A. Jordan, T. Matuz, A. Furdea, S. Halder, U. Mochty, D. J. Krusienski, T. M. Vaughan, J. R. Wolpaw, N. Birbaumer, and A. Kubler. A P300-based brain-computer interface for people with amyotrophic lateral sclerosis. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 119(8):1909, aug 2008. URL: [/pmc/articles/PMC2853977//pmc/articles/PMC2853977/?report=abstract](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2853977/) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2853977/, doi:10.1016/J.CLINPH.2008.03.034.

- [66] P. Brunner, S. Joshi, S. Briskin, J. R. Wolpaw, H. Bischof, and G. Schalk. Does the ‘P300’ speller depend on eye gaze? *Journal of Neural Engineering*, 7(5):056013, sep 2010. URL: <https://iopscience.iop.org/article/10.1088/1741-2560/7/5/056013>, doi:10.1088/1741-2560/7/5/056013.
- [67] Martin Eimer and Anna Grubert. Spatial attention can be allocated rapidly and in parallel to new visual objects. *Current biology : CB*, 24(2):193–198, jan 2014. URL: <https://pubmed.ncbi.nlm.nih.gov/24412208/>, doi:10.1016/J.CUB.2013.12.001.
- [68] Mattia Doro, Francesco Bellini, Sabrina Brigadói, Martin Eimer, and Roberto Dell’Acqua. A bilateral N2pc (N2pcb) component is elicited by search targets displayed on the vertical midline. *Psychophysiology*, 57(3), mar 2020. URL: <https://pubmed.ncbi.nlm.nih.gov/31815301/>, doi:10.1111/PSYP.13512.
- [69] Bangyu Cai, Siyuan Xiao, Lei Jiang, Yiwen Wang, and Xiaoxiang Zheng. A rapid face recognition BCI system using single-trial ERP. *International IEEE/EMBS Conference on Neural Engineering, NER*, pages 89–92, 2013. doi:10.1109/NER.2013.6695878.
- [70] Ryan Beveridge, Shane Wilson, Michael Callaghan, and Damien Coyle. Neurogaming With Motion-Onset Visual Evoked Potentials (mVEPs): Adults Versus Teenagers. *IEEE transactions on neural systems and rehabilitation engineering*, 27(4):572–581, apr 2019. URL: <https://pubmed.ncbi.nlm.nih.gov/30869627/>, doi:10.1109/TNSRE.2019.2904260.
- [71] M. Kuba, Z. Kubová, J. Kremláček, and J. Langrová. Motion-onset VEPs: Characteristics, methods, and diagnostic use. *Vision Research*, 47(2):189–202, jan 2007. URL: <https://www.sciencedirect.com/science/article/pii/S0042698906004512>, doi:10.1016/J.VISRES.2006.09.020.
- [72] Claire Kabdebon, Ana Fló, Adélaïde de Heering, and Richard Aslin. The power of rhythms: how steady-state evoked responses reveal early neurocognitive development. *NeuroImage*, 254:119150, jul 2022. URL: <https://linkinghub.elsevier.com/retrieve/pii/S1053811922002774>, doi:10.1016/J.NEUROIMAGE.2022.119150.
- [73] Danhua Zhu, Jordi Bieger, Gary Garcia Molina, and Ronald M. Aarts. A survey of stimulation methods used in SSVEP-based BCIs. *Computational Intelligence and Neuroscience*, 2010, 2010. URL: <https://europepmc.org/article/PMC/PMC2833411>, doi:10.1155/2010/702357.
- [74] Yijing Whang, Zhiguo Luo, Shaokai Zhao, Liang Xie, Minpeng Xu, Dong Ming, and Erwei Yin. Spatial localization in target detection based on decoding N2pc component. *Journal of neuroscience methods*, 369, mar 2022. URL: <https://pubmed.ncbi.nlm.nih.gov/34979193/>, doi:10.1016/J.JNEUMETH.2021.109440.
- [75] J. R. Wolpaw, H. Ramoser, D. J. McFarland, and G. Pfurtscheller. EEG-based communication: improved accuracy by response verification. *IEEE transactions on rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, 6(3):326–333, sep 1998. URL: <https://pubmed.ncbi.nlm.nih.gov/9749910/>, doi:10.1109/86.712231.
- [76] Ana Matran-Fernandez and Riccardo Poli. Brain-Computer Interfaces for Detection and Localization of Targets in Aerial Images. *IEEE Transactions on Biomedical Engineering*, 64(4):959–969, apr 2017. doi:10.1109/TBME.2016.2583200.

- [77] Yin Tian, Haiyong Zhang, Peiyang Li, and Yang Li. Multiple correlated component analysis for identifying the bilateral location of target in visual search tasks. *IEEE Access*, 7:98486–98494, 2019. doi:10.1109/ACCESS.2019.2929545.
- [78] Jovana Paunovic Pantic, Svetlana Valjarevic, Jelena Cumic, and Igor Pantic. AI-enhanced EEG signal interpretation: A novel approach using texture analysis with random forests. *Medical Hypotheses*, page 111405, jun 2024. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0306987724001488>, doi:10.1016/J.MEHY.2024.111405.
- [79] Remy Ben Messaoud and Mario Chavez. Random Forest classifier for EEG-based seizure prediction. jun 2021. URL: <https://arxiv.org/abs/2106.04510v1>, arXiv: 2106.04510.
- [80] Md Sakib Khan, Nishat Salsabil, Md Golam Rabiul Alam, M. Ali Akber Dewan, and Md Zia Uddin. CNN-XGBoost fusion-based affective state recognition using EEG spectrogram image analysis. *Scientific Reports* 2022 12:1, 12(1):1–19, aug 2022. URL: <https://www.nature.com/articles/s41598-022-18257-x>, doi:10.1038/s41598-022-18257-x.
- [81] J ; Zong, X ; Xiong, J ; Zhou, Y ; Ji, D ; Zhou, Q Zhang, Jing Zong, Xin Xiong, Jianhua Zhou, Ying Ji, Diao Zhou, and Qi Zhang. FCAN–XGBoost: A Novel Hybrid Model for EEG Emotion Recognition. *Sensors* 2023, Vol. 23, Page 5680, 23(12):5680, jun 2023. URL: <https://www.mdpi.com/1424-8220/23/12/5680> <https://www.mdpi.com/1424-8220/23/12/5680/htm> doi:10.3390/S23125680.
- [82] Niall McGuire and Dr Yashar Moshfeghi. Prediction of the Realisation of an Information Need: An EEG Study. *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, jun 2024. URL: <http://arxiv.org/abs/2406.08105>, doi:10.1145/3626772.3657981.
- [83] Yu Chen, Rui Chang, and Jifeng Guo. Emotion Recognition of EEG Signals Based on the Ensemble Learning Method: AdaBoost. *Mathematical Problems in Engineering*, 2021(1):8896062, jan 2021. URL: <https://onlinelibrary.wiley.com/doi/full/10.1155/2021/8896062> <https://onlinelibrary.wiley.com/doi/abs/10.1155/2021/8896062> <https://onlinelibrary.wiley.com/doi/10.1155/2021/8896062>, doi:10.1155/2021/8896062.
- [84] Mario Beraha, Alberto Maria Metelli, Matteo Papini, Andrea Tirinzoni, and Marcello Restelli. Feature Selection via Mutual Information: New Theoretical Insights. *Proceedings of the International Joint Conference on Neural Networks*, 2019-July, jul 2019. URL: <https://arxiv.org/abs/1907.07384v1>, arXiv: 1907.07384, doi:10.1109/IJCNN.2019.8852410.
- [85] J. Platt and Nikos Karampatziakis. Probabilistic Outputs for SVMs and Comparisons to Regularized Likelihood Methods. *Mathematics in Computer Science*, 2007. URL: <https://api.semanticscholar.org/CorpusID:56563878>.
- [86] Zhi-Hua Zhou. Machine Learning. 2021. URL: <https://link.springer.com/10.1007/978-981-15-1967-3>, doi:10.1007/978-981-15-1967-3.

- [87] Xiaoyu Li. Analysis of College Students' Psychological Education Management in Public Emergencies Based on Big Data. *Journal of environmental and public health*, 2022, 2022. URL: <https://pubmed.ncbi.nlm.nih.gov/36193395/>, doi:10.1155/2022/2654437.
- [88] scikit-learn 1.5.0 documentation. Cross-validation: evaluating estimator performance. Last accessed on 2024-06-02. URL: https://scikit-learn.org/stable/modules/cross_validation.html#multimetric-cross-validation.
- [89] Home — PsychoPy®. Last accessed on 2023-02-01. URL: <https://www.psychopy.org/>.
- [90] Usman Ghani, Nada Signal, Imran Khan Niazi, and Denise Taylor. ERP based measures of cognitive workload: A review. *Neuroscience Biobehavioral Reviews*, 118:18–26, nov 2020. doi:10.1016/J.NEUBIOREV.2020.07.020.
- [91] Leonidas Kotoulas and Ioannis Andreadis. Accurate calculation of image moments. *IEEE Transactions on Image Processing*, 16(8):2028–2037, aug 2007. doi:10.1109/TIP.2007.899621.
- [92] Kyle M. Jensen and Justin A. MacDonald. Towards thoughtful planning of ERP studies: How participants, trials, and effect magnitude interact to influence statistical power across seven ERP components. *Psychophysiology*, 60(7), jul 2023. URL: <https://pubmed.ncbi.nlm.nih.gov/36577739/>, doi:10.1111/PSYP.14245.
- [93] Jill Cohen and John Polich. On the number of trials needed for P300. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology*, 25(3):249–255, apr 1997. URL: <https://pubmed.ncbi.nlm.nih.gov/9105949/>, doi:10.1016/S0167-8760(96)00743-X.
- [94] George Chahine and Bart Krekelberg. Cortical Contributions to Saccadic Suppression. *PLOS ONE*, 4(9):e6900, sep 2009. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0006900>, doi:10.1371/JOURNAL.PONE.0006900.
- [95] OpenBCI. OpenBCI Documentation. Last accessed on 2023-01-23. URL: <https://docs.openbci.com/>.
- [96] Muse. Muse 2 | Muse™ EEG-Powered Meditation Sleep Headband. Last accessed on 2024-05-19. URL: <https://choosemuse.com/products/muse-2>.
- [97] Neurosity. Introducing the Crown. Last accessed on 2024-05-19. URL: <https://neurosity.co/>.
- [98] Micromed Group. BRAIN QUICK® LTM. Last accessed on 2024-05-19. URL: <https://micromedgroup.com/products/brainquick/brainquick-ltm/>.
- [99] Seitaro Iwama, Mitsuaki Takemi, Ryo Eguchi, Ryotaro Hirose, Masumi Morishige, and Junichi Ushiba. Two common issues in synchronized multimodal recordings with EEG: Jitter and latency. *Neuroscience Research*, dec 2023. doi:10.1016/J.NEURES.2023.12.003.
- [100] Altium Designer - PCB Design Software. Last accessed on 2024-06-24. URL: <https://www.altium.com/altium-designer>.

- [101] V-One - A Desktop PCB Printer | Voltera. Last accessed on 2024-06-24. URL: <https://www.voltera.io/v-one>.
- [102] BESDATA. EEG Caps Leading Manufacturer and Supplier. Last accessed on 2023-01-24. URL: <https://besdatatech.com/eeg-caps/>.
- [103] Boris Kotchoubey. Event-related potentials, cognition, and behavior: A biological approach. *Neuroscience and Biobehavioral Reviews*, 30(1):42–65, 2006. doi:10.1016/j.neubiorev.2005.04.002.
- [104] MNE — MNE 1.4.0 documentation. Last accessed on 2023-05-29. URL: <https://mne.tools/stable/index.html>.
- [105] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A. Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti Hämäläinen. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7(7 DEC):70133, dec 2013. URL: <http://martinos.org/mne/>, doi:10.3389/FNINS.2013.00267/BIBTEX.
- [106] E. O. Brigham and R. E. Morrow. The fast Fourier transform. *IEEE Spectrum*, 4(12):63–70, 1967. doi:10.1109/MSPEC.1967.5217220.
- [107] Craig E. Tenke and Jürgen Kayser. A convenient method for detecting electrolyte bridges in multichannel electroencephalogram and event-related potential recordings. *Clinical Neurophysiology*, 112(3):545–550, mar 2001. doi:10.1016/S1388-2457(00)00553-8.
- [108] Peter D. 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*, 15(2):70–73, 1967. doi:10.1109/TAU.1967.1161901.
- [109] M. Bach and T. Meigen. Do's and don'ts in Fourier analysis of steady-state potentials. *Documenta ophthalmologica. Advances in ophthalmology*, 99(1):69–82, 1999. URL: <https://pubmed.ncbi.nlm.nih.gov/10947010/>, doi:10.1023/A:1002648202420.
- [110] Thomas Meigen and Michael Bach. On the statistical significance of electrophysiological steady-state responses. *Documenta ophthalmologica. Advances in ophthalmology*, 98(3):207–232, 1999. URL: <https://pubmed.ncbi.nlm.nih.gov/10945442/>, doi:10.1023/A:1002097208337.
- [111] F. Perrin, J. Pernier, O. Bertrand, and J. F. Echallier. Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, 72(2):184–187, feb 1989. doi:10.1016/0013-4694(89)90180-6.
- [112] Jonathon Shlens. A Tutorial on Independent Component Analysis. apr 2014. URL: <https://arxiv.org/abs/1404.2986v1>, arXiv:1404.2986.
- [113] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. ICLLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197, sep 2019. arXiv:1901.07915, doi:10.1016/J.NEUROIMAGE.2019.05.026.

- [114] Aulikki Hyrskykari. *Eyes in Attentive Interfaces: Experiences from Creating iDict, A Gaze-Aware Reading Aid*. PhD thesis, University of Tampere, 2006. URL: <http://acta.uta.fi>.
- [115] Russell L. DeValois and Karen K. DeValois. Spatial Vision. *Spatial Vision*, pages 1–382, jan 1991. URL: <https://academic.oup.com/book/321>, doi:10.1093/ACPROF:OSO/9780195066579.001.0001.
- [116] GitHub - Human-Pose Estimation (OpenCV). Last accessed on 2024-06-24. URL: <https://github.com/quanhua92/human-pose-estimation-opencv>.
- [117] TME. PGM1201-MP . Last accessed on 2024-06-02. URL: <https://www.tme.eu/en/details/pgm1201-mp/photoresistors/token/>.
- [118] Texas Instruments. LP5907MFX-3.3-NOPB. Last accessed on 2024-06-02. URL: <https://www.alldatasheet.com/datasheet-pdf/pdf/477973/TI1/LP5907MFX-3.3-NOPB.html>.
- [119] Rateb Katmah, Fares Al-Shargie, Usman Tariq, Fabio Babiloni, Fadwa Al-Mughairbi, and Hasan Al-Nashash. A Review on Mental Stress Assessment Methods Using EEG Signals. *Sensors (Basel, Switzerland)*, 21(15), aug 2021. URL: /pmc/articles/PMC8347831//pmc/articles/PMC8347831/?report=abstract<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8347831/>, doi:10.3390/S21155043.
- [120] Orestis Papaioannou and Steven J. Luck. Effects of Eccentricity on the Attention-Related N2pc Component of the Event-Related Potential Waveform. *Psychophysiology*, 57(5):e13532, may 2020. URL: /pmc/articles/PMC7160023//pmc/articles/PMC7160023/?report=abstract<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7160023/>, doi:10.1111/PSYP.13532.
- [121] Kamila Śmigasiewicz, Matthias Liebrand, Johannes Landmesser, and Rolf Verleger. How handedness influences perceptual and attentional processes during rapid serial visual presentation. *Neuropsychologia*, 100:155–163, jun 2017. doi:10.1016/J.NEUROPSYCHOLOGIA.2017.04.030.
- [122] Nestor Matthews and Leslie Welch. Left visual field attentional advantage in judging simultaneity and temporal order. *Journal of Vision*, 15(2):7–7, feb 2015. URL: <http://www.journalofvision.org/content/15/2/7>, doi:10.1167/15.2.7.
- [123] Monika Kiss, José Van Velzen, and Martin Eimer. The N2pc component and its links to attention shifts and spatially selective visual processing. *Psychophysiology*, 45(2):240, 2008. URL: /pmc/articles/PMC2248220//pmc/articles/PMC2248220/?report=abstract<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2248220/>, doi:10.1111/J.1469-8986.2007.00611.X.
- [124] Josef Faller, Neil Weiss, Nicholas Waytowich, and Paul Sajda. Brain–Computer Interfaces for Mediating Interaction in Virtual and Augmented Reality. In Chang S. Nam, Anton Nijholt, and Fabien Lotte, editors, *Brain–Computer Interfaces Handbook*, chapter 12, pages 235–252. CRC Press, Boca Raton, 1st edition, jun 2019. URL: [https://www.taylorfrancis.com/chapters/edit/10.1201/9781351231954-12/BRAIN](https://www.taylorfrancis.com/chapters/edit/10.1201/9781351231954-12).

Appendix A

A.1 Written consent



CONSENTIMENTO INFORMADO PARA PARTICIPAR EM ESTUDO	
<p>Por favor, leia com atenção a seguinte informação. Se achar que algo está incorreto ou pouco claro, não hesite em pedir mais informações através do email: up202103370@up.pt</p> <p>Se deseja participar no estudo solicitamos que preste o seu consentimento, assinando o documento no final.</p> <p>A participação no estudo é voluntária, podendo o participante em qualquer altura cessar a sua participação sem qualquer tipo de consequência, bastando para isso contactar o responsável através do e-mail acima.</p>	
1. DESCRIÇÃO DO ESTUDO	
<p>Titulo: Estudo dos mecanismos de atenção e de localização espacial de objetos de interesse em imagens apresentadas a ritmos elevados.</p> <p>Entidade Responsável: INESC-TEC</p> <p>Responsável Direto: Nome: <u>Miguel Rocha e Costa</u> Telemóvel: <u>924355241</u></p> <p>Descrição Geral: Este estudo decorre no âmbito do desenvolvimento de uma tese de mestrado do curso de Engenharia Biomédica para o ano letivo 2023/2024.</p> <p>Este estudo pretende estudar os mecanismos de atenção associados à deteção automática de objetos de interesse presentes em imagens aéreas. Em paralelo, será também estudado o mecanismo no cérebro que relaciona onde no campo de visão surge um objeto de interesse com o mapeamento de atividade em certas localizações do córtex específicas ao processamento visual. Este mecanismo de localização espacial é gerado e interessantemente pode ser extraído antes desta informação ser reportada à parte integradora do cérebro que permite que o participante conscientemente se aperceba que o objeto de interesse está de facto na imagem.</p>	
2. TRATAMENTO DE DADOS PESSOAIS	
<p>Procedimentos gerais: Este estudo conta com a colocação de elétrodos no couro cabeludo de modo a adquirir atividade elétrica neuronal proveniente de um processamento visual de imagens específico. Para este fim, será equipado com um dispositivo de recolha de sinal cerebral denominado por Eletroencefalograma (EEG). Durante a aquisição, ser-lhe-á pedido que preste atenção a uma série de imagens que irá ser transmitida num monitor. O protocolo de aquisição contará com a apresentação de um bloco de 150 imagens diferentes a um ritmo de 5 imagens por segundo, correspondendo a um período de 30 segundos contínuos de aquisição. Cada bloco de 150 imagens irá ser repetido 14 vezes com tempo de descanso entre cada bloco definido pelo participante. Assumindo um período de descanso médio de 13 segundos entre cada bloco, o protocolo completo terá um período de 10 minutos (7 min de aquisição e 3 min de descanso).</p> <p>As imagens apresentadas foram capturadas a partir de um drone que sobrevoou o jardim da FEUP, sendo as mesmas de dois tipos diferentes: a classe 1 de imagens irá corresponder a um ser humano com uma camisola vermelha situado num local aleatório do jardim da FEUP, sendo que o mesmo poderá aparecer em qualquer um dos quadrantes da imagem; a classe 2 de imagens irá corresponder a um local aleatório do jardim da FEUP mas sem a presença do ser humano. As classes de imagens surgirão com uma ordem aleatória em cada bloco em que a primeira classe irá ocorrer com menor frequência, isto é, 10% das ocasiões (15 imagens). Ser-lhe-á pedido que, no surgimento da imagem de classe 1 (com um ser humano), conte mentalmente a sua ocorrência até ao final de cada bloco de imagens. Para a correta aquisição do sinal do EEG, ser-lhe-á pedido que tente reduzir o ato de piscar os olhos e de movimentar a cabeça durante os períodos de aquisição. Todas as instruções de utilização serão explicadas pelo investigador.</p> <p>Dados tratados: Os dados recolhidos e tratados serão exclusivamente do EEG.</p>	

Figure A.1: Page 1 of the written consent signed by every participant



Riscos: Os riscos e desconforto causado não serão maiores dos que está sujeito na utilização de qualquer equipamento médico de aquisição de sinais biomédicos, nomeadamente ECG, EMG e EEG, ou na utilização de um equipamento de realidade virtual. **Contudo, devido à rápida apresentação de imagens serão rejeitados participantes com risco de epilepsia.** Algum desconforto/secura ocular poderá surgir durante a aquisição pelo que será fornecido gotas hidratantes entre cada bloco de aquisição. Para melhorar a condutividade do sinal será aplicado uma pasta condutora sobre o couro cabeludo de modo a reduzir o ruído comumente associado à aquisição de um sinal biomédico. Salienta-se que a utilização de elétrodos certificados pode causar alguma irritação na pele.

Benefícios: A sua participação neste estudo será de extrema importância para melhorar o estado da arte na área de estudo do cérebro para desenvolver tecnologias que englobem ajudar vários aspectos da vida humana como auxiliar missões de pesquisa visual associado a imagens aéreas e indivíduos com incapacidades motoras.

Compensações e/ou custos: Não há qualquer compensação financeira, nem custos associados a participação neste estudo.

Finalidade do tratamento:

Os dados recolhidos serão tratados de acordo com a legislação nacional e da UE aplicável e apenas serão usados pelos investigadores para os fins de investigação científica.

Responsável pelo Tratamento:

INESC TEC - Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, Campus da FEUP, Rua Dr. Roberto Frias, 4200 - 465 Porto.

Confidencialidade: não

Todos os dados adquiridos serão analisados para fins científicos no âmbito do estudo em causa; todos os envolvidos neste estudo têm um compromisso de confidencialidade e de não divulgação de dados ou informação de carácter pessoal dela retirada; os dados recolhidos serão tratados de forma pseudo-anónima e em conjunto, de forma a não ser identificada a sua participação individual no estudo. Alguns passos serão tomados para proteger a sua identidade individual, nomeadamente: a) a cada participante será atribuído um código; b) todos os dados recolhidos serão guardados em servidores locais do INESC TEC (instituição acolhedora do estudo) com acesso restrito, controlado e limitado apenas aos membros do grupo de investigação envolvidos; c) os dados recolhidos destinam-se a ser anonimizados ao fim de 6 meses, sendo nessa ocasião destruída toda a informação que direta ou indiretamente possa permitir identificar participantes d) qualquer vídeo contendo imagens que identifiquem os participantes serão guardados de forma segura, durante os 6 meses de projeto, e apenas é permitido o seu acesso a investigadores autorizados. No entanto, os participantes podem, ou não, permitir a utilização em apresentações ilustrativas / educativas bem como a partilha de imagens a investigadores terceiros, indicando "SIM" ou "NÃO" na frase seguinte:

- Compreendo que os investigadores possam querer utilizar alguma parte de um vídeo, apenas para fins científicos e/ou educativos. Dou autorização para tal, desde que o meu nome e rosto não sejam visíveis.

Por favor indique:

SIM NAO

Partilha de Dados Pessoais: Os dados/resultados anónimos e em conjunto poderão ser divulgados/publicados no âmbito de publicações científicas, podendo envolver equipas de investigação de diversas instituições.

Encarregado de Proteção de Dados: Para quaisquer questões, exercício de direitos do titular dos dados pessoais, pedidos ou reclamações relativas ao tratamento de dados pessoais, contacte por favor, o nosso encarregado de proteção de dados, através do contacto: dpo@inesctec.pt

Adicionalmente, se algum problema ou anomalia fisiológica for detetada nos dados recolhidos, deseja ser informado dessa situação?

Figure A.2: Page 2 of the written consent signed by every participant



<p>Por favor indique: _____ SIM _____ NÃO</p> <p>Caso responda positivamente indique o seu contacto (email ou tel): _____</p> <p>Direitos do Titular dos Dados: Enquanto titular dos dados, a lei reconhece-lhe os seguintes direitos: Informação, Acesso, Retificação, Apagamento, Oposição e Portabilidade. No caso de desistência não há prejuízo no tratamento de dados recolhidos até então, sendo estes tornados totalmente anonimizados. Para o exercício de algum dos seus direitos utilize o seguinte endereço de e-mail: miguel.r.costa@inesctec.pt</p> <p>O exercício dos direitos poderá ver-se afastado ou limitado, no respeito pelos termos e condições previstos na legislação nacional e da UE aplicável, na medida em que tal exercício seja suscetível de tornar impossível ou prejudicar gravemente a obtenção dos objetivos do tratamento para fins de investigação e apenas na medida do necessário para a prossecução daqueles fins.</p> <p>A lei confere-lhe, igualmente, o direito de apresentação de queixas perante uma Autoridade europeia de supervisão, sendo que em Portugal a Autoridade competente é a Comissão Nacional de Proteção de Dados (www.cnpd.pt).</p>		
<p>3. TERMO DE CONSENTIMENTO INFORMADO</p> <p>1. Li e compreendi a informação sobre o estudo, incluindo a identidade do Responsável, o tipo de dados que vai ser recolhido, o objetivo da recolha e do respetivo tratamento. <input type="checkbox"/></p> <p>2. Li e compreendi a informação sobre como os dados serão armazenados e durante quanto tempo, incluindo o que acontecerá aos meus dados no caso de desistir de participar no estudo. <input type="checkbox"/></p> <p>3. Foi-me dada a oportunidade de fazer perguntas e de esclarecer todas as dúvidas sobre este estudo. <input type="checkbox"/></p> <p>4. Compreendo que posso desistir desta participação em qualquer momento, sem necessitar de dar justificações e sem que sofra penalizações ou que questionem as minhas razões. <input type="checkbox"/></p> <p>5. Percebi de que forma poderei comunicar a minha desistência, bem como exercer os meus direitos enquanto titular dos dados. <input type="checkbox"/></p>		
<p>O Participante: <i>Declaro ter lido e compreendido este documento, bem como as informações verbais que me foram fornecidas previamente. Desta forma, aceito participar neste estudo e permito a utilização dos dados que forneço de forma voluntária.</i></p> <p>Nome: Assinatura: Data: / /</p>		

Figure A.3: Page 3 of the written consent signed by every participant

A.2 Signal quality estimation analysis

A.2.1 Frequency spectra

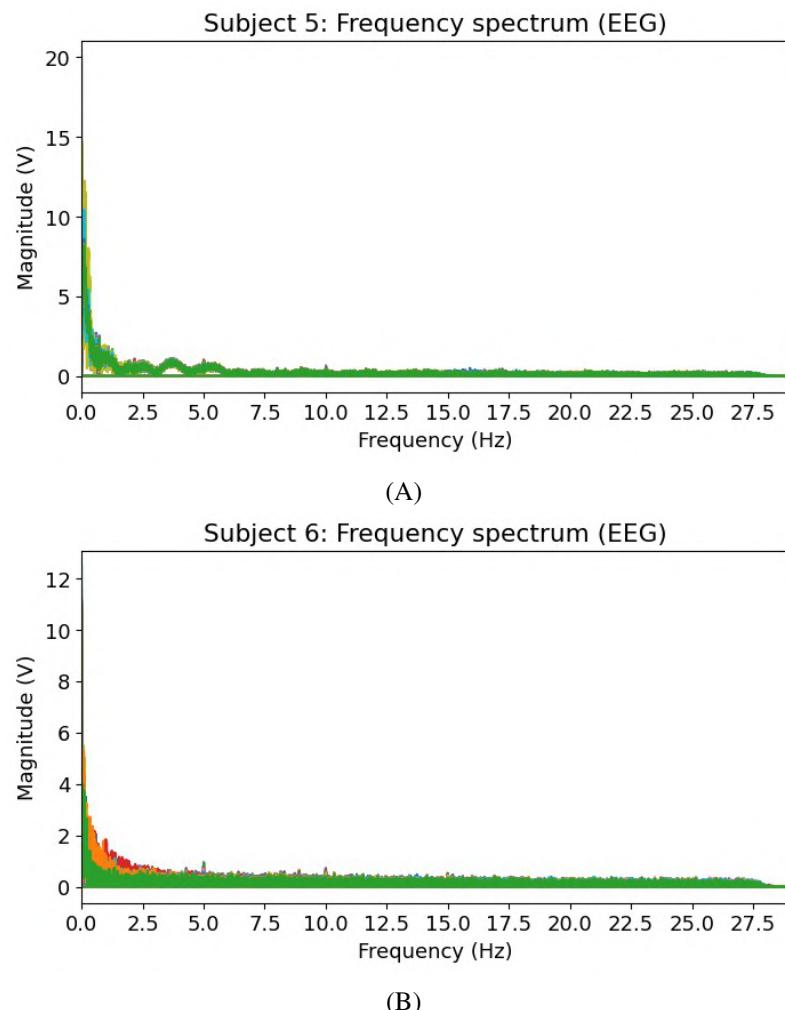


Figure A.4: From left to right, the frequency spectra of subjects 5 and 6.

A.2.2 Electrode bridging

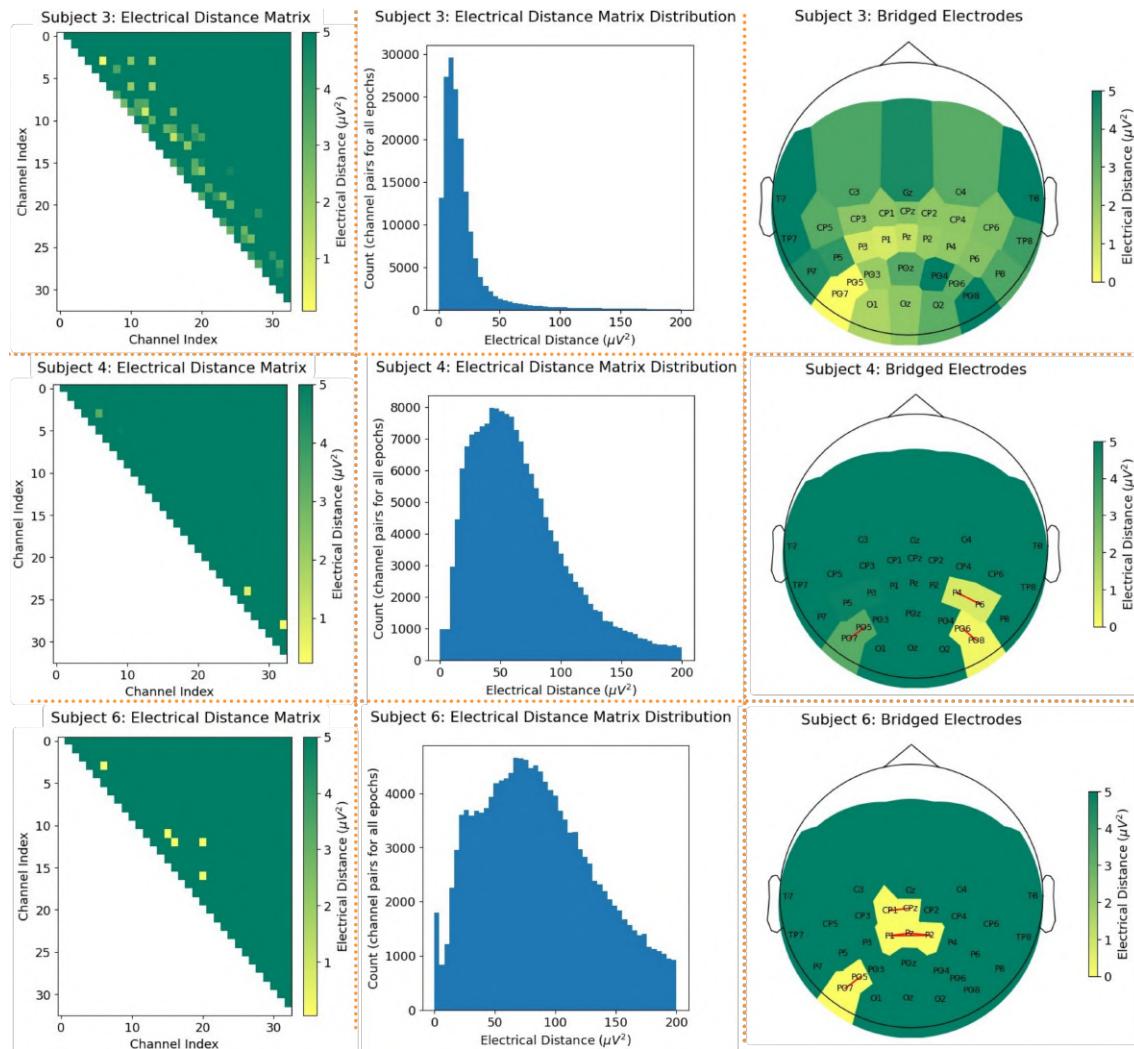


Figure A.5: From top to bottom, electrode bridging analysis for subjects 3, 4, and 6.

A.2.3 SSVEP SNR estimation

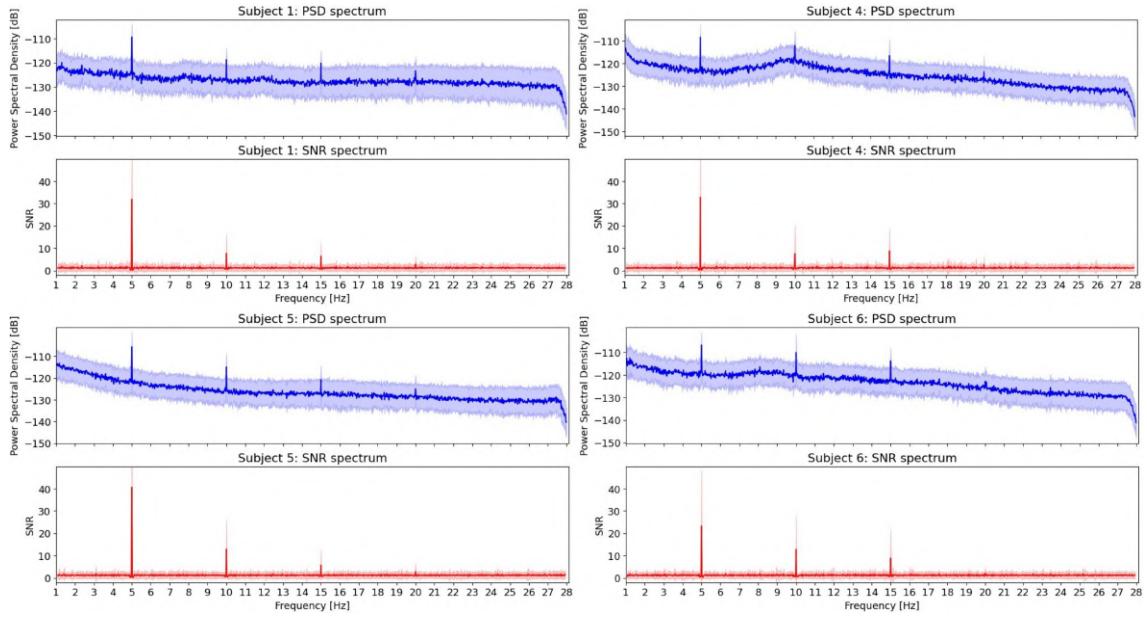


Figure A.6: The PSD and SNR estimation metrics of subjects 1 (top left corner), 4 (top right corner), 5 (lower left corner), and 6 (lower right corner).

A.3 ERP Grand-Averages (*Testing paradigm*)

A.3.1 P3

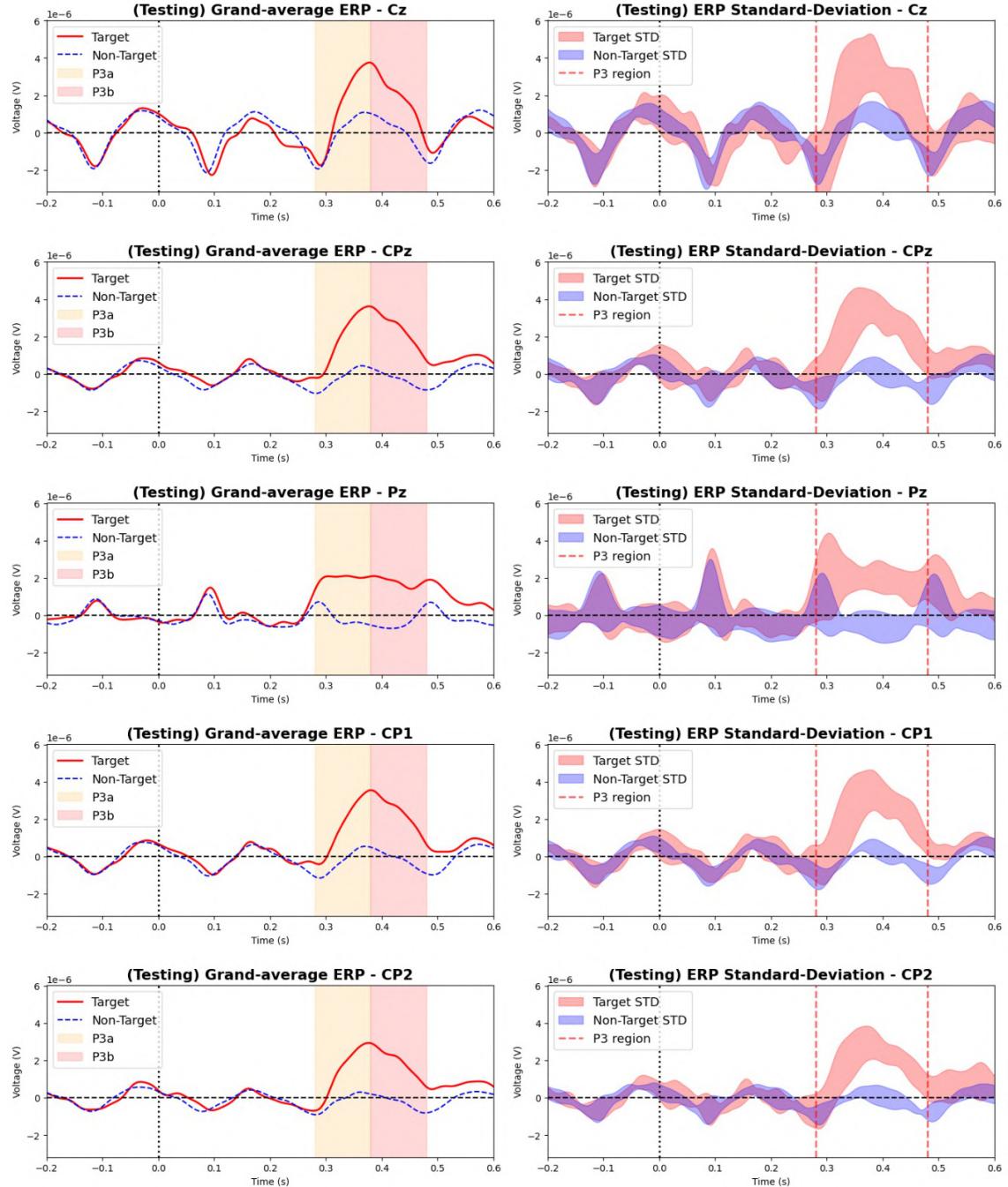


Figure A.7: The Grand-Average P3 (left) and its standard deviation (right) along the 5 analyzed channels for the **Testing** experiment.

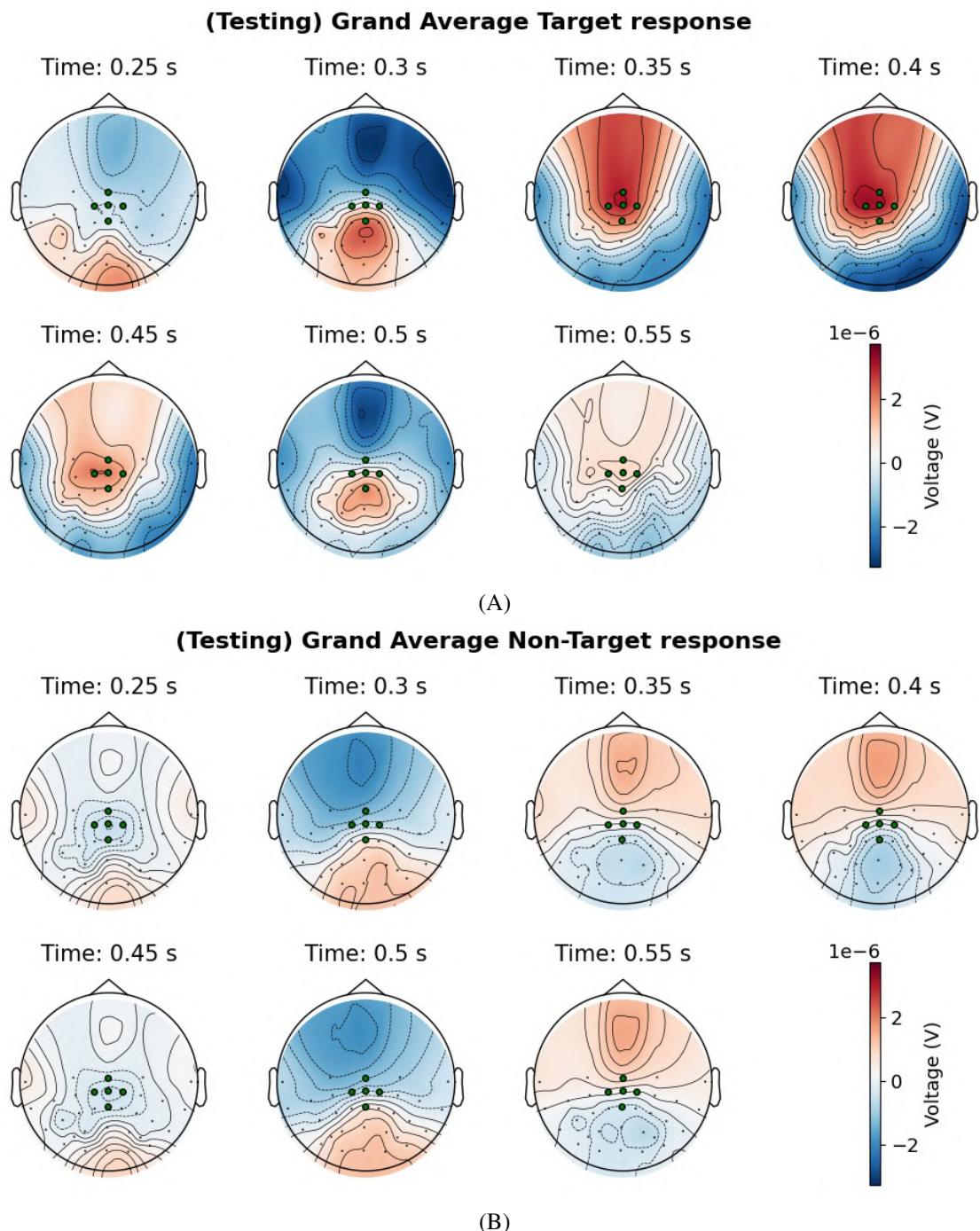


Figure A.8: Grand average topographies of the **Testing** paradigm.

A.3.2 N2pc

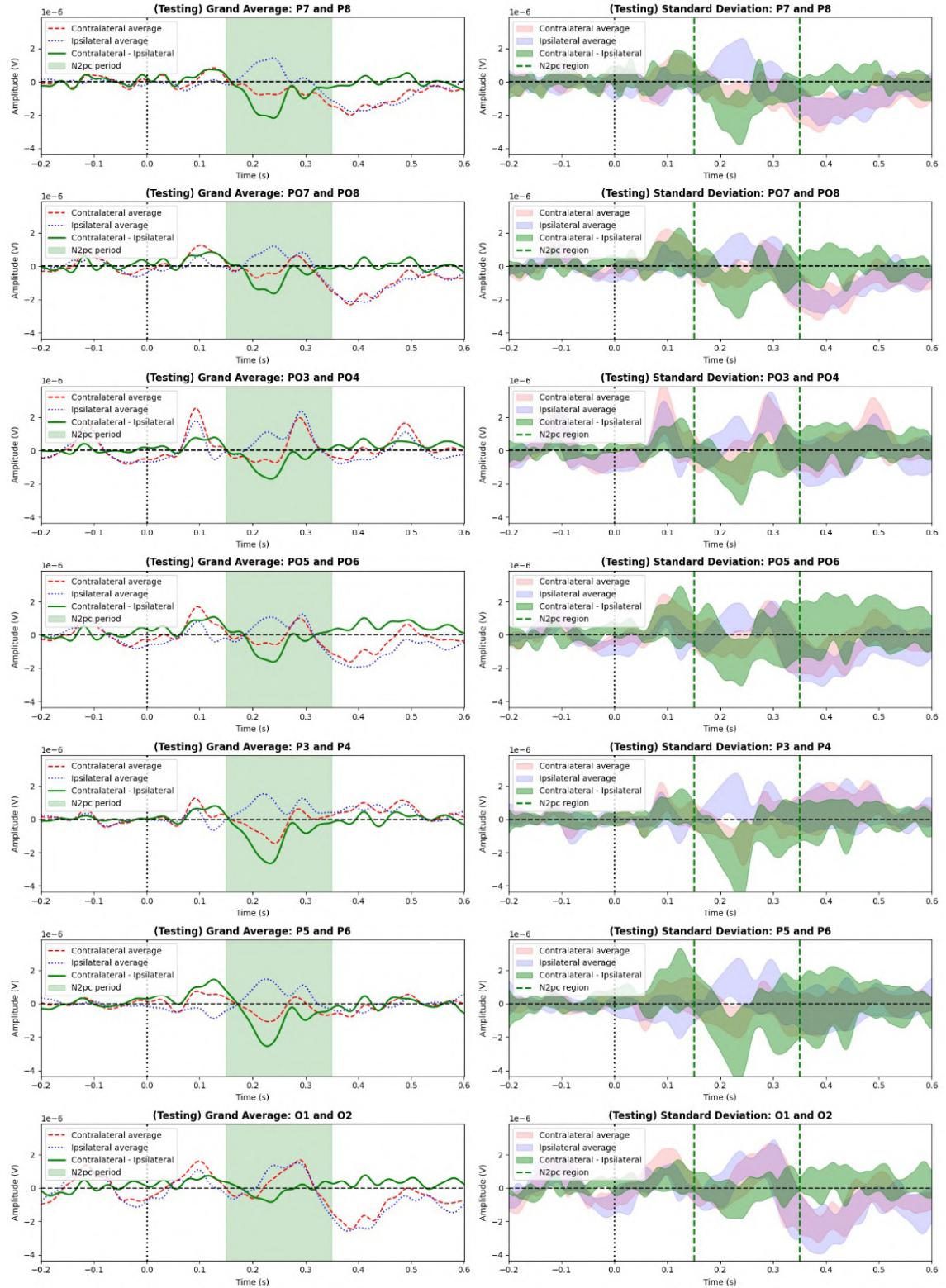


Figure A.9: The Grand-Average N2pc (left) and its standard deviation (right) along the 7 analyzed channels pairs for the **Testing** experiment.

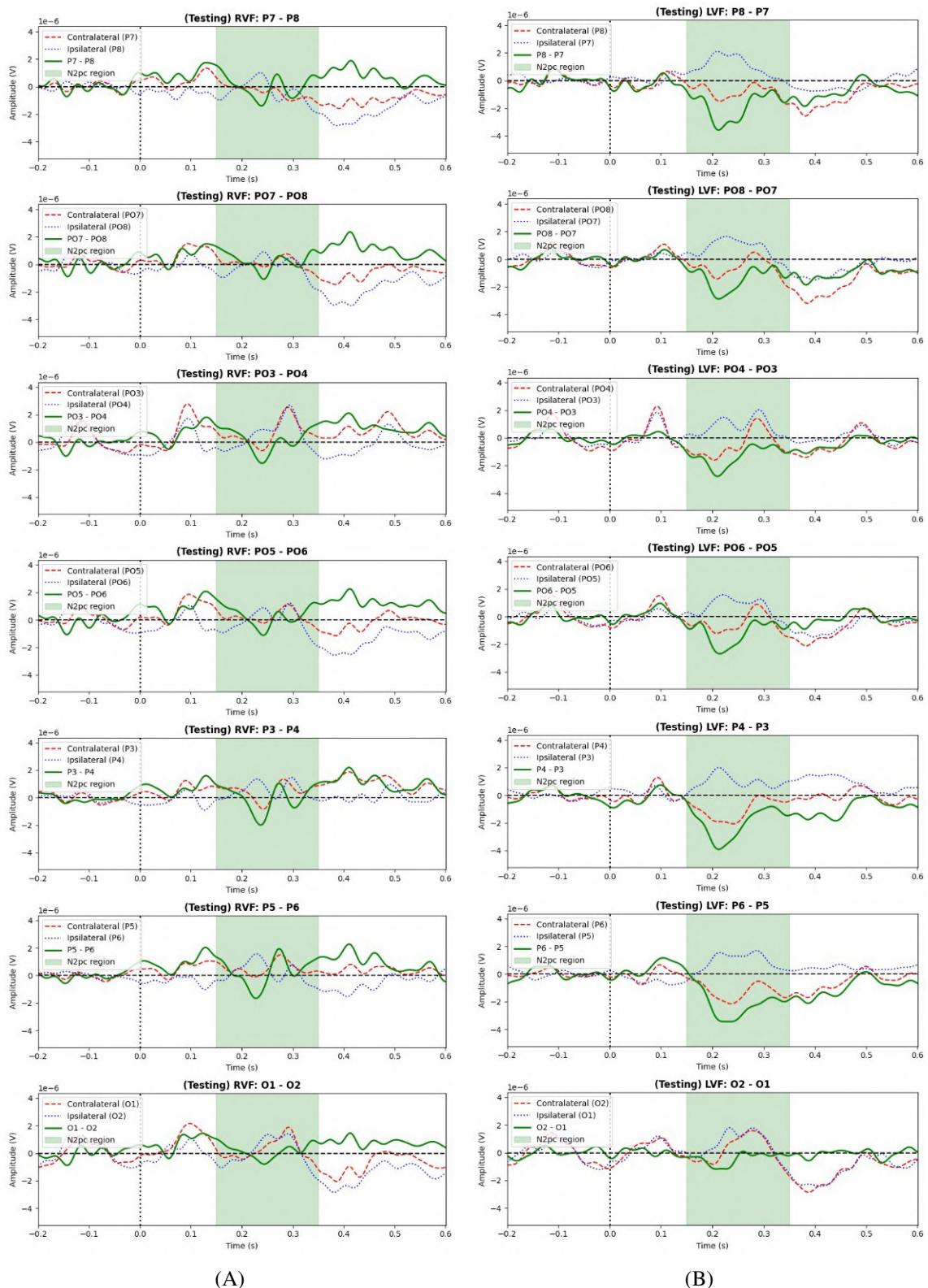


Figure A.10: The grand averaged RVF N2pc (A) and LVF N2pc for the **Testing** session.

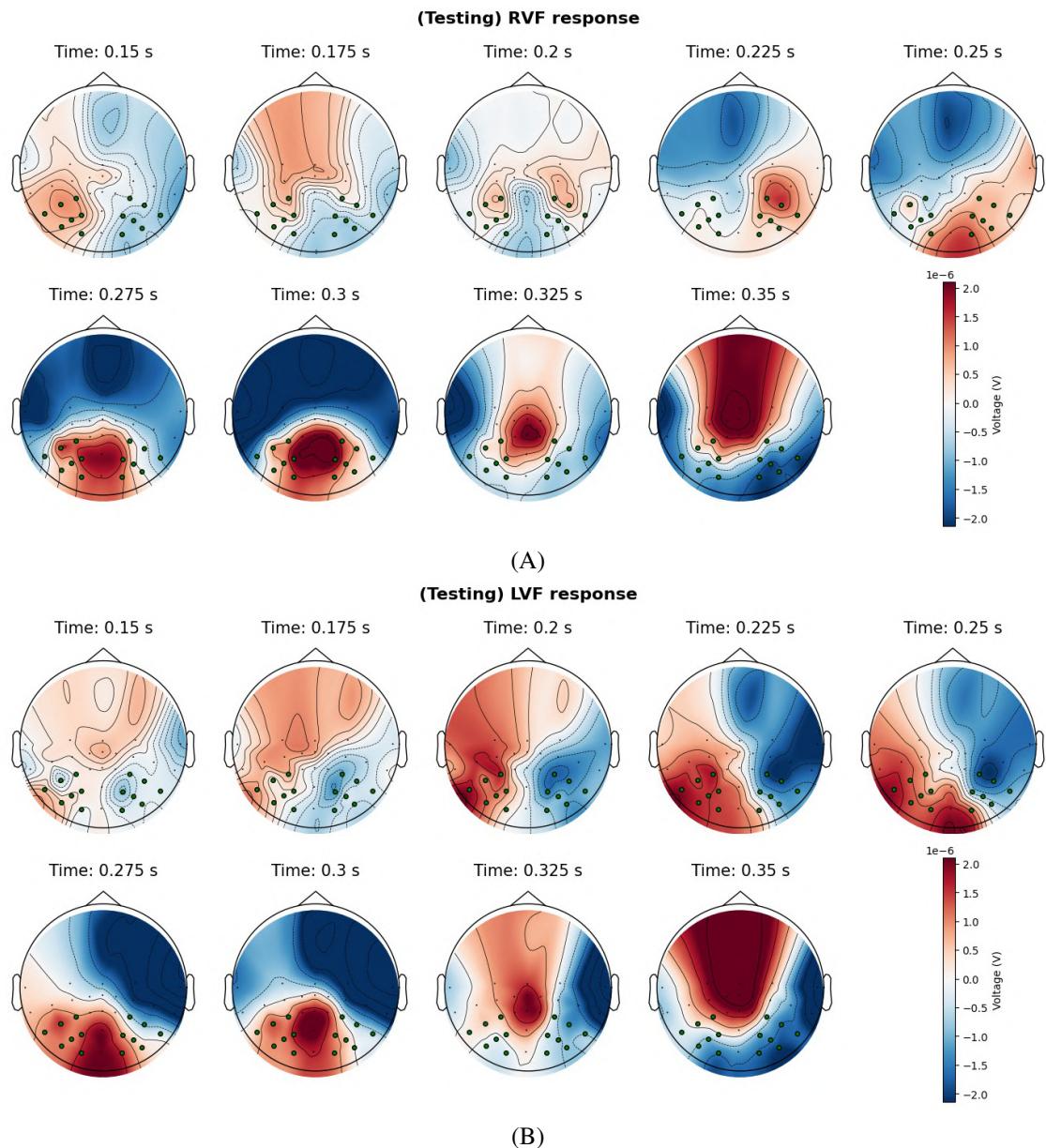


Figure A.11: The grand averaged RVF (A) and LVF (B) topographies for the **Testing** session.

A.3.3 N2pcb

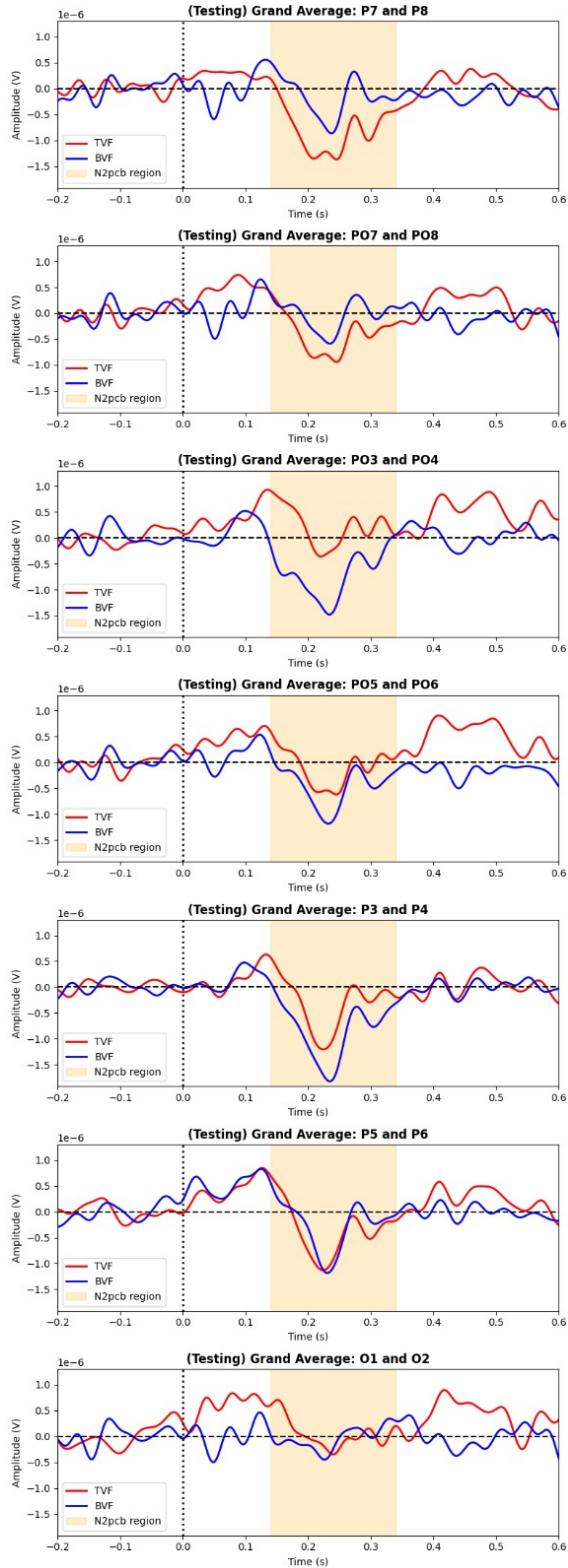


Figure A.12: The Grand-Average N2pcb along the 7 analyzed channels pairs for the **Testing** experiment.

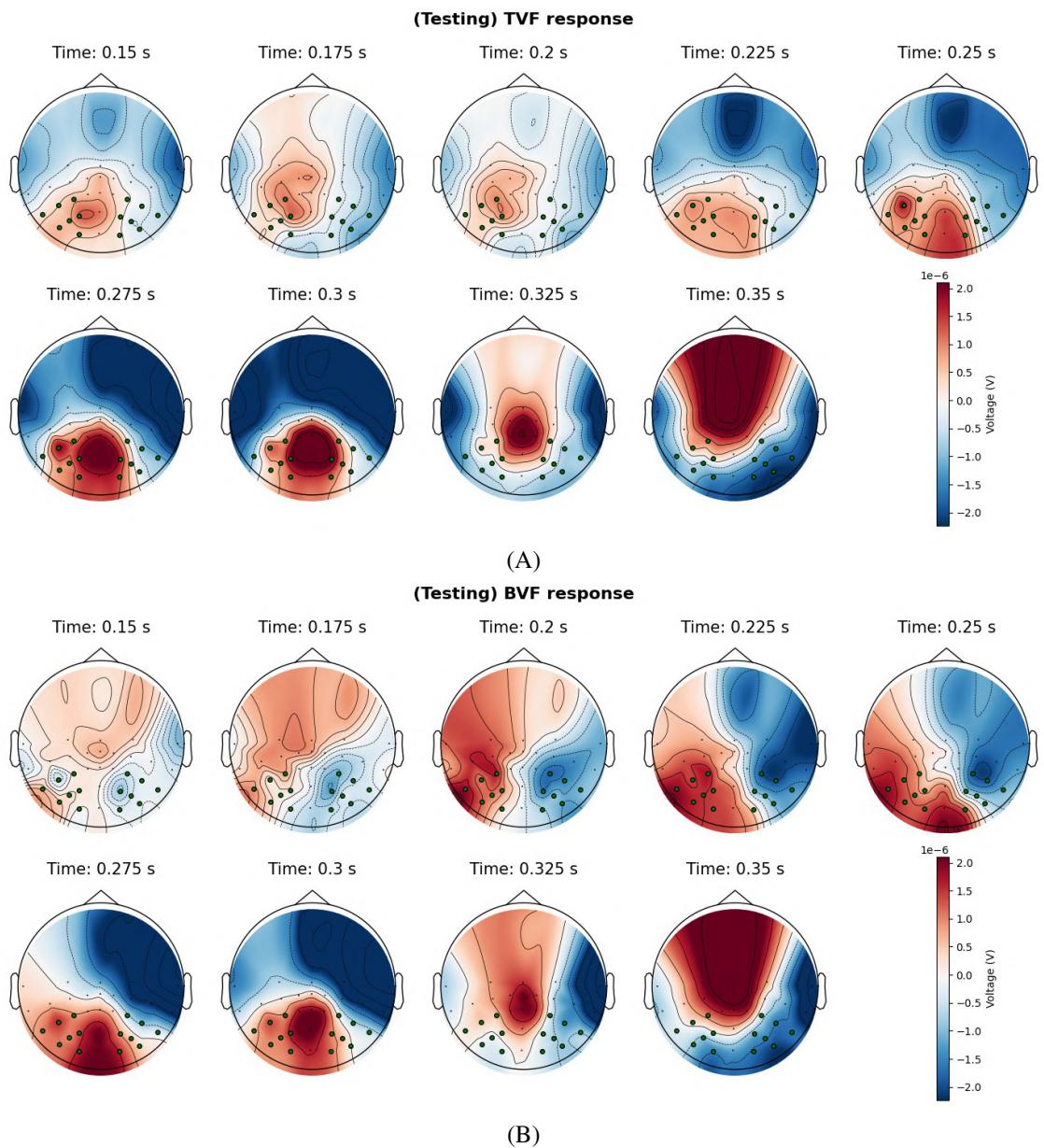


Figure A.13: The grand averaged TVF (A) and BVF (B) topographies for the **Testing** session.

A.3.4 Four quadrant N2pc

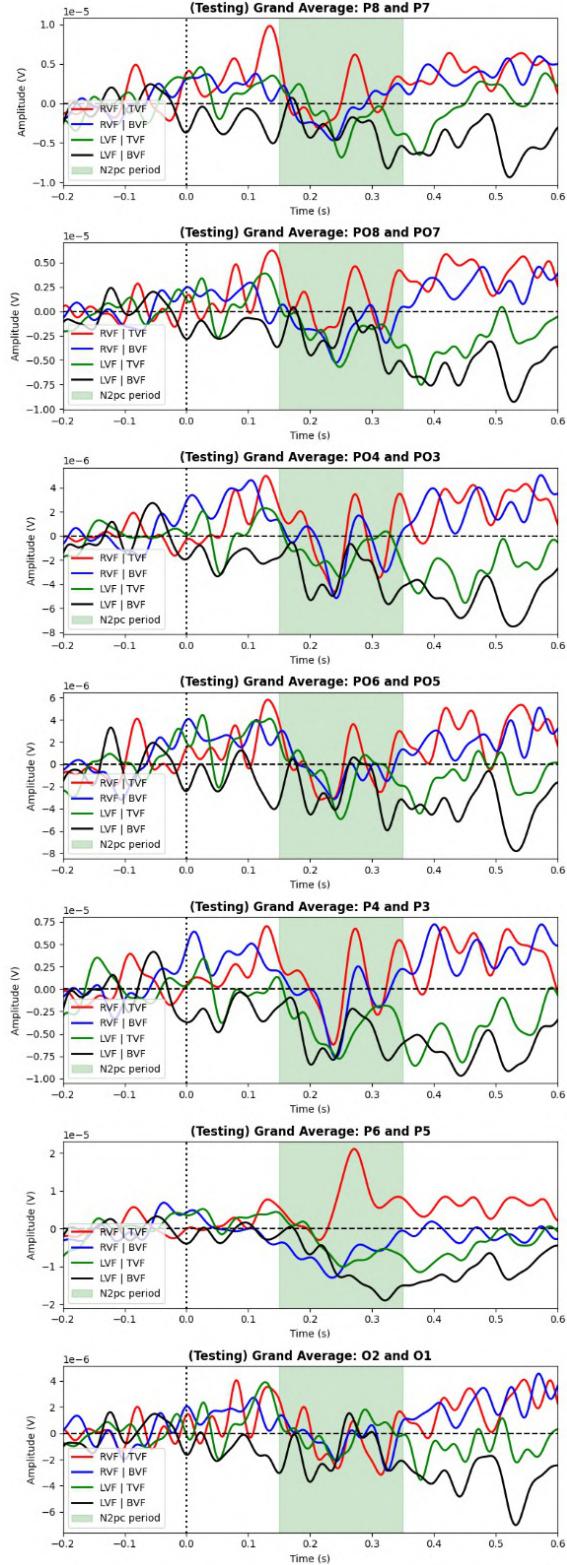


Figure A.14: The Grand-Average N2pc along the 7 channel pairs divided into four quadrants (RVF-TVF vs LVF-TVF vs RVF-BVF vs LVF-BVF) for the **Testing** experiment.

A.4 BCI performance

A.4.1 Horizontal centroid estimation (N2pc) - Train/Validation set

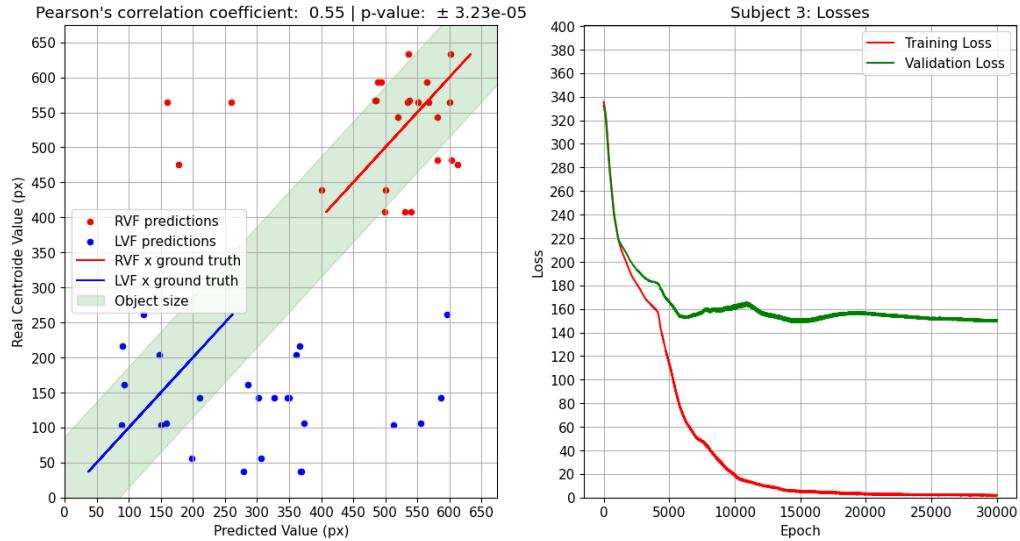


Figure A.15: Subject 3 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model's train and validation loss (right).

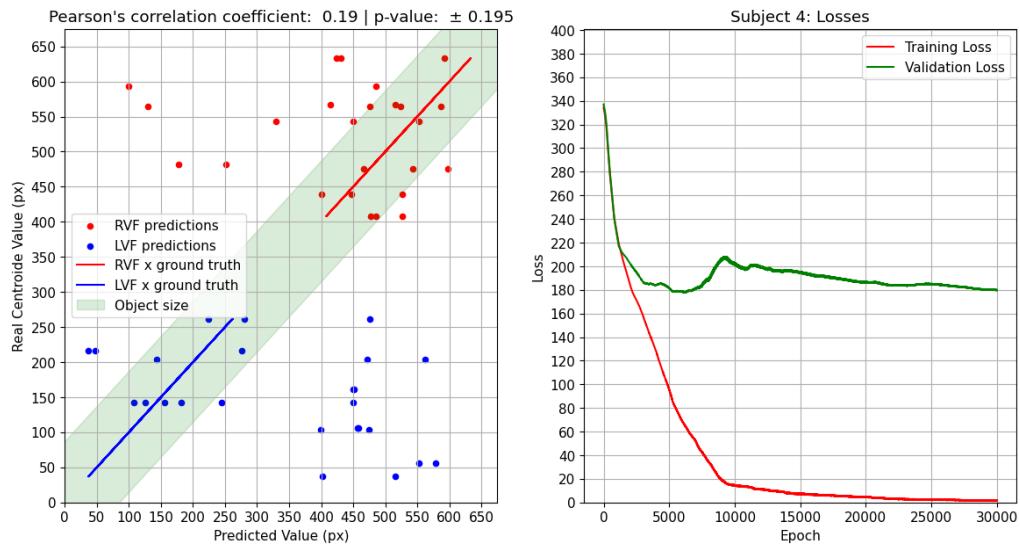


Figure A.16: Subject 4 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model's train and validation loss (right).

A.4.2 Horizontal centroid estimation (N2pc) - Testing set

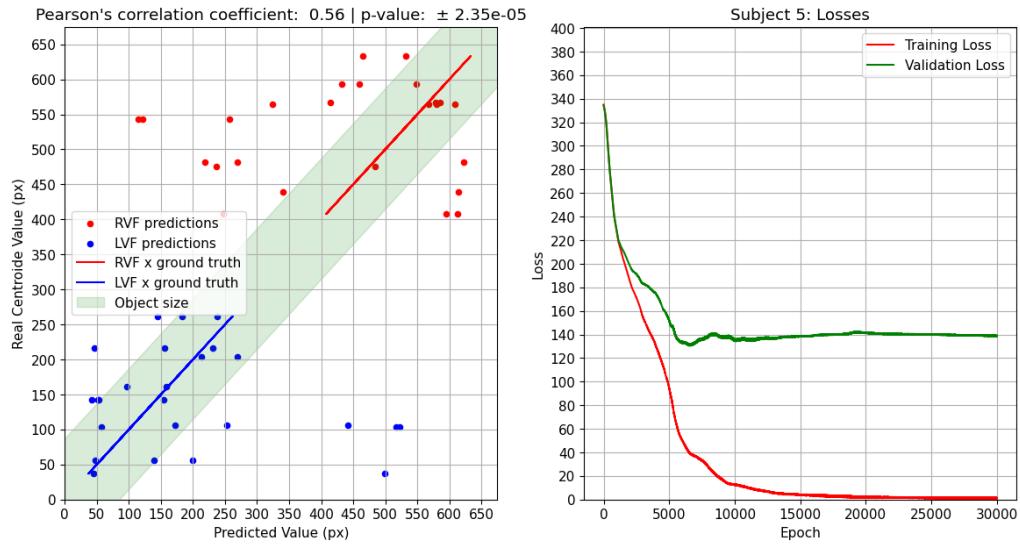


Figure A.17: Subject 5 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model's train and validation loss (right).

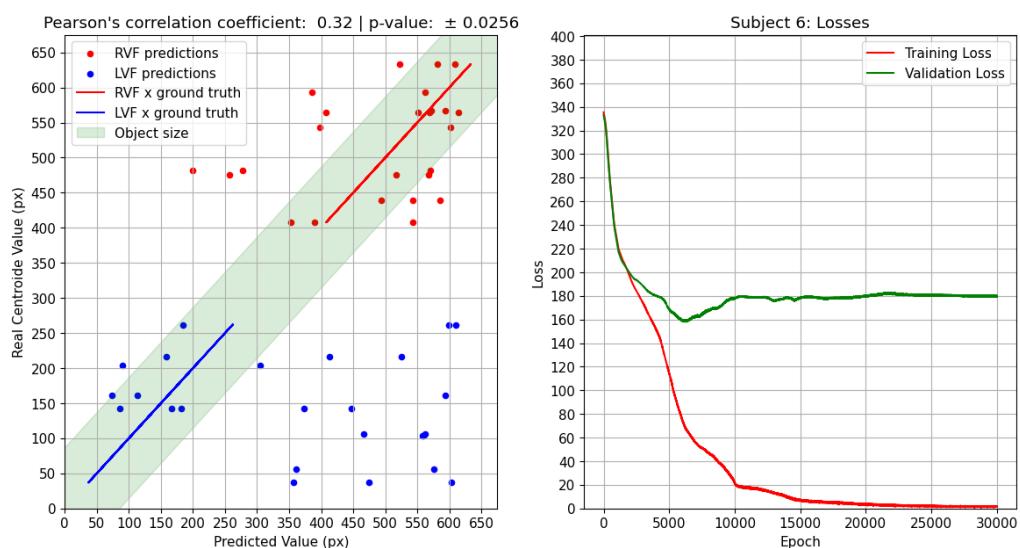


Figure A.18: Subject 6 performance in predicting the correct pixel where a lateral target appears within the image (left) and the model's train and validation loss (right).

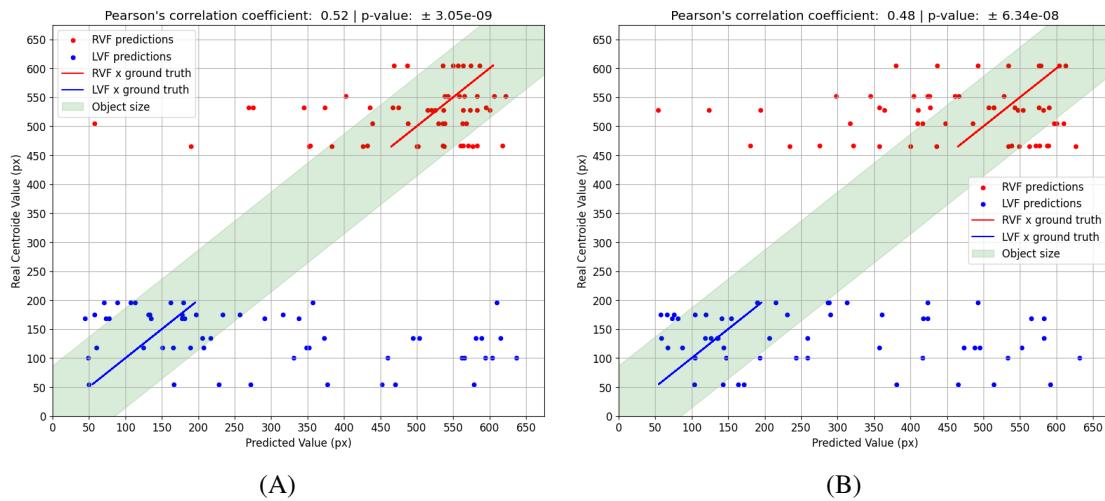


Figure A.19: Subject 3 (A) and 4 (B) performance in predicting the correct pixel where a lateral target appears within the image for the single-trial testing data.

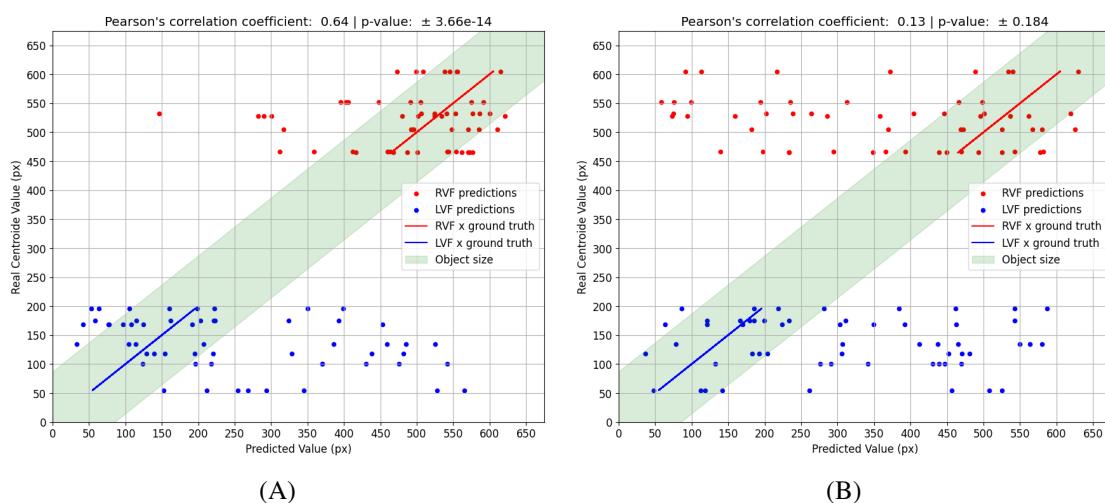


Figure A.20: Subject 5 (A) and 6 (B) performance in predicting the correct pixel where a lateral target appears within the image for the single-trial testing data.