上海交通大学博士学位论文

头皮脑电伪迹消除与脑源定位特征提取方法研究

姓　　　名：Eva Sarah Serene FLURY

学　　　号：0130209126

导　　　师：朱向阳教授

院　　　系：机械与动力工程学院

学 科/专 业：机械电子工程

申 请 学 位：工学博士

2024 年 2 月 23 日

A Dissertation Submitted to

Shanghai Jiao Tong University for Doctoral Degree

INNOVATIVE ARTEFACT ELIMINATION AND SOURCE LOCALIZATION-BASED FEATURE EXTRACTION FOR EEG BCI PIPELINES

Author: Eva Sarah Serene FLURY

Supervisor: Prof. Xiangyang Zhu

School of Mechanical Engineering

Shanghai Jiao Tong University

Shanghai, P.R. China

February 23rd, 2024

上海交通大学

学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全知晓本声明的法律后果由本人承担。

学位论文作者签名：

日期：2023 年 11 月 27 日

上海交通大学

学位论文使用授权书

本人同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。

本学位论文属于：

☑**公开论文**

□**内部论文**，保密□1 年/□2 年/□3 年，过保密期后适用本授权书。

□**秘密论文**，保密____ 年（不超过 10 年），过保密期后适用本授权书。

□**机密论文**，保密____ 年（不超过 20 年），过保密期后适用本授权书。

（请在以上方框内选择打 "✓"）

学位论文作者签名：　　　　　　　指导教师签名：

日期：2023年 11 月 27日　　　　日期：2023 年 11 月 27日

摘　要

在蓬勃发展的神经科学领域，脑机接口（Brain-Computer Interfaces, BCI）已经成为一项突破性的技术，促进了人脑与外部设备或应用程序之间的直接通信。通过利用神经信号的力量，脑机接口有可能彻底改变医疗健康、辅助技术和人机交互等多个领域。多种记录模式可用于捕捉大脑活动，其中脑电图（Electroencephalography, EEG）设备因其佩戴方便、舒适且具有成本效益而备受青睐，它提供了高时间分辨率来无创地监测用户意图。

这项研究致力于对想象运动进行深入研究，特别关注于右手和左手的抓握运动相关的辨别模式。本研究的最终目标是在不需要神经反馈或额外间接信息的情况下，实现对用户意图的最高精度预测。考虑到脑电在信噪比和空间分辨率方面存在局限性，研究有必要使用源定位技术以揭示大脑内复杂的动态相互作用和连接。为了清晰地区分本研究中引入的新算法与现有算法，原始缩写词与名称将用粗体标注。

在绝大多数情况下，确保脑电图数据分析的准确性和可靠性需要电极功能的均匀性和信号的清洁度。这需要解决两个主要目标：消除统计上显著的（和间接的）伪迹与纠正任何坏通道。自发眨眼代表一种主要的生理干扰，与感兴趣的神经信号同时发生，平均概率为 10%。

基于以上考虑因素，本研究设计了一个多模态数据记录，旨在全面记录眨眼的各方面信息，其中包括使用脑电图、眼电图、肌电图、眼动仪和高速相机能够保存的信息。另外，本研究涉及的范式还包括运动执行、P300 和稳态视觉诱发电位，是因为纳入这些范式将有助于增进对想象运动期间潜在神经机制的全面理解。

本研究首先在前瞻性能量分析中使用拟合分布蒙特卡罗（Fitted Distribution Monte Carlo, **FDMC**）模拟方法进行先验样本量估计，以确保收集到足够的数据，以深入理解皮层现象的复杂性。与正态分布相比，**FDMC** 对样本量的要求降低了 35%，表现出优越性。特别是在资源或时间有限的情况下，**FDMC** 方法具有显著的吸引力。

由此产生的多模态数据集有助于开发一个稳健的模型，用于准确地分类和校正眨眼伪迹，以及从 EEG 数据中去除坏通道。自适应眨眼校正与漂移去噪（Adaptive Blink Correction and De-drifting, **ABCD**）算法已被证明能够提高整体数据质量，并展现出比传统的技术发展现状方法（即独立分量分析，Independent Component Analysis, ICA）或伪迹空间重建校正（Artifact Subspace Reconstruction, ASR）更好的结果。分类准确率及其 95% 置信水平下的置信区间显示，**ABCD** 算法的平均分类准确率为 93.81%

[74.81%；98.76%]，而 ICA 的平均分类正确率为 79.29% [57.41%；92.89%] 与 ASR 为 84.05% [62.88%；95.31%]。

本研究使用了多种创新算法的组合用以提取跨时间、空间和频率域中最显著的感兴趣信号（Signal of Interest, SOI）。首先，基于 **ABCD** 清理后的脑电图数据计算脑源定位。本研究应用了脑源定位时空特征（Source Localized Spatio-Temporal, **SLST**）提取方法，通过分析多个试次间的相似性来捕捉 SOI 的时空特征。为了进一步增强分类过程，利用从 **SLST** 特征提取算法中空间和时间位置所导出的协方差矩阵，实现了双分类器（Dual Classifier, **DC**）。每个具体类别的 Fréchet 均值以及所有其他协方差矩阵，在核心通道选择过程中计算得到，其中是基于 Meta 分析确定的。

本研究同时对比了不同频带（delta、theta、alpha 和 beta）、滤波器类型（Butterworth、Chebyshev 和 Elliptic）以及重新参考方法（例如，共同平均参考、大和小拉普拉斯空间滤波器）之间的分类效果，以找到这些处理步骤的最佳组合。为了确保公平的比较，我们首先使用 **ABCD** 进行数据预处理，然后使用大拉普拉斯空间滤波器进行重新参考以及 beta 频带的 Butterworth 带通滤波器进行处理。这些滤波器的组合已被证明能够获得最佳的准确率。

新引入的一致性计算用于验证分类器的稳定性，因此可以用于确认它们之间的层次结构，结果表明，我们的 **SLST+DC** 方法的平均分类准确率为 95.03%（SD = 3.41%）。常用的公共空间模式（Common Spatial Pattern, CSP）与线性判别分析（Linear Discriminant Analysis, LDA）结合，当与支持向量机（Support Vector Machine, SVM）相结合时，平均分类准确度为 89.16%（SD = 1.76%）或 88.99%（SD = 1.60%）。还与较新的方法进行了比较，如达到 81.13%（SD = 4.64%）的最小离黎曼平均值（Minimum Distance to Riemannian Mean, MDRM）和 86.09%（SD = 4.48%）的切线空间（Tangent Space, TS）。然后使用混淆矩阵对不同方法所得到的结果进行了对比。

概括地说，混淆矩阵（Confusion Matrix, CM）由两个独立的样本（正样本和负样本）组成，每个样本都遵循二项式分布。为了提高 CM 的效用，本研究计算与每个样本相关的置信区间（Confidence Interval, CI）。利用依赖于所选 CI 近似的距离分离（Distance Separation, **DS**）方法，展示了最小样本量的互补估计。**DS** 方法可以作为 **FDMC** 的一种可行的实验设计替代方案，专门用于评估具有已知准确率的两个分类器之间的统计差异。最后，本研究还提出了一种通过 CI 的标准偏差（Standard Deviation of CI, **SDCI**）方法来估计准确率，并将其应用于 CM 的视觉表示，称为混淆椭圆（Confusion Ellipses, **CE**）。经过与 PyCM（一个专门用于 CM 评估和比较的

Python 库）和 SDCI 进行比较后发现，PyCM 的计算似乎低估了零附近的 CI 平均宽度。本研究还测试了各种 CI 近似值，以评估 CI 近似值选择的影响，这可作为设计新实验的参考，以比较具有已知全局分类准确率的两个处理方法。这些新颖方法（即 **SDCI、DS、CE**）不仅能适用于脑机接口领域，而且能够应用于广泛的领域和学科。

关键词：脑机接口，脑电图，自发性眨眼，伪迹去除，试验样本量估计，通道选择，脑源定位，eLORETA，混淆矩阵，准确率的置信区间

ABSTRACT

In the thriving realm of neuroscience, Brain-Computer Interfaces (BCI) have emerged as a groundbreaking technology that facilitates direct communication between the human brain and external devices or applications. By harnessing the power of neural signals, BCI have the potential to revolutionize various domains, including healthcare, assistive technology, and human-computer interaction. Multiple recording modalities are available for capturing brain activity. Among them, electroencephalography (EEG) devices stand out as a wearable, comfortable, and cost-effective solution that provides high temporal resolution to non-invasively monitor basic user intentions.

This research is dedicated to the rigorous investigation of motor imagery, with a specific focus on discerning patterns associated with grasping movements of the right and left hands. The ultimate objective is to achieve the highest possible classification accuracy of user intent without the need for neurofeedback or additional indirect information. The limitations of EEG in terms of signal-to-noise ratio and spatial resolution necessitate the use of source localization techniques, which can unveil the intricate dynamic interactions and connections within the brain. To easily distinguish between existing algorithms and the new ones introduced in this research, the original acronyms and appellations will be noted in bold.

Ensuring the accuracy and reliability of EEG data analysis requires the homogeneity of electrode functioning and the cleanliness of signals in the vast majority of cases. This entails addressing two primary objectives: eliminating statistically significant (and consequential) artifacts and rectifying any malfunctioning electrodes. The spontaneous blink represents a major physiological disturbance, occurring simultaneously with the neural signal of interest with an average probability of 10%.

These considerations led to the design of a multi-modal dataset targeted at comprehensively recording every aspect of the blinks using EEG, electrooculography, electromyography, eye tracker, and high-speed camera. Additional paradigms, including motor execution, P300, and steady-state visual evoked potentials have also been investigated due to their relevance in inferring a comprehensive understanding of the underlying neural mechanisms during MI.

The Fitted Distribution Monte Carlo (**FDMC**) simulation is first conducted for a priori

sample size estimation during a prospective power analysis. The goal is to ensure that sufficient data will be collected to truly comprehend the intricacies of the cortical phenomenon. **FDMC** demonstrates its superiority by lowering the sample size requirement by 35% compared to the Normal distribution (and by seven times compared to traditional power tables). **FDMC** is of particular interest when only limited resources or time are available.

The resulting multi-modal dataset facilitated the development of a robust model for accurately classifying and correcting blink artifacts as well as removing bad channels from the EEG data. The Adaptive Blink Correction and De-drifting (**ABCD**) algorithm has proved to improve the overall data quality and displays significantly better results than the state-of-the-art, i.e., Independent Component Analysis (ICA), or Artifact Subspace Reconstruction (ASR). The classification accuracy, along with its confidence interval at 95% confidence level, reveals a mean classification accuracy of 93.81% [74.81%; 98.76%] for **ABCD** against 79.29% [57.41%; 92.89%] for ICA or 84.05% [62.88%; 95.31%] for ASR.

A combination of various innovative algorithms is implemented to extract the most prominent signal of interest (SOI) across the temporal, spatial, and frequential domains. Source localization is first computed on the **ABCD**-cleaned EEG data. The Source Localized Spatio-Temporal (**SLST**) features approach is applied to capture the spatio-temporal characteristics of the SOI by analyzing the similarities across multiple trials. To further enhance the classification process, a Dual Classifier (**DC**) is implemented, utilizing both the spatial locations and time-derived covariance matrices from the **SLST** feature extraction algorithm. The class-specific Fréchet means, along with all other covariance matrices, are computed at the **Core Channel Selection**, which is determined based on a comprehensive meta-analysis.

Comparisons across different frequency bands (delta, theta, alpha, and beta), filter types (Butterworth, Chebyshev, and Elliptic), and re-referencing, e.g., Common Average Reference, Large and Small Laplacian spatial filters, are also carried out to reveal the optimized combination of all these processing steps. To ensure a fair comparison, all data are first pre-processed with **ABCD**, re-referenced with the Large Laplacian filter and filtered with a beta pass-band Butterworth filter, as this combination of various filters proved to yield the best accuracy.

A new introduced **Consistency** computation serves to validate classifiers' stability and can thus be used to confirm the hierarchy between them, revealing that our **SLST+DC** method

yields a mean classification accuracy of 95.03% (SD = 3.41%). The commonly employed Common Spatial Pattern (CSP) coupled with Linear Discriminant Analysis (LDA) yields 89.16% (SD = 1.76%) or 88.99% (SD = 1.60%) when coupled with Support Vector Machine (SVM). Comparisons were also made with newer methods such as the Minimum Distance to Riemannian Mean (MDRM) that achieved 81.13% (SD = 4.64%) and Tangent Space (TS) 86.09% (SD = 4.48%). All the results issued from the various pipelines are then compared using their confusion matrices.

More generally, confusion matrices (CM) consist of two independent samples (positive and negative), each following a Binomial distribution. To extend their utility, the confidence intervals (CI) are calculated for the probabilities associated with each sample. A complementary estimation of the minimum sample size is presented based on the Distance Separation (**DS**) method relying on the chosen CI approximation. **DS** can serve as a viable alternative to **FDMC** for experimental design, specifically aimed at evaluating the statistical distinction between two classifiers possessing known accuracies. Finally, a method to estimate the accuracy though standard deviation of CI (**SDCI**) is presented and applied to visual representations of CM, called Confusion Ellipses (**CE**).

Comparisons with PyCM, a Python library dedicated to CM evaluation and comparison, shows that their computations seem to underestimate the CI mean widths around zero compared to **SDCI**. Various CI approximations are also tested to assess the influence of the CI approximation choice and can serve as reference when designing a new experiment with the goal of comparing two pipelines with known global classification accuracies. The applicability of these novel methodologies (i.e., **SDCI**, **DS**, **CE**) extends far beyond the confines of the BCI field, encompassing a wide range of domains and disciplines.

Key words: Brain-Computer Interface (BCI), Electroencephalography (EEG), Spontaneous blinking, Artifact removal, Sample size calculation, Channel selection, Source Localization, eLORETA, Confusion Matrix (CM), Accuracy confidence interval

Contents

List of Figures

List of Tables

List of Algorithms

Nomenclature

ABCD[*]	Adaptive Blink Correction and De-drifting	自适应眨眼校正与漂移去噪
ACI	Asymptotic Confidence Interval	渐近置信区间
BCI	Brain-Computer Interface	脑机接口
CAR	Common Average Reference	共同平均参考
CCS	Core Channel Selection	核心通道选择
CE	Confusion Ellipses	混淆椭圆
CI	Confidence Interval	置信区间
CM	Confusion Matrix	混淆矩阵
CSP	Common Spatial Patterns	公共空间模式
DC	Dual Classifier	双分类器
DS	Distance Separation	距离分离
ECoG	Electrocorticography	脑皮层电图
EEG	Electroencephalography	脑电图
EMG	Electromyography	肌电图
EOG	Electrooculography	眼电图
FDMC	Fitted Distribution Monte Carlo	拟合分布蒙特卡罗
fMRI	Functional Magnetic Resonance Imaging	功能性磁共振成像
FN	False Negative	伪阴性
FP	False Positive	伪阳性
ICA	Independent Component Analysis	独立分量分析
LDA	Linear Discriminant Analysis	线性判别分析
LFP	Local Field Potential	局部场电位
MDRM	Minimum Distance to Riemannian Mean	最小离黎曼平均值
ME	Motor Execution	运动执行
MEG	Magnetoencephalography	脑磁波仪
MI	Motor Imagery	想象运动
NIRS	Near-Infrared Spectroscopy	近红外线分光光度法
PET	Positron Emission Tomography	正子断层造影

SDCI	Standard Deviation Confidence Interval	置信区间的标准偏差
SLST	Source Localized Spatio-Temporal	脑源定位时空特征
SNR	Signal-to-Noise Ratio	信噪比
SOI	Signal Of Interest	感兴趣信号
SSVEP	Steady-State Visual Evoked Potentials	稳态视觉诱发电位
SVM	Support Vector Machine	支持向量机
TN	True Negative	真阴性
TP	True Positive	真阳性
TS	Tangent Space	切线空间

* Original acronyms are noted in bold

Chapter 1 Introduction

1.1 Motivation

The primary objective of this doctoral thesis is to contribute to the advancement of Brain-Computer Interfaces (BCI) by developing a comprehensive pipeline allowing for the real-time, efficient, and non-invasive utilization of brain signals for monitoring basic user intentions. To achieve this goal, the focus of this work is dedicated to employing electroencephalography (EEG) techniques to rigorously analyze the Motor Imagery (MI) paradigm, specifically in relation to movements of the right and left hands.

The multifaceted approach undertaken encompasses the development of novel algorithms, the integration of advanced signal processing techniques, and the evaluation of system performance through rigorous experimentation and analysis. This thesis will start by presenting a comprehensive overview of the pathway that has been followed to address the various existing limitations. The subsequent chapters will delve into the specific details of the original contributions made, providing a thorough examination and analysis of each advancement.

1.2 Background

Various methods have been implemented to remove artifacts from EEG-based BCI systems and identify recurrent brain activity. The innovative algorithms presented in this work will be written in bold to easily differentiate from the already existing ones. Two new systematic reviews are carried out in this section to provide a global overview of the large diversity of algorithms available for pre-processing and processing.

1.2.1 The strange case of Brain-Computer Interface (BCI) and electroencephalography (EEG)

The primary objective of BCI technology is to use brain signals to establish a real-time interaction between a person and the outside world. Users can perform tasks that enhance or replace muscle activation for various functions such as movement, speech, writing, or expressing reactions (either in a conscious or vegetative state)[1]. Current research tries to build upon previous work dedicated to motor or cognitive understanding only with more

recent integration of different aspects of emotional and mind states (e.g., neurofeedback). The validation of a comprehensive procedure or pipeline requires several essential steps.

The signals to be analyzed must possess significance on two levels. Firstly, they should accurately represent the phenomenon of interest, and secondly, they should exhibit statistical properties that allow for generalization with a certain level of confidence. The brain signals can manifest as electric, magnetic, or biochemical signals, with the choice of recording modality depending on factors such as the required response time, spatial and temporal resolution, practicality, and patient comfort. Practicality can encompass physical or behavioral considerations as well as cost-effectiveness.
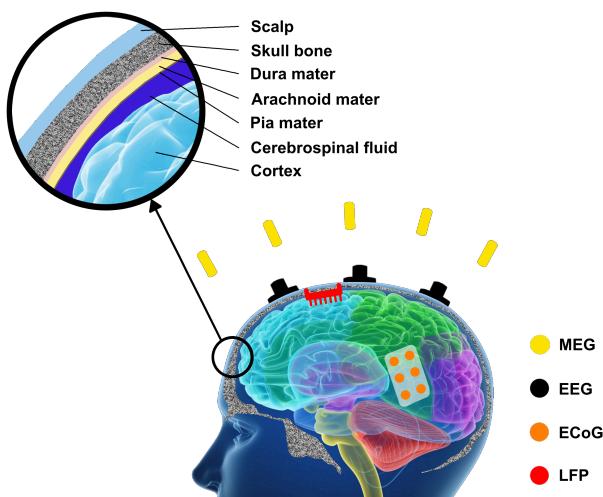
Figure 1–1 Invasive (ECoG, LFP) and non-invasive (EEG, EMG) imaging methods recording the brain electrophysiological activity

Among the available recording techniques (described both in Figure 1–1 and Table 1–1), EEG devices offer a wearable, relatively comfortable, and cost-effective solution with high temporal resolution. These characteristics make EEG devices a favorable choice for capturing brain signals in BCI applications[2].

The current state-of-the-art approach involves the use of Electrocorticography (ECoG) to capture the relevant signals[3]. This is due to the limitations of EEG in terms of signal-to-noise ratio (SNR) and spatial resolution, as well as the general impracticality of Magnetoencephalography (MEG) or functional magnetic resonance imaging (fMRI). However, the invasiveness of ECoG poses a significant challenge. Consequently, our research endeavors have focused on developing a purely non-invasive EEG-based solution.

Table 1–1 Recording methods' characteristics; * depends on the number of electrodes; ** is only for restricted regions depending on the electrodes' placement

Recording method	Measure	Risk	Electrode size (diameter)	Spatial reach	Spatial resolution	Typical values	Sampling rate (temporal resolution)	Freq. of interest
LFP	Electrical (Direct)	Invasive	5–120 μm (optimal > 16 μm x 16 μm)	0.1–1 mm	10^{-3}–1 mm^3	Potential amplitude 0.1–1.5mV	20 kHz	1–300 Hz
EcoG	Electrical (Direct)	Invasive	150 μm – 1 cm (optimal 2–3 mm)	0.5–5 mm	1–20 mm^3	Potential amplitude 50–100μV	1200 Hz	0–200 Hz
EEG	Electrical (Direct)	Non Invasive	2–10 mm (commercial Quick-Cap 7 mm)	15–60 mm	6–37 cm^3 *	Potential amplitude 1–10μV	100 – 5000 Hz (typical 256–512 Hz)	0.1–80 Hz (mostly 0.1–30 Hz)
MEG	Magnetic (Direct)	Non Invasive	None	15–60 mm (below the scalp)	10^3–10^4 mm^3	Magnetic fields 50–500fT	1–60 kHz (typical 1–2 kHz)	0.1–90 Hz (mostly 0.1–30 Hz)
fMRI	Hemo-dynamic (Indirect)	Non Invasive	None	Whole brain	0.5 – 200 mm^3 (typic. 2–3 mm^3)	Reg. cereb. blood vol. -20 – 20 %	0.3–20 Hz (typical 0.5–1 Hz)	< 0.5 Hz (mostly < 0.1Hz)
NIRS	Hemo-dynamic (Indirect)	Non Invasive	1.5–5 cm (emitter-detector pair) (optimal 3–3.5cm for adults, 2–2.5cm for infants)	7–24 mm	1–3 cm^3 * **	Conc. chg. (de)oxy-Hb -0.1–0.1 $mmol.mm$	0.5–125 Hz (typical ~10 Hz)	< 3 Hz (mostly 0.01–0.5 Hz)
PET	Hemo-dynamic (Indirect)	Non Invasive (injection of radio-active tracers)	None	Whole brain	1–10 mm^3	Standard. Uptake Val. (changes in reg. cereb. blood flow) 0–10 SUV	0.001–0.02 Hz	< 0.02 Hz

3

1.2.2   The Secret Signal of Interest

In order to maintain a rigorous approach, it is essential to thoroughly evaluate the necessity and benefits of each step involved in the research process. This evaluation involves a comprehensive assessment of various aspects, including the accurate representation of the targeted phenomenon.

The initial requirement is to define the chosen phenomenon and the reasons behind its selection, which can arise from a multitude of motivations. These motives may include addressing problems that lack existing solutions, improving existing procedures, developing new procedures that are simpler, more efficient, or more cost- and time-effective, or utilizing known phenomena to establish and quantify more universally applicable procedures. This list is, of course, non-exhaustive.

To illustrate one of the key choices we have explored, consider the scenario where a user intends to perform an action using either their right or left hand. This intention is commonly studied using the Motor Imagery (MI) paradigm. Detecting this intention has the potential to enhance the activation of weakened muscles (e.g., muscular dystrophy) or enable the control of a prosthetic apparatus. Numerous research laboratories worldwide are actively working towards developing real-time solutions for this application[4].

Our objective is to achieve a classification accuracy higher than 90% for user intent, without the need for neurofeedback or additional indirect information such as postural information. We anticipate that the remaining incertitude can be addressed through complementary techniques. Our approach involves an original application that reveals dynamic cortical patterns. This innovative utilization provides insights into the dynamic interactions and connections within the brain and subsequently enables feature extraction.

1.2.3   The Tell-Tale Source Localization

EEG signals represent the electrical activity of the brain but are not easily interpretable. The signals recorded from the scalp are a mixture of the activity from numerous neurons, and this activity is further distorted by the skull and scalp. Utilizing these signals directly can lead to inaccurate or ambiguous results due to the difficulty of pinpointing the exact origin within the brain. This is commonly referred to as the "inverse problem", depicted in Figure 1–2[5].

Solving the inverse problem involves addressing an underdetermined linear system that computes the cortical sources from the signals recorded by the EEG channels. In our case, this calculation is based on a general head model that transforms the 62 theoretically functioning EEG inputs into several thousands of cortical sources, also known as voxels. This approach provides more precise information about the active regions of the brain. Source localization techniques can significantly enhance the performance of BCIs by offering more accurate insights into brain activity.

To ensure accurate and reliable results, it is crucial to ensure the electrodes function homogeneously and that the signals to analyze are clean. Signal cleaning entails two primary objectives: eliminate statistically significant (and consequential) artifacts, as well as address any malfunctioning electrodes. By

achieving these goals, we can enhance the reliability and quality of the signals used in the analysis.
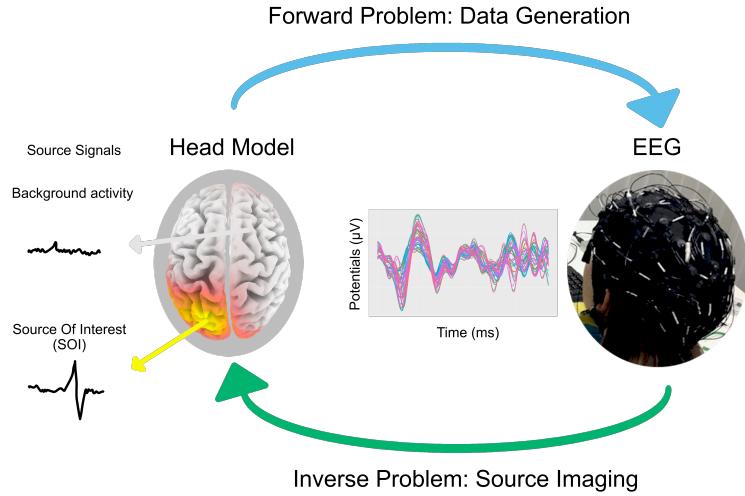
Figure 1–2 Illustration of the forward and inverse problems for EEG

1.2.4   Interview with the Blink

A major and repetitive artifact commonly observed in EEG signals is caused by spontaneous blinks (simply called blinks hereafter), illustrated in Figure 1–3. Three key metrics are considered to assess their impact: the frequency, the duration of the period during which the ratio of amplitude between the signal of interest and the artifact is significant, and the potential interference caused by them. Blinks occur at an average frequency of 20 times per minute, with a duration of approximately 0.6 seconds. The potential interference from blinks persists for more than 0.3 seconds. In a one-minute epoch, this implies that 6 seconds are affected by potential interference between the signal of interest and blinks, representing a 10% probability. While these evaluations provide rough approximations, they emphasize the need to address blink artifacts through appropriate correction methods.

Blind Source Separation (BSS) methods, such as Independent Component Analysis (ICA) or Artifact Subspace Reconstruction (ASR), are commonly employed for artifact correction. However, when tested within our pipeline, neither ICA nor ASR yielded satisfactory results. This observation has been reported in previous studies, yet a clear theoretical explanation is still lacking[6].

These limitations underscore the need for a deeper understanding of the origin, shape, distribution, transmission, and variability of blinks. This led us to design a multi-modal dataset targeted at comprehensively recording every aspect of the blinks. This multi-modal approach allowed us to construct an accurate model of the blinks, which played a crucial role in successfully addressing and mitigating these artifacts within our pipeline. Using this dataset, we were also able to implement robust correction techniques and enhance the overall quality and reliability of the EEG signals used in our analysis.
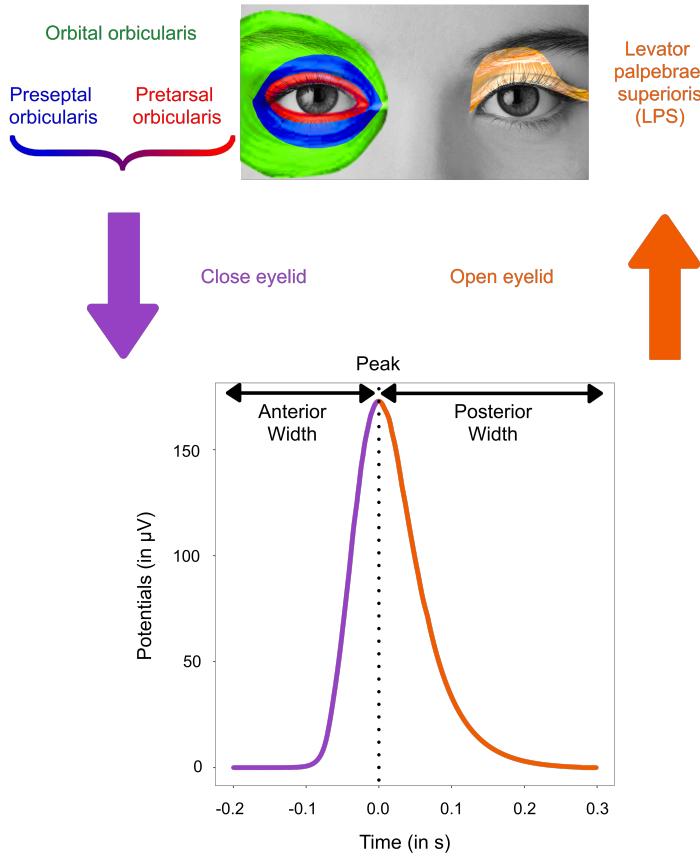
Figure 1–3 EEG recording (on a frontopolar channel) of a blink and its anatomic relationship

1.2.5 One Flew Over a Multi-Modal Dataset

Motor Imagery (MI) serves as the central paradigm of investigation in our research, given its practical relevance. However, in order to acquire a more comprehensive understanding of the underlying neural mechanisms occurring during MI, additional paradigms are explored, including motor execution (ME), P300, and steady-state visual evoked potentials (SSVEP). The inclusion of ME is particularly valuable due to the resemblance of its signal to MI. Additionally, the examination of P300 provides insights into the high temporal correlation between the stimulus and the brain response, while the exploration of SSVEP offers valuable information regarding the high frequency correlation in this context. These additional paradigms will serve as complementary approaches to validate and verify the effectiveness, reliability, and robustness of the proposed methodologies developed for the main paradigm. From a statistical standpoint, a priori sample size estimation is a crucial step as it ensures adequate statistical power to detect meaningful effects or differences in the study[7]. This guarantees that sufficient data is collected to truly comprehend the intricacies of the phenomenon and draw reliable conclusions. The minimum sample size required to detect the effect size of interest with a desired level of statistical power can be estimated by taking into

account factors such as the expected effect size, the desired level of significance, and the variability within the data.

Despite the importance of a priori sample size estimation, several studies and reviews have pointed out the complete absence of reported sample size calculations in neuroscience in general[8] and EEG research in particular (0% of the selected studies in[9]). This omission poses a substantial problem as it compromises methodological rigor, research reproducibility, and the ability to conduct meta-analyses. The absence of transparent reporting regarding sample size estimation hinders the evaluation of statistical power and the reliability of findings. It may also lead to underpowered studies that result in false-negative results or the inability to detect meaningful effects.

From a hardware perspective, the combination of EOG, EEG, EMG, eye tracking data, and high-speed videos during blinks has yielded valuable insights that are often unexpressed or even refuted. A notable result is that blinks are muscular artifacts, as evidenced by their synchronized replication in the frontopolar electrodes FP1 and FP2, which perfectly align with the EOG signals. Additionally, each individual exhibits a unique attenuation factor that can be modeled based on the distance between the electrodes.

These findings have enabled the development of an innovative correction method called the Adaptive Blink Correction and De-drifting (**ABCD**) algorithm. This approach not only corrects blink artifacts but also provides a means to identify and address problematic channels. By recalibrating or removing these problematic channels, the overall data quality is improved.

Furthermore, these findings offer empirical support for the theoretical understanding of the limitations associated with ICA and similar methods (such as ASR). The insights derived from the analysis of multi-modal dataset shed light on the specific reasons why ICA or ASR may struggle to adequately correct blink artifacts.

1.2.6   The Chronicles of Signals Cleaning

Over the years, various algorithms have been created or modified to remove unavoidable artifacts. To identify the research hotspots, a comprehensive review listing the major research topics was carried out using a bibliometric analysis. Several multidisciplinary databases are available online, such as Google Scholar, JSTOR, Scopus, or Web of Science (WoS). The latter, currently maintained by Clarivate Analytics, provides bibliometric data on published scientific articles starting from 1945.

We queried WoS in May 2022 with the following group of keywords: "(EEG OR electroencephalogra*) AND (artifact* OR artefact*)". This search returned 5136 papers, from which any article not related strictly to human EEG and biological artifacts was removed. Articles focusing on diseases or conducted in intensive care facilities were also discarded on the account that the artifacts would be specific to the medical environment. Finally, papers that did not explicitly mention how the artifacts were corrected in the keywords or in the abstract were also eliminated. The remaining 1465 'artifact' papers were further searched for explicit mention of blinks or ocular artifacts, resulting in 714 'blink' papers. Both groups

were arbitrarily categorized into five techniques: (1) geometrical features, (2) filtering, (3) source decomposition, (4) blind source separation (BSS), and (5) hybrid with machine learning. Each group is based on the core concept that underlies the algorithm. Specifically, hybrid approaches combine two (or more) artifact correction techniques or adapt machine learning methods to enhance the pre-processing step. Figure 1–4 illustrates these categories for both groups.
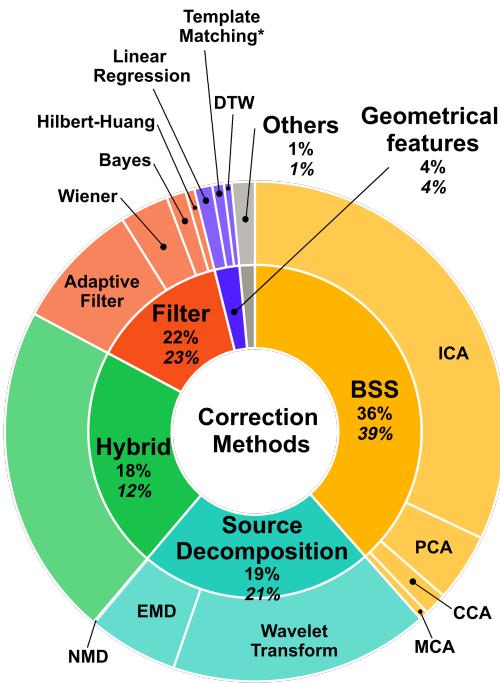
Figure 1–4 Common artifact preprocessing and blink correction (in italic) methods with the corresponding percentage of articles published per category; the category corresponding to the blink correction algorithm developed in this thesis is marked by an asterisk

This thesis Adaptive Blink Correction and De-drifting (**ABCD**) algorithm integrates both stationary and non-stationary characteristics and thus effectively addresses the challenges posed by blink artifacts in EEG signals. The stationary component accounts for consistent patterns and statistical properties of blink artifacts, while the non-stationary component captures the variability and temporal dynamics associated with blinks.

The consistent patterns observed in blink artifacts might be attributed to the repetitive nature of blinks, analogous to the regular and intensive nature of cardiac activity. To effectively correct these patterns, our approach employs a template-based modeling strategy. A second consideration is that spontaneous blinks exhibit varying levels of muscular fiber recruitment, independent of external stimuli. To account for this variability, our approach continually adapts the selected template to the specific shape of each blink artifact, facilitating the construction of a robust model.

Consequently, our method offers a more precise and reliable correction of blink artifacts, even in the presence of fatigue-induced changes. In contrast, statistical methods that rely on averaging the entirety of a session's data may not adequately capture the evolving characteristics of blink artifacts. This adaptive template-based modeling approach exhibits efficient computational processing, rapid adaptability to individual subjects, and the potential for online training during the setup phase for new users.

To evaluate the robustness and applicability of the proposed **ABCD** method, we conducted a comprehensive assessment across several commonly employed paradigms in EEG research (detailed previously). The experimental results substantiated the superiority of the proposed method over the conventional ICA approach, particularly in terms of extracting accurate source localization features. Notably, the remnants of blink artifacts that persisted even after ICA correction were effectively mitigated using **ABCD**. This improvement was evident in the enhanced quality of the reconstructed EEG signals, which improved fidelity to the underlying neural activity, as could be assessed by the upgraded classification results. The source localization that ultimately allowed for classification is influenced by three types of feature extraction, namely temporal, spatial, and spectral information.

1.2.7 The Three Influences on Feature Extraction

The eLORETA algorithm has been selected as the default method for source localization in our analysis pipeline[10]. This decision is based on several reviews that have demonstrated its superiority over various existing techniques. Additionally, its compatibility with the MNE Python library further supports its suitability for our purposes, as it allows for seamless integration into our existing analysis framework.

When considering the consistency of EEG measures in the source space, several important factors are crucial to acknowledge. One such factor is the physiological state of participants during data acquisition, as it can influence the noise variance used during the computations of the source localization. The resting condition, often used as a reference for noise, poses challenges for source-level analysis and is likely a significant contributor to the observed inconsistencies in source localization and connectivity outcomes.

These inconsistencies have been observed to impact the physiological significance of source localization and can alter connectivity estimates across different methods and modalities. Consequently, there exists notable variability in both within-subject (same subject, different sessions) and between-subject analysis, further emphasizing the need to carefully consider and account for these sources of variability in EEG research. However, it is important to note that despite these positional variations, the relative spatio-temporal relationship between the identified sources remains consistent. In other words, while the specific locations of the sources may change, their relative patterns of activation and temporal dynamics remain stable. These considerations led to the original implementation of a spatio-temporal similarities pattern recognition algorithm that displayed competitive classification results.

The selection of electrodes for classification plays a crucial role in addition to temporal features. However, there is a lack of comprehensive reviews that provide optimized subsets of electrodes for the

different paradigms under investigation in this thesis. To bridge this gap, a meta-analysis was conducted to systematically extract and synthesize information from various imaging modalities. A comparative analysis was performed to assess the classification performance of the Core Channel Selection (**CCS**) subsets in relation to commonly employed electrodes and subject-specific electrode configurations. The results of this analysis demonstrated the superior performance of the **CCS** subsets in achieving accurate classification outcomes.

Another factor that greatly impacts feature extraction is the spectral information. Neuronal synchronization refers to the coordinated activity of neurons, resulting in rhythmic oscillations at specific frequencies that can be detected by EEG. Brain rhythms, which are based on their frequency range such as delta (0.5 –4 Hz), theta (4 –7.5 Hz), alpha (7.5 –12.5 Hz), beta (12.5 –30 Hz), and gamma (30 –50 Hz) oscillations, exert a significant influence on brain function. These rhythms are believed to arise from various underlying dynamical mechanisms that are not yet fully understood.

Notably, such synchronization often occurs during task performance or specific states of consciousness. In the context of Motor Imagery (MI), beta band activity is frequently observed in the motor cortex. Consequently, filtering the EEG signals to focus on the beta band can facilitate the identification of MI, thereby enhancing its use as a BCI control. The main classical types of band-pass filters are Butterworth, Chebyshev, and Elliptic filters[11].

The tweaking of all three aspects of feature extraction, namely temporal, spatial, and spectral components, significantly impact the selection of EEG signal, which is then mapped to a corresponding covariance matrix. This mapping principle relies on the assumption that the power and spatial distribution of EEG sources exhibit a certain degree of stability during a specific mental state. It further posits that this valuable information can be effectively encoded by utilizing covariance matrices, which provide a discrete multilinear approximation of the intricate "landscape" of scalp signals.

1.2.8 A Song of Feature Extraction and Classification

Feature extraction techniques transform the pre-processed signal into a low-dimensional set of features that are assumed to characterize the underlying brain signal(s) of interest. Methods may use time, frequency, time-frequency, or spatial domains to select the proper characteristics (see[12],[13] for a review). Hereafter, we arbitrarily differentiate these techniques into seven categories, namely: (1) statistical, (2) Hjorth and autoregressive, (3) spectral, (4) dimensionality reduction, (5) spatial, (6) wavelet, and (7) chaotic parameters.

Classification algorithms then focus on discriminating these features between two or more distinctive classes. Several characteristics can describe these methods, such as supervised vs. unsupervised learning, the parametric vs. nonparametric model, or linearity vs. nonlinearity. Classification algorithms may be divided into six main categories based on the use of (1) hyperplane, (2) Bayesian probability, (3) clustering, (4) ensemble, (5) manifold, and (6) deep learning.

10
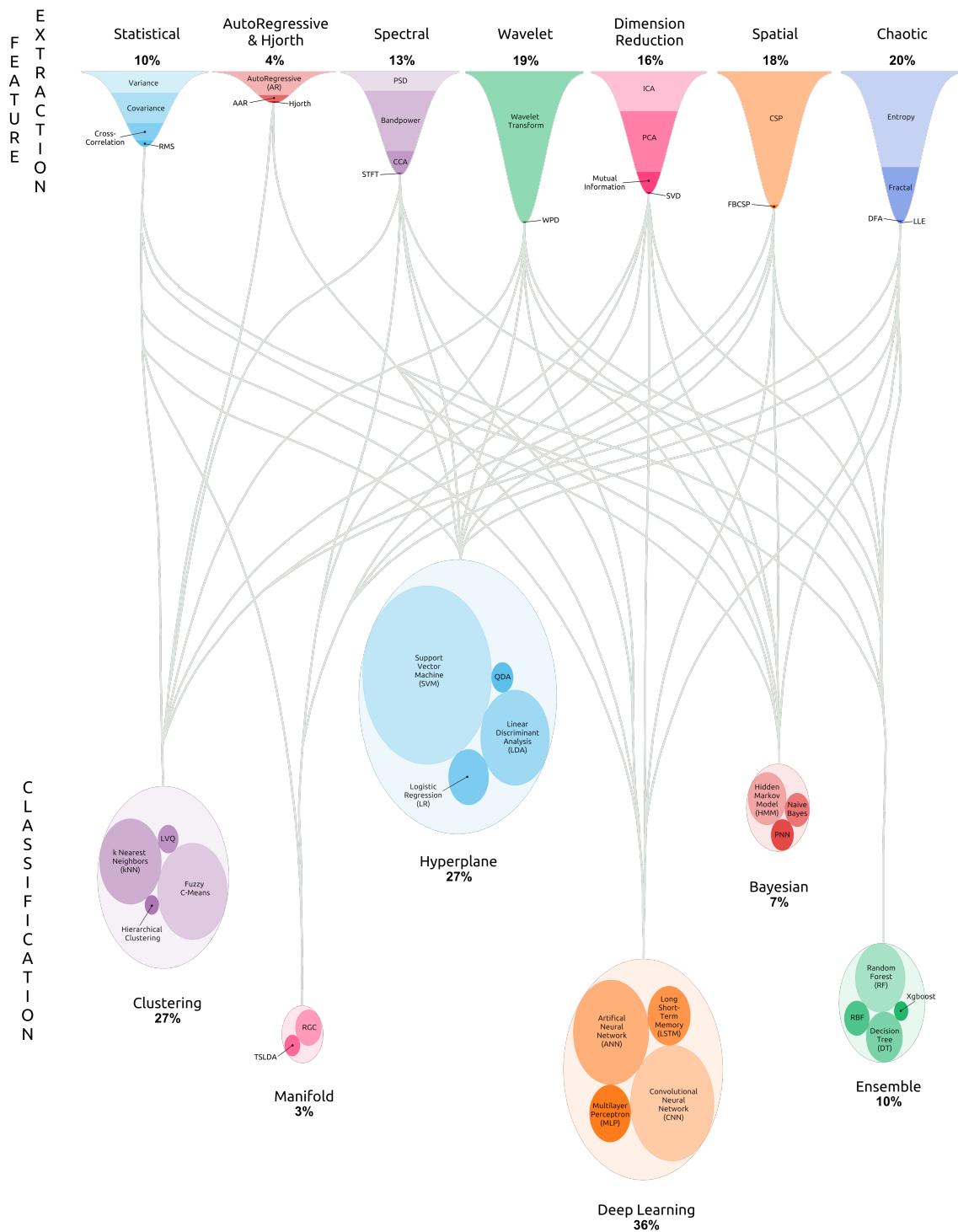
Figure 1–5 Main feature extraction and classification algorithms with the corresponding percentage of articles published per category for each step

With the same methodology as described previously, we queried WoS in May 2022 with the following group of keywords: "(EEG OR electroencephalogra*) AND (classification OR pattern recognition OR feature)". This search returned 32590 articles from which any paper not related strictly to human EEG, feature selection, or classification was removed. Additional exclusion criteria were similar to the ones used previously, namely, disease-focused (e.g., epilepsy) or hospital facility-oriented (e.g., intensive care units).

Our goal is here to give a quick overview of the available feature selection and classification methods, but is by no means a systematic review. That is why only the abstract and keywords have been screened. Hence, if neither precisely mentioned the selected algorithm(s), the corresponding paper was rejected. This drastic exclusion criterion resulted in 4076 articles for the feature selection step and 4368 papers for the classification part. Figure 1–5 outlines the different categories corresponding to these two processing stages, along with their main sub-categories.

1.2.9 Much Ado About Confusion Matrix

Confusion matrices are commonly employed to assess the performance of classifiers. They provide a comprehensive representation of the classification results by organizing them into a matrix format. Typically, a confusion matrix consists of four cells representing the counts or probabilities of True Positive (TP), False Negative (FN), True Negative (TN), and False Positive (FP) instances (see Figure 1–6).

Figure 1–6 Confusion matrix for a simple binary classification

They consist of two independent samples, each following a Binomial distribution. These samples can be denoted as follows:

- a positive sample (P) comprising True Positive (TP) and False Negative (FN) instances, with the corresponding probability of TP/(TP+FN);
- a negative sample (N) consisting of True Negative (TN) and False Positive (FP) instances, with the corresponding probability of TN/(TN+FP).

However, confusion matrices have certain limitations that need to be addressed. Firstly, they provide nominal values, which lack scientific validity and do not provide a quantitative measure of performance. Secondly, interpreting the results of confusion matrices can be challenging, particularly when dealing with small sample sizes, extreme probabilities near 0 or 1, and imbalanced datasets.

To overcome these limitations, efforts have been made to extend the utility of confusion matrices. One novel approach involves estimating confidence intervals for the probabilities associated with each sample (positive and negative). This allows for a more robust assessment of the uncertainty associated with the classification results.

Additionally, new visual representations have been developed to enhance the interpretation of confusion matrices. One such representation is the confusion ellipse, which provides a synthetic visualization of the classification performance. This ellipse takes into account the uncertainty in the classification results and provides a more comprehensive understanding of the classifier's performance.

From these generalized confidence intervals, an innovative methodology to ascertain the minimum sample size necessary to assess the statistical difference between two classifiers has been successfully formulated, along with a novel normalized quantification approach to measure the extent of their difference. This rigorous framework facilitates a more robust and standardized assessment of the comparative performance exhibited by the classifiers.

Overall, these advancements in the analysis and visualization of confusion matrices contribute to a more comprehensive and reliable assessment of classifier performance. They provide researchers and practitioners with valuable tools for evaluating and comparing classification results in various domains, not limited to BCIs.

1.2.10 Existing Limitations and Thesis Contents

BCIs hold immense potential for transformative advancements across diverse domains, including healthcare, assistive technology, and human-computer interaction. Of particular interest are non-invasive BCIs, which offer the convenience of real-time implementation and accessibility to a wide range of users. However, a common limitation of EEG-based BCIs is the low quality of the signal-to-noise ratio (SNR) and the localized nature of the features of interest. Therefore, the utilization of algorithms that enhance signal quality (e.g., effective artifact removal) and enable global information processing across the scalp (i.e., source localization) is of particular importance. In particular, EEG-based BCIs encounter several significant shortcomings that hinder their widespread adoption and curtail their effectiveness.

Shortcoming 1: EEG signals are contaminated with various artifacts, which can be particularly problematic for real-time applications. Among them, the spontaneous blink represents a significant physiological disturbance, with an average 10% probability that it occurs simultaneously with the neural signal of interest. Despite their prevalence, the characteristics of blinks are often overlooked and rarely recorded in a multi-modal fashion that would enable a comprehensive understanding of their nature as muscular

artifacts.

Shortcoming 2: A critical deficiency to validate EEG research results pertains to the absence of reported sample size calculations, which are essential for ensuring adequate statistical power to detect meaningful effects or differences in a study. In other areas of neuroscience-related research, it is common practice to rely on approximations derived from the Normal distribution for sample size calculations. However, this approach may lead to substantial discrepancies when compared to results obtained using a more appropriate distribution that is specifically fitted to the data at hand.

Shortcoming 3: A problematic limitation in artifact correction methods within EEG research is the inadequate performance of commonly used algorithms when it comes to correcting blink artifacts. ICA, while intended to separate the underlying cortical components from artifacts, often inadvertently removes genuine cortical aspects of the signal while retaining remnants of the blink artifact. The variability observed across different reconstructions of the same raw EEG data arises due to the inherent stochastic nature of ICA, resulting in different outcomes across multiple iterations or runs. This, combined with the continuous correction of the original signal, even in the absence of detected blinks, introduces unwanted variability and can potentially distort the underlying neural activity. Such inaccuracies may lead to erroneous interpretations and conclusions that can be especially consequential for real-time applications.

Shortcoming 4: A significant limitation in EEG measures in the source space (i.e., after source localization) lies in the considerable variability observed intra-subject (same subject, different sessions) and inter-subject. This variability poses challenges for efficient pattern recognition in the temporal domain, as it hampers the ability to consistently identify and extract meaningful patterns of brain activity over time. The spatial domain is also affected by this variability, as no optimized subsets of electrodes have been reported to our knowledge. Finally, the impact of various filtering approaches on the consistency and reliability of source localization results remains largely unexplored. These three knowledge gaps hinder the establishment of standardized and optimized feature extraction methods that can enhance the stability and robustness of source localization outcomes.

Shortcoming 5: Confusion matrices are a commonly employed method for evaluating classifiers' performance, yet they only report nominal values, which lack scientific validity and a quantitative measure of performance. Interpreting them can be challenging, especially when dealing with small sample sizes, extreme probabilities near 0 or 1, and imbalanced datasets. Additionally, there is currently no established method for determining the minimum sample size required to evaluate the statistical difference between two classifiers.

Chapter 2   A Large Multi-Modal Dataset of Four BCI Paradigms using EEG, Eye-Tracker, and High-Speed Camera

This chapter is dedicated to providing reliable data that facilitates the understanding of spontaneous blinks and the development of a realistic blink model for correcting EEG signals. The resulting dataset should effectively capture this major physiological disturbance caused by spontaneous blinks, which can impede the user's intent (e.g., during movement imagination). Given the uncertainty regarding the possible differential impact of blinks on EEG signals depending on the BCI paradigm, several tasks are performed during this experiment.

This large multi-modal dataset is the only online-available one (to the best of our knowledge) that simultaneously records electrophysiology and videos, along with synchronized eye-tracking. To estimate the required sample size, an innovative Fitted Distribution Monte Carlo (**FDMC**) simulation is conducted during a prospective power analysis. The **FDMC** analysis reveals that a minimum of 63 sessions is necessary, which is seven times fewer than what is suggested by power tables and 35% less than the estimation provided by G*Power 3, a widely used software that relies on the Normal distribution for a priori sample size estimation.

The online-available dataset encompasses a cohort of 31 subjects (both left and right-handed individuals) for a total of 63 sessions, resulting in 2520 examples of MI, ME, and SSVEP, and 5670 examples of P300. The quality of the data is validated through segmented signal-to-noise ratio (SNR) and source localization plots. This dataset will serve as the foundation for the development of innovative pre-processing and processing algorithms in this thesis. It has been made publicly available with the intention of facilitating analogous eye-related movement analysis or pre-processing and processing algorithms evaluation.

2.1   Introduction

Publicly accessible BCI datasets play a crucial role in promoting transparency and reproducibility in research. Their availability enables the evaluation and comparison of various approaches while also facilitating replication, verification, and enhancement of existing work. This fosters a collaborative environment, encourages the development of standardized evaluation protocols, and ultimately contributes to the advancement of BCI technology.

Statistical, practical, and technical aspects should be highlighted to offer insights into the dataset's potential strengths and limitations. This will enable other researchers to make well-informed decisions when using the dataset for their own studies. This notably includes details about the necessary number of subjects or sessions to guarantee the validity of hypothesis testing on this data.

2.1.1 Online BCI Datasets

The availability of online EEG BCI datasets has been steadily increasing, encompassing various applications such as motor imagery (often accompanied by motor execution), event-related potentials (such as P300), and steady-state visual evoked potentials (SSVEP). Additionally, datasets focusing on emotion BCI, error-related potentials, visually evoked potentials, slow-cortical potentials, resting state, as well as eye-blinks and movements can be accessed through https://openbci.com/community/publicly-available-eeg-datasets/ or https://physionet.org/about/database/. Another example is the BCI Competition IV dataset, which has become widely used in the EEG BCI data processing literature.

The availability of publicly accessible BCI datasets provides valuable resources for BCI researchers. However, it is worth noting that datasets dedicated solely to blink recordings capture EEG signals from a limited number of channels (typically ranging from 2 to 14 electrodes) without incorporating paradigms suitable for device control. Conversely, datasets recording motor imagery (MI), motor execution (ME), P300, or SSVEP lack simultaneous acquisition of other physiological data that could be utilized to validate blink detection or provide insights into its characteristics. Most datasets are also characterized by a relatively small size (number of subjects) and limited quantity (number of sessions/trials), which poses a risk of overfitting when developing or evaluation new methods. As a result, the scarcity and inadequacy of publicly accessible datasets may hinder the reliability and reproducibility of BCI research[14].

2.1.2 Sample Size Requirements

Analyzing the effect and sample sizes of past research on a particular topic should be the first step in any study aiming to replicate or expand upon earlier investigations. In the absence of a power analysis, this step is usually bypassed by adopting "non-rational bases" for determining sample sizes, as Cohen called them[15]. This encompasses adhering to past practice, making decisions based on data availability, relying on intuition or personal experience, and following statistical rules of thumb [16].

None of the datasets we encountered incorporated an explicit rationale for the determination of the number of subjects or sessions. Likewise, only the most recent articles make reference to certain effect size calculations[17-18], with one meta-analysis specifically examining MI in healthy participants[19]. However, only about half of the articles reviewed focused on the differences between left- and right-hand MI, and approximately 40% of the papers employed a range of 1 to 8 EEG channels. Yet, source localization, which will be used for pattern recognition in Chapter 4, is highly influenced by the number of sensors, and its accuracy diminishes significantly as the number of electrodes decreases[20]. That is why the effect size extracted from this meta-analysis could not directly be applied to our a priori sample size estimation.

2.2 The Great Sample Size

2.2.1 Goal

The user's intent is expressed by the specific task that the subject will perform during the experiment. For the sake of simplicity, MI will represent the task to be performed, although the approach detailed in the following sections applies equivalently to other paradigms.

The primary objective of this section is to estimate the magnitude and direction of effects that occur when cortical sources are activated through MI. In this context, the term "effect" refers to any results the study revealed linked to the imagined hand movement. The specific hand involved in the movement is irrelevant, as it cannot be assumed that findings for one hand can be generalized to the other. Certain brain sources may either be linked to the left hand, right hand, or both. Therefore, it is necessary to conduct independent investigations for each distinct imaginary movement (or, more generally, for each specific task).

2.2.2 Statistical Significance, Effect Size, and Statistical Power

To enhance our understanding of actual effects, regardless of the phenomenon under study, it is crucial to interpret the statistical significance and the practical significance of the results. Statistical significance reflects the improbability of the findings, while practical significance relates to their meaningfulness. It is unfortunate, but not uncommon, for a result to be statistically significant yet trivial. This means that, although a result may possess statistical significance when a sufficiently large sample size is employed, it may lack practical significance and therefore have no actionable implications. A statistically significant result indicates that the observed outcome is unlikely to be due to chance and is inferred from the *precision of the estimate*. However, a practically significant result holds meaning in the real world and is inferred from the *size of the effect*.

The statistical significance of any result is influenced by both the effect size and the sample size used to estimate it. Smaller samples are less likely to yield statistically significant results, regardless of the effect size. Another confounding problem is that estimates need to be tested in order to be detected. Statistical power thus refers to the *likelihood of detecting a genuine effect* in a study. Unfortunately, surveys consistently demonstrate that most studies, particularly in BCI, where effect sizes tend to be small, lack the necessary power to detect the desired effects. a literature review focusing on human Electroencephalography (EEG) and Event-Related Potential (ERP) showed that 0% of the selected studies reported sample size calculations[9].

A well-designed study typically incorporates a prospective analysis of statistical power. For the current investigation, previous research revealed that MI has an effect on the cortical sources. This means that, while we acknowledge the presence of a genuine effect to be detected, we cannot be certain whether the algorithms developed in this thesis are able to accurately identify it. It is reasonable to assume that

this effect is likely to be both small (as previous studies often attribute its undetection to BCI illiteracy) and significant (as efficient detection of MI would have numerous applications). Therefore, it is probable that this probably tiny effect would go unnoticed if the sample size was too small.

Cohen's d index is one of the most commonly used dimensionless effect size metrics, assessing the statistical level of distinction between two variables belonging to two groups with means M_1 and M_2 [21]. The greater the distance in terms of standard deviations (σ) units, the easier these variables can be separated:

$$Cohen's \ \ d = \frac{M_1 - M_2}{\sigma} = \frac{M_1 - M_2}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}} = \frac{M_1 - M_2}{\sqrt{\frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1+n_2-2}}} \tag{2-1}$$

Typically, studies are designed with an 80% probability of detecting a genuine effect (or a 20% probability of committing a Type II error, i.e., failing to detect the genuine effect). This design choice is based on the notion that Type I errors should be treated four times more seriously than Type II errors. If the alpha significance level is set at .05 (indicating a 5% chance of a Type I error), the beta level should be set at .20 (representing a 20% chance of a Type II error).

Statistical tests are assumed to be nondirectional (two-tailed), and the effect size (when not reported in previous studies) can be derived from a prospective power analysis, where the current study is initially conducted on a few subjects. The effect size, sample size (or number of observations), and Type I and Type II errors are interrelated, allowing the determination of any parameter based on the other three.

2.2.3 Prospective Power Analysis

Assuming that the algorithms presented later in this thesis possess the capability to accurately discern the Signal of Interest (SOI) associated with MI, and that this outcome can be represented as a univariate variable, the initial phase of the study involves conducting preliminary experiments on a limited number of subjects. The purpose of these preliminary experiments is to compute the a priori sample size based on the analysis of this specific variable. By analyzing the data obtained from these initial subjects, the necessary sample size for subsequent investigations can be determined in advance, ensuring adequate statistical power for the study.

All calculations presented hereafter are performed under the assumption of the worst-case scenario, i.e., for the hand yielding the smallest effect sizes. This conservative approach ensures that the analysis accounts for the most challenging conditions, thereby providing a robust and rigorous evaluation of any related phenomenon under investigation.

2.2.3.1 Usual A Priori Sample Size Computations

To detect an effect size of $r = .03$ (derived from the mean effect size obtained in the prospective power analysis) using a two-tailed test, the estimation of sample size based on approximate correlation power calculation indicates a minimum requirement of $N = 8,718$ observations given con-

ventional alpha and power levels for one (This result can be simply found using the following code in

R: `pwr::pwr.r.test(n = NULL, r = 0.03, sig.level = 0.05, `**`power`**` = 0.8, alternative = "two.sided")`).
This implies that if there are 20 imaginary hand movements per session, a minimum of sessions would be
necessary, which in turn necessitates the participation of subjects, each engaging in 3 sessions.

A more precise approach to calculating the required sample size for a specified power is to utilize
G*Power 3[22]. The computations are based on the common procedure that investigates a null hypothesis,
stating that two independent samples' population means are equal ($H_0 : m_A - m_B = 0$). The t test on
means can then be computed and compared to the critical value (corresponding to a specific area under
the t-distribution curve). By assessing the effect size from differences observed in the prospective analysis
(matched pairs), an effect size of $dz = 0.064$ is determined, leading to a minimum sample size of $N = 1929$
observations with standard alpha and power levels. This would require at least $N = 97$ sessions or $N = 33$
subjects participating in 3 sessions each.

It is important to note that G*Power 3 assumes normality in the distribution. If the data distribution
is skewed, the power indicated in the results may not hold for small sample sizes, as the calculations are
based on the normality assumption. An alternative approach is to employ statistical methods that can
accommodate non-normal and heteroscedastic data. Regarding our distribution, it is first necessary to
examine whether it is homogeneous across subjects and/or sessions (i.e., exhibiting a single population
mean) and whether it is normally distributed across subjects and/or sessions.

2.2.3.2   Homogeneity

Homogeneity refers to the assumption that the populations or groups being compared possess sim-
ilar distributions and parameters. Figure 2–1 depicts the values obtained from the prospective power
analysis conducted across multiple subjects, as well as for the three individual sessions of a specific sub-
ject. Following the identification of the signal of interest (with methods detailed in later chapters), the
relevant features of interest (FOI) are extracted. The median of these FOI (limited to those that our al-
gorithms successfully extracted) can then be computed, and each FOI's distance from the median FOI is
then calculated.

Table 2–1 presents the pooled effect sizes, both Cohen's d and r, pertaining to the comparison of the
variable of interest during imagined movement versus non-movement conditions across each session of a
specific subject.

A test of homogeneity across sessions is conducted on four subjects who participated in three sessions
each during the prospective analysis. The weighted mean effect size $\bar{r} = \frac{\sum n_i r_i}{\sum n_i}$ is computed along with the
variance of the sample of correlations $v_r = \frac{\sum n_i (r_i - \bar{r})^2}{\sum n_i}$ and the variability in all groups in the calculation of
the standard error $SE_{\bar{r}} = \sqrt{\frac{v_r}{k}}$ where k is the number of sessions.

The r score can be converted to a z-score with $z = \frac{|\bar{r}|}{SE_{\bar{r}}}$, which in this case equals 3.99. With a
predetermined significance level fixed at $\alpha = 0.005$, the null hypothesis would be rejected in a two-tailed
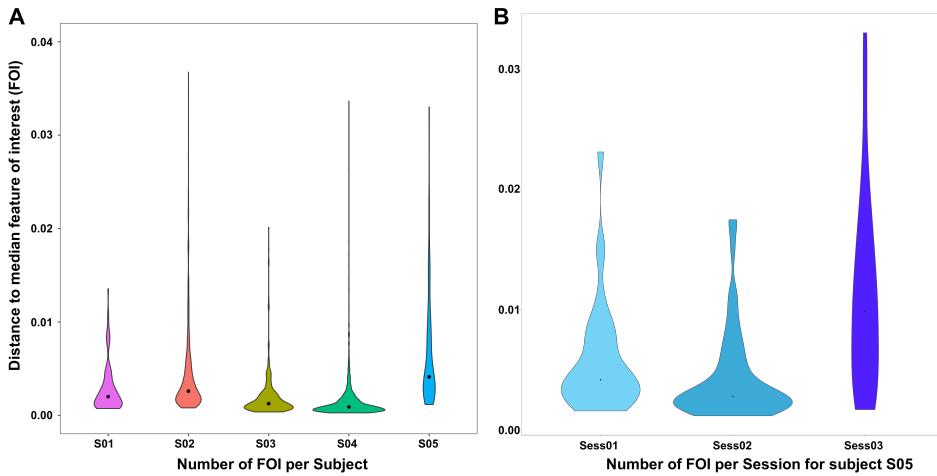
Figure 2–1 Violin plots showing the distance between each feature of interest (FOI) and the median FOI in (A) five subjects and (B) all three sessions for one subject (S05); the black dot represents the median

Table 2–1 Pooled effect sizes (Cohen's d and r) during imagined movement versus non-movement conditions for subject S05

Session	Number of detected task-related brain source activations	Pooled Cohen's d effect size with a 95% CI	Pooled r effect size with a 95% CI
Sess01	58	0.26 [-0.11, 0.63]	0.13 [-0.05, 0.3]
Sess02	55	0.39 [0.02, 0.77]	0.19 [0.01, 0.36]
Sess03	39	0.78 [-0.32, 1.24]	0.36 [0.16, 0.53]

test whenever the z-score exceeds $z_{\alpha/2} = 1.96$. Since the calculated z-score is greater than the critical value of z for the chosen standard of significance, the mean is statistically significant. Therefore, the mean correlation with a 95% confidence interval ($CI = \bar{r} \pm z_{\alpha/2} * SE_{\bar{r}}$) across sessions is $\bar{r} = 0.211[0.107, 0.314]$.

A small confidence interval indicates that the distribution of effect sizes is likely to be homogeneous. To test this hypothesis (i.e., that there is only a single population mean), the Q statistic can be calculated to quantify the degree of difference between the observed and expected effect sizes with $Q = \sum (n_i - 1)(r_i - \bar{r})^2$ which in this example is equal to 1.241. The critical values that intersect $\alpha = 0.05$ and three degrees of freedom is 7.815. As the computed Q statistic does not exceed this critical value, the homogeneity hypothesis cannot be rejected.

For every subject that underwent three experiments in the prospective analysis, the corresponding Q statistic does not exceed this critical value. This means that the population of effect sizes is homogeneous, or, in other words, similar effects are being observed across sessions (as one would expect).

Similar calculations across the five subjects yield a Q statistic equal to 11.019 which does not exceed

the critical value of 11.070 for five degrees of freedom. This implies that the homogeneity of the effect size population across subjects is present, albeit with a lower level of confidence compared to the homogeneity observed within sessions of an individual subject.

2.2.3.3 Fitting to Common Distribution

Having established the presence of homogeneity in the data, the focus shifts towards investigating the distribution, with a particular emphasis on assessing whether the data follows a normal distribution. Understanding the distribution of the data is crucial, as it facilitates informed decision-making regarding the selection of suitable statistical tests. If the data is normally distributed, the results obtained with G*Power 3 can be used. On the other hand, if the data deviates from normality, we need to devise alternative distribution-free methods to ensure accurate and reliable analyses.

Testing for all known distributions would be incredibly time-consuming. Instead, common distributions are arbitrarily chosen to test whether the CP distribution can be fitted to one of them. The univariateML library is used in R to test for nearly 30 common distributions[23].

Table 2–2 Fitted common distributions for the distance per subject

Subject	CVM	AD	KS	AIC	BIC	Choice
S01	No distribution	No distribution	No distribution	Skew Generalized Error	Pareto	Skew Generalized Error
S02	No distribution	No distribution	No distribution	Skew Student-t	Inverse Weibull	Skew Generalized Error
S03	No distribution	No distribution	No distribution	Skew Student-t	Inverse Weibull	Skew Generalized Error
S04	No distribution	No distribution	No distribution	Skew Student-t	Inverse Weibull	Skew Generalized Error
S05	No distribution	No distribution	No distribution	Skew Generalized Error	Skew Generalized Error	Skew Generalized Error

The Kolmogorov-Smirnov (KS), Cramer von Mises (CVM), and Anderson-Darling (AD) statistics are popular goodness-of-fit tests. The KS distance is defined as the largest absolute difference between the CP distribution CDF and the hypothesized underlying distribution CDF evaluated at any point. The CVM distance is obtained by taking the square root of the sum of the squared differences between the two

distributions. Finally, the AD test uses the fact that if the data arises from the hypothesized distribution, the CDF can be assumed to follow a uniform distribution.

To complete this analysis, Akaike information criterion (AIC) and Bayesian information criterion (BIC) are also computed. AIC estimates the relative amount of information lost by a given model, and thus deals with the trade-off between the goodness of fit and the simplicity of the model. BIC is based, in part, on the likelihood function, and is closely related to AIC. Both AIC and BIC try to avoid overfitting but do not provide a test of a model in the sense of testing a null hypothesis. They merely give intel on the relative quality to other models.

By definition, AIC and BIC always give a result, since they rank the tested statistical models. The preferred model is merely the one with the minimum value. On the other hand, CVM, AD and KS rely on testing a null hypothesis. For a candidate model, if the P-value is "small", the null hypothesis will be rejected, meaning that the test failed to show that the considered distribution follows the candidate distribution.

All these tests cannot find the "best" distribution. On the contrary, by considering all these statistics, one can infer which distribution is the least less likely. This is the reason why these different tests are computed along with visual inspection of the Q-Q and P-P plots to provide an informed choice on the fitted distribution. The results are summarized in Table 2–2.
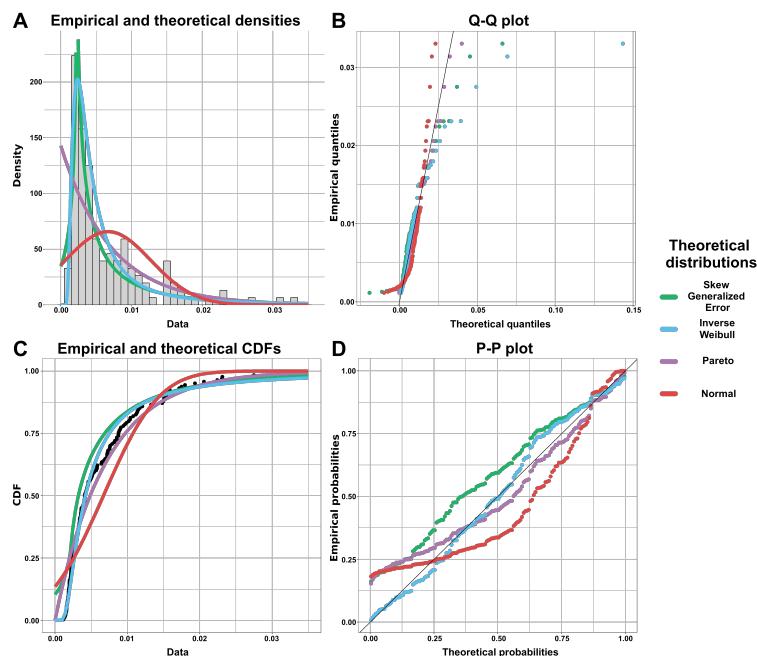
Figure 2–2 Visual comparison of four fitted distributions (three from Table 2–2 plus the normal one) for subject (S05) (A) empirical and theoretical densities, (B) Q-Q plot, (C) empirical and theoretical CDFs, and (D) P-P plot

The empirical and theoretical distributions, as well as the CDFs, Q-Q plot, and P-P plot, can be displayed using an adapted code from the fitdistrplus library (see Figure 2–2). Among the four represented distributions, the normal one gives the worst results for S03, while the Inverse Weibull distribution gives the worst fits for S01. The skew generalized error distribution, followed by the Pareto one, gives the most constant results across subjects.

The Skew Generalized Error Distribution (SGED) possesses the ability to effectively represent data exhibiting skewness and heavy tails. This characteristic is particularly of interest due to the asymmetrical nature of the empirical distance. The SGED offers greater flexibility, allowing for the inclusion of exceptional observations without exerting excessive influence on the estimation of distribution parameters[24]. Consequently, the SGED is chosen to represent the distance distribution, resulting in the least less likely distribution that will provide a better goodness of fit for the empirical data.

2.2.3.4 Fitted Distribution Monte Carlo (FDMC) Simulation for A Priori Sample Size Estimation

Neither the distance distribution per subject nor the overall distance distribution can be assumed to be Gaussian. Hence, it would be incorrect to directly use the Central Limit Theorem. Instead, Monte Carlo simulation (or parametric bootstrapping) can generate samples from the fitted to the empirical distance distribution (here the skew generalized error) and calculate the sample size required to reach a given power for a new EEG experiment.
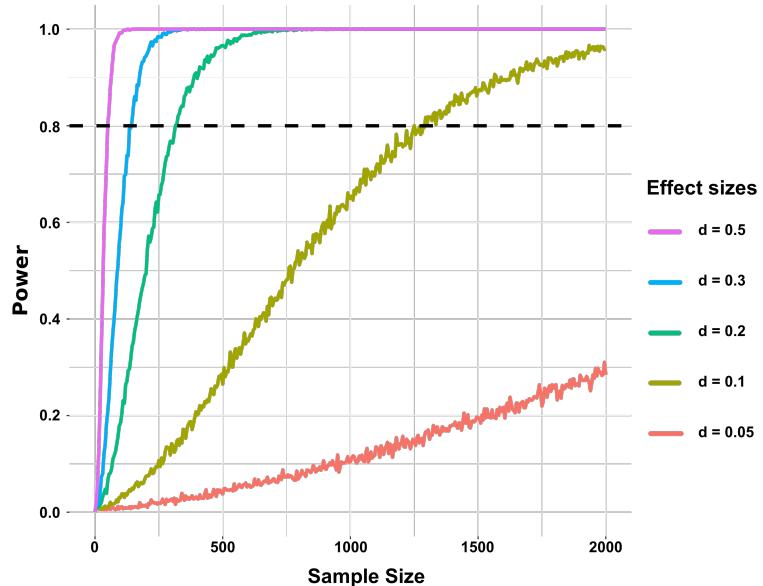
Figure 2–3 Monte Carlo simulation using the skew generalized error distribution (SGED) for five different Cohen's d effect sizes (prospective power analysis conservative effect size $d = 0.1$

Several effect sizes derived from the prospective power analysis are tested to compute the minimum sample size required to achieve an 80% power target (from very small to medium: $d = 0.05$, $d = 0.1$, $d = 0.2$, $d = 0.3$, and $d = 0.5$). As usual, the significance level is fixed at $\alpha = 0.05$, and the power at 0.8 ($\beta = 0.2$). The Monte Carlo power analysis using the SGED as the fitted distribution for these five different effect sizes is displayed in Figure 2–3.

The **FDMC** simulation highly depends on the effect size estimated from the prospective power analysis. Given that the smallest effect size was found to be around in this analysis, this conservative Cohen's d was deemed representative and employed for the a priori sample size estimation. In that case, the minimum number of observations should be equal to $N = 1,250$ observations, assuming standard alpha and power levels. This would necessitate a minimum of $N = 63$ sessions or $N = 21$ subjects, with each subject participating in 3 sessions.

2.3 The Story of Participants

The study protocol was approved by the local ethics committee of Shanghai Jiaotong University. All volunteers signed an informed consent form explaining the experiment that allows to publish their anonymized data. Thirty-one healthy individuals voluntarily participated in the study (11 women and 20 men, mean age 29 ± 7, range $20 - 57$). Among them, 14 participants completed a single session, 2 attended two sessions, and 15 achieved three sessions, for a total of 63 sessions.
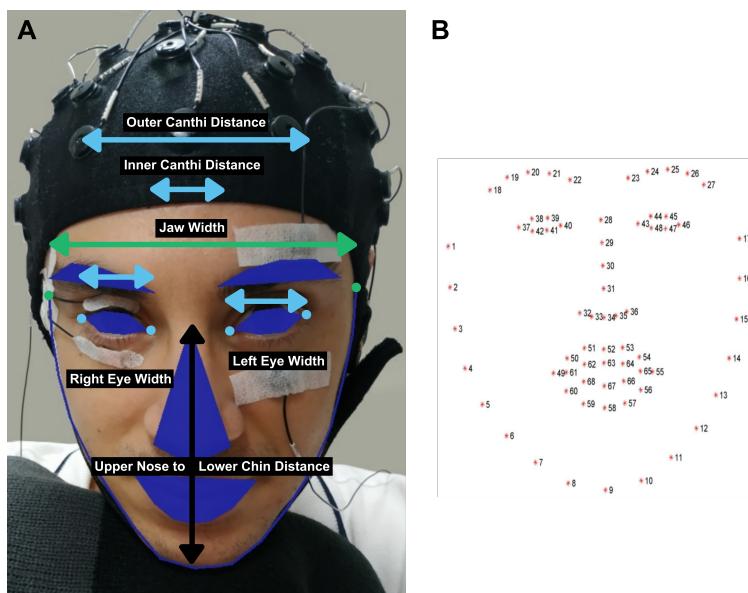
Figure 2–4 (A) Five subject-specific distances computed from (B) the 68 points iBUG 300-W dataset defined face regions. Informed consent was obtained from the individual in the figure for the publication of the images.

At the beginning of the first session, subjects were handed out a comprehensive questionnaire[25] completed with physical characteristics (such as height, skull diameter, eye correction) and the Edimburg Handedness Questionnaire[26]. All subjects were kept in the study, regardless of their handedness or familiarity with BCI. The participants are hereafter identified only by their aliases "S01" to "S31" and some of their characteristics are summarized in Table 2–3.

Table 2–3 Demographic description and a few physical characteristics of the participants; R. for right, L. for left, D. for decile

Subjects	Age	Gender	Height	Head circumference (cm)	Nasion - Inion (cm)	Eye correction Left (K dioptre)	Eye correction Right (K dioptre)	Laterality Index Handedness	Decile Edimburg Handedness	Augmented (15 item) index	Mother tongue	Familiarity with fast displays	Familiarity with BCI
S01	30	F	176	56	37	500	550	75	4th R.	76.67	FR	Very	Yes
S02	30	M	176	57	38	250	230	95	9th R.	90	FR	A little	Yes
S03	25	M	181	58	36	200	200	100	10th R.	100	ES	Very	No
S04	25	M	170	57.5	37	600	600	-30	1st L.	-26.67	ZH	A little	No
S05	23	F	163	56	35	850	900	80	5th R.	86.67	ZH	No	Yes
S06	25	F	168	57	33	0	0	25	Middle	33.33	ZH	Very	No
S07	20	M	165	58	36	200	150	60	2nd R.	66.67	ZH	Very	Yes
S08	38	F	160	53	31	500	500	90	7th R.	73.33	ZH	Very	No
S09	29	M	185	59	38	175	150	75	4th R.	76.67	FA	Very	Yes
S10	28	M	186	56	37	200	100	100	10th R.	100	FR	Very	No
S11	29	F	170	54	33	100	100	75	4th R.	76.67	ZH	A little	Yes
S12	24	F	165	55	35	75	50	85	6th R.	83.33	RU	Very	Yes
S13	23	M	179	57	37	425	375	-40	1st L.	-46.67	ZH	A little	Yes
S14	27	M	180	61	39	300	300	95	9th R.	96.67	ZH	A little	Yes
S15	24	F	170	53	34	750	750	90	7th R.	93.33	ZH	Very	No
S16	22	M	172	60	34	900	800	95	9th R.	90	ZH	Very	No
S17	24	M	186	61	39	400	400	-25	Middle	-26.67	ZH	A little	Yes
S18	26	M	182	57	37	175	175	80	5th R.	86.67	EN	Very	Yes
S19	32	M	180	57	36	0	0	55	1st R.	50	FR	Very	No
S20	27	F	159	56	33	350	350	100	10th D.	96.67	ZH	A little	No
S21	36	F	178	55	38	100	100	40	Middle	33.33	FR	A little	No
S22	35	M	184	57	34	0	0	75	4th R.	70	FR	A little	No
S23	23	M	182	57	37	550	550	100	10th R.	100	FR	A little	No
S24	30	M	178	59	35	800	800	100	10th R.	100	ZH	Very	Yes
S25	26	M	158	59	38	0	-200	80	5th R.	86.67	FR	Very	No
S26	33	M	182	57	38	0	0	100	10th R.	100	FR	Very	No
S27	25	M	175	57	33	300	200	35	Middle	23.33	ZH	Very	Yes
S28	32	F	160	54	34	NA	NA	100	10th R.	96.67	FR	No	No
S29	32	M	168	57	36	0	0	65	2nd R.	73.33	FR	Very	No
S30	57	F	170	56	34	NA	NA	50	1st R.	60	FR	A little	No
S31	28	M	186	57	36	NA	NA	80	5th R.	80	EN & FR	A little	No

Pictures were also taken for each subject. A facial landmark detector trained on the 68 points anno-

tated dataset from the 300 Faces In-The-Wild Challenge (iBUG 300-W)[27] was implemented to extract individual facial features. Figure 2–4 shows the resulting facial landmarks from which seven face regions are defined: jaw, mouth, nose, left and right eye, and eyebrow. Since these facial landmarks are calculated in pixels, distances in the metric system are needed to compare data from different photographs shot from the same distance and angle. Five subject-specific measures are then computed: Left and Right Eye Width, Inner and Outer Canthi Distance, and Upper Nose to Lower Chin Distance.

2.4   The Paradigms Code

The subjects performed all the four following BCI tasks: Motor Imagery (MI), Motor Execution (ME), Steady-state visual evoked potentials (SSVEP) and P300 visual evoked potentials. Constrained by hardware storage capacities and to reduce user fatigue, each BCI segment has been limited to a maximum of 14 min recording. The P300 speller task, lasting for 24 min, has been divided into two parts: one for spelling four-letter words (P3004L) and the other one for spelling five-letter words (P3005L). Every session is composed of these five paradigms, randomly ordered at the beginning of the session. A complete session lasts in average 45 minutes, not counting the breaks and time dedicated to the questionnaires. The P3005L is the only task consisting of 50 trials. The other tasks all contain 40 trials.

All BCI tasks have a similar structure design. Each paradigm starts with a message greeting the participants, while a 300 trigger code is sent and a Welcome cue is recorded. Trials vary depending on the BCI task. At the end of all trials, a message thanks the volunteers for their participation. This final step is recorded with a 300 trigger code and a Goodbye cue. Trials last for a fixed time linked to the paradigm. On the other hand, the Welcome and Goodbye recording times vary depending on when the experimenter launches and stops the Python code.

2.4.1   Paradigm #1: MI

Subjects are instructed to perform a grasping with kinesthetic motor imagination, using the left or right hand with all the fingers. The 40 trials are randomly distributed between 20 Left MI and 20 Right MI. At the start of each trial, a white fixation cross appears in the center of the screen for 2s. A red rectangle cue then appears randomly on the left or the right side of the cross for 4s. At this cue onset, subjects start to repeatedly imagine grasping the corresponding hand 3 times with a self-paced speed of approximately 1 Hz.

The trial ends with the fixation cross and the red rectangle cue disappearing during a random rest of 1 −1.5s, allowing for relaxation while avoiding the subject's adaptation. The trigger code sent indicates the number of the trial, whereas the cue is respectively recorded as "Fixation", "Left" or "Right", and "Break Random" (see Figure 2–5).
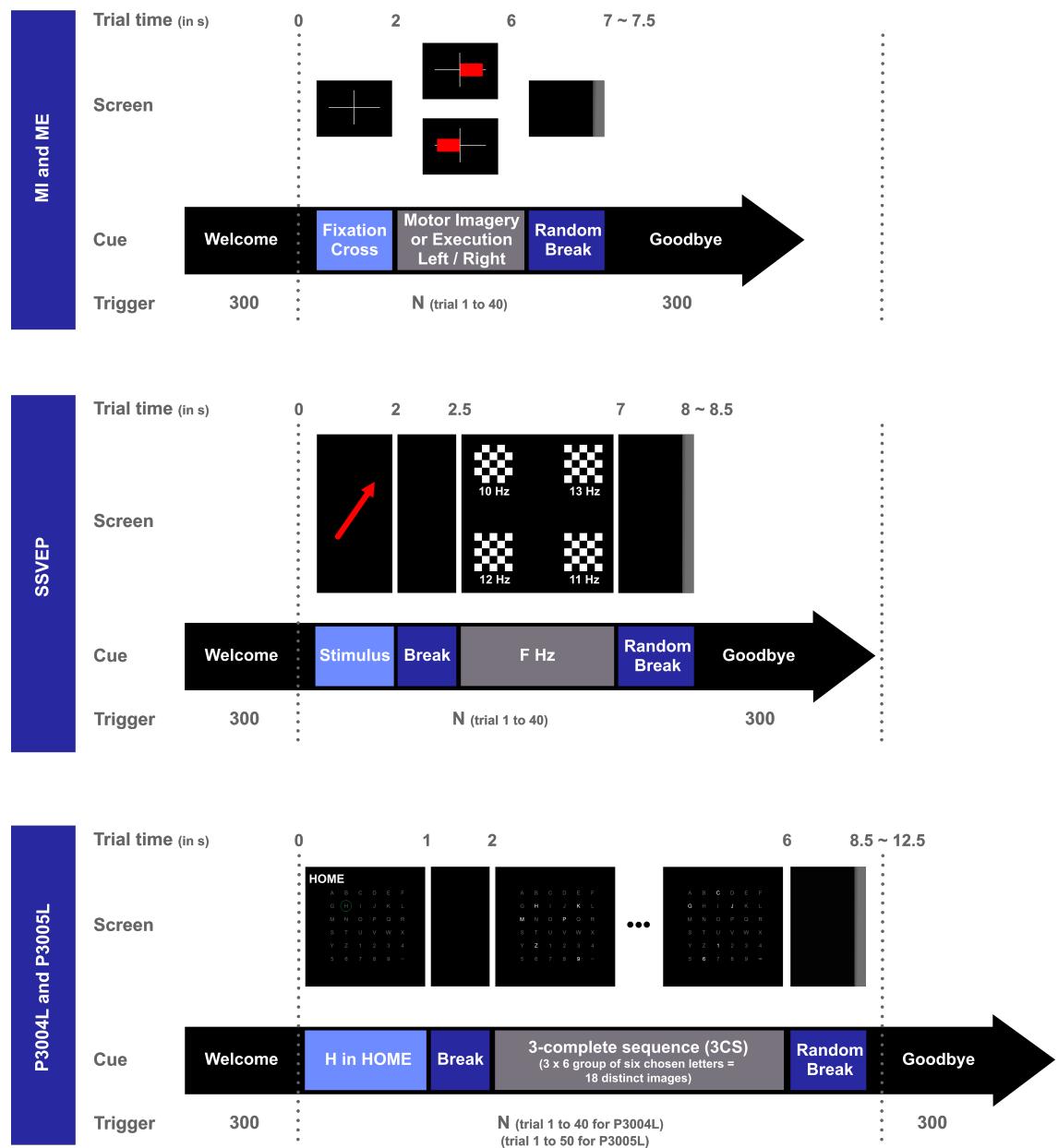
Figure 2–5 Chronology of the five paradigms: Motor Imagery and Motor Execution of a 1 Hz frequency hand grasping; multi-frequencies Steady-State Visual Evoked Potentials; P300 speller for four-letter and five-letter words

2.4.2 Paradigm #2: ME

The experimental paradigm for motor execution (ME) is the same as the one used for MI (Figure 2–5). The trigger codes and cues are also unchanged. The only difference is that the participants physically perform the hand grasping movements.

2.4.3 Paradigm #3: SSVEP

Stimuli consist of four checkerboards reversing black and white at a fixed frequency (10 Hz, 13 Hz, 12 Hz, and 11 Hz, respectively), each located in a monitor quadrant (upper-left, upper-right, lower-left, lower-right). The 40 trials are randomly ordered and equally distributed between each target frequency. At the beginning of each trial, a red arrow indicates during 2s where to gaze. This is followed by a brief black screen for 0.5s.

Participants then fixate the target stimulus during 4.5s. Finally, a black screen is displayed during a random rest of 1 – 1.5s. The trigger code corresponds to the number of the trial, whereas the cue is respectively recorded as "Stimulus", "Break", "F Hz" (where F is either 10, 11, 12 or 13), and "Break Random" (Figure 2–5).

2.4.4 Paradigm #4: P3004L (P300 for Four Letters Word)

2.4.4.1 Stimulation Sequence

In the classic visual oddball paradigm, the detection of the P3b is used to infer the targeted stimulus. In particular, the widely used Farwell and Donchin speller consists of a matrix whose cells are flashed alternately. A sequence is complete when all cells have been highlighted in the matrix. Each image generally highlights six symbols, and six images form a sequence. Three consecutive sequences usually constitute a trial. Originally, the six cells are grouped according to the row/column paradigm (RCP), where a whole row or a whole column is flashed[28]. The checkerboard paradigm (CBP) has demonstrated a significant increase in performance over the RCP by avoiding vertical or horizontal adjacent letters[29-30].

In this experiment, we used the traditional white/gray flicker matrix containing the 26 letters of the Latin alphabet followed by the Arabic numerals from 1 to 9 and the hyphen-minus. For each trial, the three sequences are randomly chosen from 120 sequences generated using the CBP principle. Once all 36 symbols have been split into six groups, this algorithm is repeated 120 times. This algorithm can generate approximately $250\,000$ compatible images (from $C_{36}^6 \approx 2\,000\,000$ images including adjacent characters) yielding around 4000 CBP sequences. A sequence is built from the following algorithm 2–1 generating six valid images.

Algorithm 2–1 Pseudo-code to generate CBP sequence for BCI P300 speller

Data: a list of 36 characters C; AD list of horizontal and vertical adjacent characters.

$AS(c_i) = c_{i-6}, c_{i-1}, c_i, c_{i+1}, c_{i+6}$ if they exist: c_{i-6} if $i > 6$, c_{i-1} if $modulo(i, 6) > 1$, c_{i+1} if $modulo(i, 6) > 0$, c_{i+6} if $i < 31$

Result: a sequence S of six images, each highlighting six characters

1 Initialize sequence and list of available characters: $S = \{\}; AC = C$

2 **for** $i = 1$ *to 6* **do**

3 $\quad ACI = AC; J = \{\}; FCI = \{\};$ `    // Initialize available, chosen, and`
`                                         forbidden characters for the current`
`                                         image`

4 $\quad$ **for** $j = 1$ *to 6* **do**

5 $\quad\quad$ **if** $FCI \cap AC = \varnothing$ `        // No forbidden character should be in the`
`                                         available ones`

6 $\quad\quad$ **then**

7 $\quad\quad\quad c = random(ACI)$ `        // Randomly choose a symbol`

8 $\quad\quad\quad IM = IM \cup \{c\}$ `        // Saving chosen symbol`

9 $\quad\quad\quad ACI = ACI - AD(c)$ `        // Remove adjacent symbols from available`
`                                         ones`

10 $\quad\quad\quad AC = AC - c$ `            // Remove chosen character from available`
`                                         ones for all remaining images`

11 $\quad\quad\quad FCI = FCI \cup AD(c)$ `    // Saving forbidden characters`

12 $\quad\quad\quad$ **if** $j = 6$ **then**

13 $\quad\quad\quad\quad S = S \cup \{IM\}$ `        // Save image containing six characters`

14 $\quad\quad\quad$ **end**

15 $\quad\quad$ **else**

16 $\quad\quad\quad$ **break** `                  // Break when not enough remaining symbols`

17 $\quad\quad$ **end**

18 $\quad$ **end**

19 **end**

20 **if** $card(S) = 6$ **then**

21 $\quad$ **return** S

22 **end**

2.4.4.2 Cue and Trigger

Subjects are asked to spell words by focusing on each letter's color change. The letter is originally gray on a dark background and becomes white three times. The participant mentally counts up to three every time the letter is highlighted. The target letter is first circled with a green ellipse. This ellipse disappears as soon as the letters start to be highlighted. Participants are requested to spell 10 four-letter words: HOME, WITH, WHAT, GOOD, YOUR, FROM, MUCH, THEM, 6-17, and 2345. These words

are randomly ordered and result in 40 letters to spell. At the beginning of each trial, the current word is displayed in white in the upper-left quadrant of the screen, whereas the green ellipse surrounds the target letter for 1s. A black screen then appears for 1s, followed by a 3-complete sequence (3CS) of flashing letters that lasts for 4s.

2.4.5 Paradigm #5: P3005L (P300 for Five Letters Word)

The experimental paradigm for P3005L is the same as the one used for P3004L (Figure 2–5). The only difference is the change from four to five-letter words: ABOUT, BLACK, ENJOY, PRIZE, EQUAL, FALSE, HEAVY, EXACT, JUN88, and 13-59.

2.5 A Walk in the Hardware Acquisition

Three devices simultaneously record subjects sitting at an approximate 80 cm distance from a 23-inch TFT monitor (see Figure 2–6). The sound-attenuated and electromagnetically-shielded chamber is designed specifically for EEG recordings. To avoid head movements, participants are asked to keep their head on a chin rest while performing a BCI task. The headrest is fixed to the table with an angle of around 20°. Between two paradigms, subjects can freely move to relax. The monitor is a screen unit of the Tobii TX300 Eye-Tracker (Tobii Technology AB, Stockholm, Sweden) with a 1920 × 1080 pixel resolution at a refresh rate of 60 Hz.
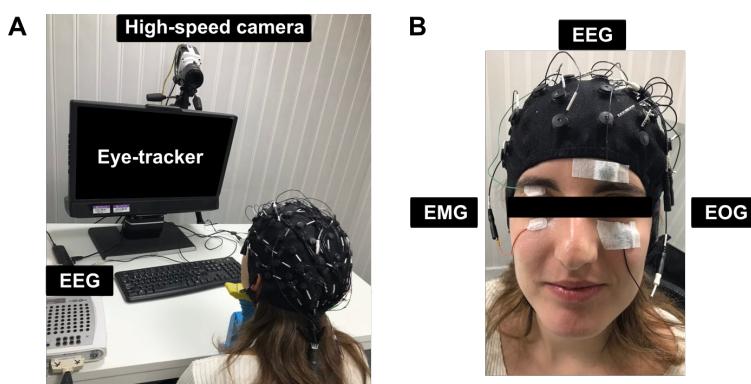
Figure 2–6 Data acquisition environment. Informed consent was obtained from the individual in the figure for the publication of the images.

2.5.1 EEG

A 65-channel Quik-cap acquires signals through a SynAmps2 system connected to an amplifier (Compumedics, Neuroscan). 62 EEG electrodes are placed according to the extended 10/20 system. The reference is located on the right mastoid (M1), with the ground electrode on the forehead. A bipolar vertical

EOG channel records the potentials with two electrodes placed above and below the left eye. The electrical activity produced by the Levator Palpebrae Superioris and the Orbicularis Oculi muscles around the right eye are recorded by two EMG electrodes. These electrodes are attached at the middle of the upper and lower eyelids, respectively. EMG activity is typically recommended to be sampled at a minimum of 1000 Hz, whereas EEG is commonly recorded at 250 Hz. Since the electrodes send data to the same system, the protocol globally records the continuous signals at a 1000 Hz sampling frequency.

2.5.2 Eye-Tracker

An eye-tracker (Tobii TX300) collects gaze-related data at 300 Hz. Eyeball position is calculated using the data from the pupil eye tracking system. This system uses an illuminator first placed closer, then further away to the optical axis of the imaging device, causing the pupil to appear subsequently lit up and black, respectively.

2.5.3 High-Speed Camera

A high-speed camera (Phantom Miro M310) captures a single eye with a resolution of 320*240 pixels. This is the smallest mode, allowing for a whole eye to be framed with a little margin in case of minor head movements. The focus is on the left eye, considering that the attached EOG electrodes are further away from the eyelids compared to the ones on the right eye. The eyelid position can then be extracted from the video, with both eyes being assumed to blink symmetrically.

2.5.4 Multi-Modal Acquisition

The cues are generated by the E-Prime software and displayed on the presentation screen. To avoid hardware limitations, the Tobii TX300 inner clock is used. By default, gaze-related data is recorded every 3.3 milliseconds. Thus, other sampling rates should be proportional to this amount. The Phantom high-speed camera has a maximum internal memory of 10 GB. With the mentioned resolution and a similar frequency to Tobii's (300 fps), the maximum recording time is around seven minutes.

To prevent data loss, the video sampling frequency is chosen at 150 fps. A binary trigger is sent from E-Prime every 6.6 milliseconds to an Arduino Nano (Atmega 328). This signal is then converted into a square wave, used as an external input to control the opening of the Phantom camera shutter. High-speed videos are directly recorded on the Phantom internal memory. After each task, the data is transferred and saved on a computer through the Phantom software.

The same E-Prime-generated trigger is also sent directly to the Neuroscan software, allowing for synchronousness between EEG and video recordings. The Tobii TX300 can not receive triggers with such a high frequency. To bypass this limitation, a light sensor detects the appearance of a white rectangle at the bottom right of the presentation screen at the start of each trial. Its output simultaneously sends a trigger

to the Tobii software using the Cedrus StimTracker, in order to provide a common time base across the three devices.
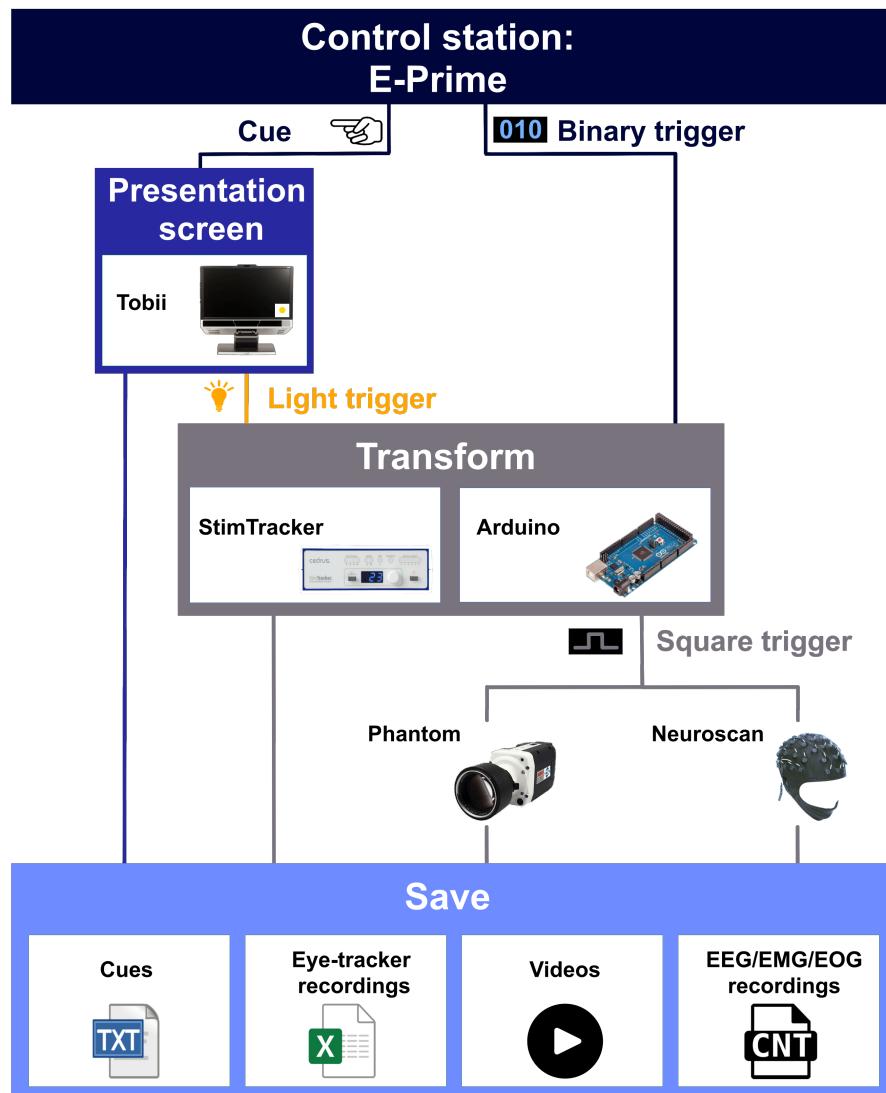
Figure 2–7 Experiment flowchart ensuring a common time base across the three devices: EEG, eye-tracking, and high-speed camera

The four programs (E-Prime, Neuroscan, Phantom, and Tobii) are run on two computers to prevent any RAM-related issues. The E-Prime computer is designated as the control station. It displays the cues on the presentation screen and records the eye-tracker data using the Tobii software. The second computer runs the two remaining softwares. A Python code is launched at the beginning of the experiment to automatically start and stop the recordings. At the end of each session, data synchronization is checked for consistency. The whole experiment flowchart is summarized in Figure 2–7.

A session started by randomly choosing the order of the paradigms. Participants were asked a few questions to assess their global condition (e.g., sleepiness, coffee intake, hunger). A session was punctuated by regular breaks to limit tiredness. During these interruptions, subjects were also asked about their degree of alertness (DOA), potential mistakes, or whether any exterior distraction occurred.

Technical details about the experiment and all necessary explanations about the global setup as well as the computer codes to reproduce the experiment are available online at https://github.com/QinXinlan/EEG-experiment-to-understand-differences-in-blinking. All data is also directly downloadable at https://www.kaggle.com/qinxinlaneva/datasets.

2.6 When Signal-to-Noise Ratio (SNR) Met Plots

Large-scale patterns of synchronized neuronal activity are inherently nonlinear and non- stationary. When considering a specific task, the mean and covariance properties generally change over different time segments. Theoretically, this nonstationarity reflects the different stages of a self-organized process[31]. The high inter-segment (or, more largely, inter-trial) variability coupled with possible time-varying relationships between the different measurement channels induces the non-Gaussianity of the signals.

2.6.1 Noise Characteristics

For an EEG electrode, the measured time series x is recorded at each instant t. Hereafter, the discrete time series will be noted symbolically as much as possible, without referring explicitly to the time: $x = x(t)$. The measured signal x is assumed to consist of the signal of interest (SOI) s and some additive noise n^a, the latter being generally interpreted as anything that is not the SOI: $x = s + n^a$.

This SOI happens when the desired cortical sources are activated by a specific event, for instance, in response to a certain stimulus. The investigated brain source(s) can be obscured by the brain's spontaneous activity and other observational noise. In particular, nonsystematic noise may arise from the recordings, the subject, or the task. In its broader meaning, noise can be decomposed into four components: (1) basic noise n^b, (2) added event-generated noise, (3) subtracted event-generated noise, and (4) signal variance-generated noise. The first part is the noise that is always present, regardless of whether the event occurred.

Any activated cortical source is likely to occur at shifting times and varying intensities, depending on the trial. Throughout an event, many cortical sources are activated, along with the SOI brain source(s). Some of them can be specifically generated by the event (added noise), while others will be suppressed because of it (subtracted noise). Both event- generated noises will be referred to altogether as n^e.

Assuming that the SOI cortical source can be detected (in spite of the latency variability), its intensity will still differ depending on the epoch. Traditionally, the SOI is presumed uncorrelated to the background EEG activity. Thus, when there is no event, the SOI should be equal to zero.

To compare similar signals across epochs, the signal is commonly first centered. Every epoch is

subtracted by the mean of the signal calculated before the SOI happens during a time length T, for instance, between $t_1 = 100 \ ms$ before the event to the start of the event at $t_2 = 0 \ ms$[32]. The baseline-corrected measured signal at an epoch x_i^c is then given by:

$$x_i^c = s_i + n_i^b + n_i^e = \bar{s} + n_i \tag{2--2}$$

Where $\bar{s}$ is the mean of the SOI. To simplify, afterward x will refer to the baseline- corrected measure signal instead of writing x^c. The whole noise during an epoch n_i is thus defined as:

$$n_i = s_i - \bar{s} + n_i^b + n_i^e = n_i^b + n_i^s \tag{2--3}$$

The centered basic noise n^b is assumed to be stationary and ergodic and can be represented by its variance since its mean is zero. On the other hand, since they are time-dependent, the added and subtracted event-generated noises n^e can not be assumed to be ergodic. Similarly, it is unlikely that the signal variance-generated noise is ergodic. For instance, the subject might get used to the task during the experiment, and the habituation would change the intensity of the SOI. Also, the electrodes could move slightly, or a myriad of reasons could cause this variability.

The measurement of specific components time series can be carried out through peaks or areas. The peak measurement approach presumes that the effect under investigation is large enough to be easily identified on both single trials and ERPs. Depending on the time series characteristics, the global maximum, several local maxima (resp. minima), or other parametric choices could be of interest. For smaller effect sizes, the high epochs' variability may render the analysis of single points meaningless. In this case, the area under the curve is usually favored[33].

To distinguish between the theoretical value of the expected mean $\mathbb{E}$ (resp. variance $\mathbb{V}$) and its practical computation, the symbol changes from a double-struck to a single-struck E (resp. V). In practice, the mean (resp. variance) cannot be calculated on an infinity number of trials. The sample mean is an unbiased[34] estimator of the population mean, which can be computed with:

$$E(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \int_0^T x_i(t) \tag{2--4}$$

And the unbiased variance is:

$$V(x) = \frac{1}{N-1} \sum_{i=1}^{N} \frac{1}{T} \int_0^T (x_i(t) - \bar{x}(t))^2 \tag{2--5}$$

2.6.2   Signal-to-Noise Ratio

By definition, the signal-to-noise ratio (SNR) is the dimensionless ratio between the level of a desired signal (i.e., signal power) and the level of background noise (i.e., noise power). It can be measured for a

single trial SNR_T or for the grand average (GA) of all trials SNR_{GA}. The GA of an event-related potential (ERP) usually exhibits a higher signal power for a lower noise power, and hence a better SNR.

When the brain patterns under investigation are buried in the rest of the signal (e.g., background EEG), the SNR will be small and the relevant patterns hard to detect (small effect size). On the other hand, a large SNR will simplify the BCI's detection and classification. Enhancing the SNR of event-related brain responses effectively and reliably is therefore of paramount importance to obtain a reliable estimate for single trials.

For a single trial, the measured signal can be expressed as the sum of the mean SOI and the whole noise $x_i = \bar{s} + n_i$, or equivalently the sum of the (varying) SOI and the centered basic noise $x_i = s_i + n_i^b$. The mean SOI is constant by assumption and thus independent of the noise. Similarly, the centered basic noise is postulated to be stationary and ergodic, and hence independent of the SOI. The power $\mathbb{P}$ of the measured signal, defined as the expected value of the squared signal, can then be written as:

$$
\begin{aligned}
\mathbb{P}(x_i) = \mathbb{E}[x_i^2] &= \mathbb{E}[s_i^2 + 2 * s_i * n_i^b + (n_i^b)^2] \\
&= \mathbb{E}[s_i^2] + 2 * \mathbb{E}[s_i] * \mathbb{E}[n_i^b] + \mathbb{E}[(n_i^b)^2]
\end{aligned}
\tag{2–6}
$$

Since the mean of the basic noise is assumed to be equal to zero, $\mathbb{E}[n_i^b] = \overline{n^b} = 0$, the power of the measured signal is simply:

$$
\mathbb{P}(x_i) = \mathbb{E}[s_i^2] + \mathbb{E}[(n_i^b)^2] = \mathbb{P}(s_i) + \mathbb{P}(n_i^b)
\tag{2–7}
$$

The SNR of a single trial can then be expressed as the ratio between the signal power $\mathbb{P}(s_i)$ and the basic noise power $\mathbb{P}(n_i^b)$:

$$
SNR_T = \frac{\mathbb{P}(n_T^{signal})}{\mathbb{P}(n_T^{noise})} = \frac{\mathbb{P}(s_i)}{\mathbb{P}(n_i^b)} = \frac{\mathbb{P}(x_i) - \mathbb{P}(n_i^b)}{\mathbb{P}(n_i^b)} = \frac{\mathbb{P}(x_i)}{\mathbb{P}(n_i^b)} - 1
\tag{2–8}
$$

The centered basic noise n^b being postulated as stationary and ergodic, its power can be calculated on any time interval. This means that the noise power can be calculated directly from the recorded signal before (or after) the SOI. With t_1 being the start of the noise, the SNR can then be computed with:

$$
SNR_T = \frac{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \int_0^T x_i^2(t)}{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \int_{t_1}^{t_1+T} x_i^2(t)} - 1
\tag{2–9}
$$

2.7 Summary of Results and Discussion

2.7.1 A Priori Sample Size Estimation

A prospective power analysis was conducted on five subjects to estimate the necessary total number of sessions (and subjects) needed to achieve statistically significant results. Three different approaches,

each based on different assumptions, were employed to estimate the required a priori sample size. The traditional method of approximating correlation power using power tables is a simplified and straightforward method but has limited accuracy. G*Power 3, a widely used software for statistical power analysis, provides a fast and convenient means of estimating the sample size but also relies on the Gaussian assumption.

To address these limitations, a new method called Fitted Distribution Monte Carlo (**FDMC**) simulation was developed in this thesis for estimating the a priori sample size. **FDMC** selects the least likely distribution from a pool of nearly 30 common distributions and treats it as an input of equal importance to the effect size, significance level, and power. Compared to the power tables, **FDMC** yielded a required sample size that was approximately seven times smaller, and 35% smaller than the sample size estimated by G*Power 3 (i.e., with the Normal distribution) (Table 2–4).

Table 2–4 A priori sample size estimation comparison from three methods (* fitted distribution from the prospective power analysis)

Method	Variability	Distribution	Minimum sample size (number of sessions)	Computational time
Power tables	No	Normal	436	~ 1 s
G*Power 3	Yes	Normal	97	~ 1 min
Fitted Distribution Monte Carlo (**FDMC**)	Yes	Fitted (SGED*)	63	~ 5 h

It is important to note that if the fitted distribution is indeed Normal, then **FDMC** will produce similar results to G*Power 3. In such cases, G*Power 3 should be preferred due to its significantly shorter computational time. However, in most real-life scenarios, the distribution of interest is unlikely to be Gaussian, especially when dealing with small sample sizes such as the ones usually encountered in BCI applications. Therefore, **FDMC** will yield a more accurate and easier to implement a priori sample size.

2.7.2 Segmented SNR and Source Localization Plots

Event Related Desynchronization (ERD, reduction in power) in the beta band during MI has been extensively documented in the literature. After cleaning and removal of bad channels (detailed in Chapter 3) and identification of the SOI (outlined in Chapter 4), the EEG data is averaged per subject, and the SNR plots are computed using Equation (2–9). The topographical distribution of SNR in the beta band (12.5-30 Hz) for both Left and Right MI is illustrated in Figure 2–8.

The SNR plot represents the average over time at each electrode. Source localization can also be
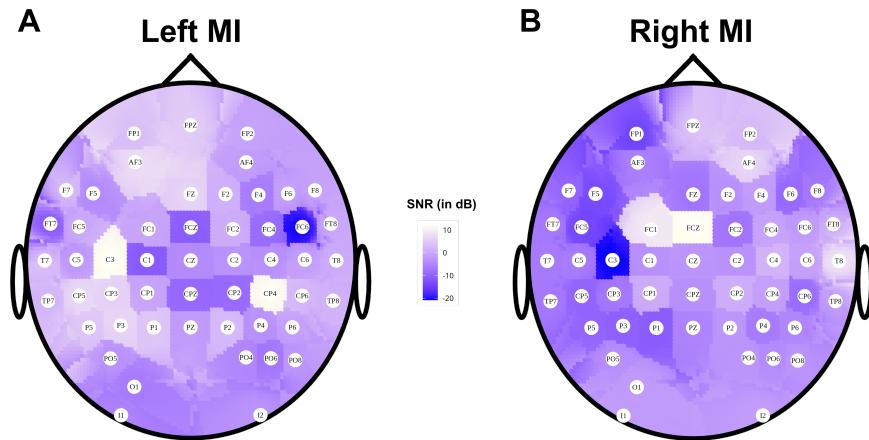
Figure 2–8 Segmented SNR topographies in the beta band (12.5-30 Hz) for both Left and Right MI (only the "good" electrodes are represented)

calculated at each time point with eLORETA and plotted at the time of interest (TOI) corresponding to the SOI (see Chapter 4 for more details). The eLORETA 3D representation of the TOI in the beta band for both Left and Right MI is depicted in Figure 2–9.
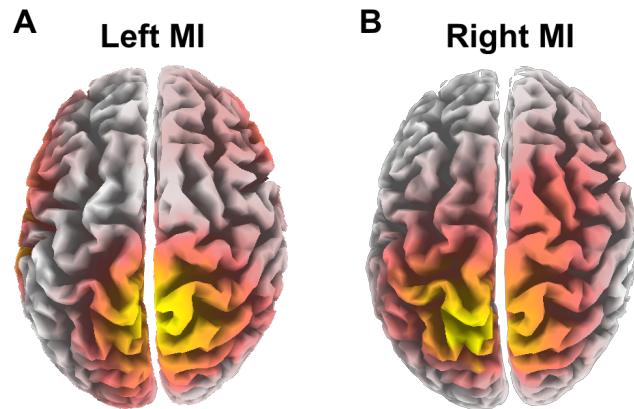
Figure 2–9 Source localization plot in the beta band (12.5-30 Hz) for both Left and Right MI

The juxtaposition of these two plot types accentuates the uniformity, whether intentionally chosen or arising from the computations, observed in the majority of conventional plots. However, we argue that the nature of the recorded EEG signals is characterized by irregularity and should thus be represented as such. This decision does not impact subsequent calculations but serves as a reminder of the distinctive nature of the signals under investigation.

2.8 Conclusion and Highlights

In this work, we publish a large dataset simultaneously recording electrophysiology and videos, along with synchronized eye-tracking, facilitating the investigation of spontaneous blinks. Data include a combination of Electroencephalography (EEG), Electrooculography (EOG), and Electromyography (EMG), which capture this major physiological disturbance impeding the user's intent.

The dataset includes four common BCI tasks: 1) motor imagery (MI) of the right or left hand, 2) motor execution (ME) of the right or left hand, 3) steady-state visual evoked potential (SSVEP), and 4) P300 BCI speller. An a priori sample estimation led to the data collection from 31 subjects, 63 sessions, resulting in 2520 examples of MI, ME, and SSVEP, and 5670 examples of P300. To the best of our knowledge, the current dataset is the sole existing resource focused on capturing various aspects of blinks through a multi-modal setup, allowing for the validation (or rejection) of insights previously derived from EEG alone.

The crucial points to be emphasized are:

- The Fitted Distribution Monte Carlo (**FDMC**) simulation for a priori sample size estimation developed in this thesis gives more practical and accurate results than other existing methods (seven times less sessions compared to power tables, and 35% less compared to G*Power 3).

- Participants characteristics are gathered from questionnaires, measures, and pictures (facial landmarks' extraction). Such data is rarely reported in datasets yet can be very useful for the comprehension of physiological signals.

- For replicability and robustness, the whole experimental process is detailed, and the code is shared at https://github.com/QinXinlan/EEG-experiment-to-understand-differences-in-blinking.

- The multimodal dataset from the thirty-one subjects (62 EEG, 2 EMG, 1 EOG electrodes, 1 eye-tracker, 1 high-speed video camera) is also directly available at https://www.kaggle.com/qinxinlaneva/datasets.

This dataset will be used throughout this thesis to develop innovative algorithms for pre- processing and processing. It has been collected to overcome the first two shortcomings listed in Chapter 1. New (offline and real-time) EEG artifact correction or classification algorithms may be efficiently tested on these data, and acquisition of similar multi-modal datasets can be easily reproduced as all required code and data is provided online.

Chapter 3 Adaptive Blink Correction and De-drifting (ABCD) Algorithm

This chapter aims at developing a novel method for accurate blink correction in EEG signals. The objective is to create an algorithm that can be used in real-time, without the need for additional sensors, and that only corrects the identified artifacts, avoiding systematic corrections. To achieve these goals, a realistic blink model needs to be derived from the data obtained in the experiment described in Chapter 2.

Blinks are detected from both EEG signals and the high-speed videos, allowing for verification by cross-checking the blinks detected from both modalities. The displacement of the upper eyelid is found to be correlated with the potential maximum during a blink. The resulting distribution of blink peaks exhibits significant inter- and intra-subject variability. This warrants the classification of blinks per subject and per class, where a class consists of all blinks falling within a specific peak range. The subsequent blink model is derived from the smoothed Grand Median, which is then adapted to each individual blink before being subtracted. This entire process is referred to as the Adaptive Blink Correction and De- Drifting (**ABCD**) algorithm, as it also eliminates drift curves from the EEG signals. **ABCD** also enables the removal of bad channels by analyzing Blink-Related Potentials (BRP) on adequately-chosen subsets of electrodes.

The **ABCD** algorithm is applied to all 31 subjects (for a total of 63 sessions). To assess its accuracy in comparison with other existing methods, such as Independent Component Analysis (ICA) or Artifact Subspace Reconstruction (ASR), each resulting cleaned data is further processed and classified using the methods detailed in Chapter 4. The classification accuracy, along with its confidence interval at 95% confidence level, (for details refer to Chapter 5) reveals a mean classification accuracy of 93.81% [74.81%; 98.76%] for **ABCD**, and 79.29% [57.41%; 92.89%] for ICA, and 84.05% [62.88%; 95.31%] for ASR.

The significant improvement in classification accuracy, as well as the comparative analysis of single blink data on frontal electrodes, highlights the superiority of **ABCD** over ICA and ASR. These results provide highly promising evidence that **ABCD** accurately correct blinks and should be tested on additional datasets. Further research could explore the application of **ABCD** in real-time analysis.

3.1 Introduction

3.1.1 Previous Explanations of Blink Effects in EEG

Traditionally, the influence of blinking on EEG signals has been attributed to an upward rotation of the eyeball during eyelid closure, where the eyeball is considered a dipole with corneal positivity and retinal negativity. This theory, known as Bell's phenomenon, originated from observations of attempted forced eyelid closure in patients with mechanical or neural lesions and from personal experiences of upward

corneal movement during rapid blinking[35].

However, with the advent of newer technologies, such as the magnetic search coil technique, the accuracy of this theory has been called into question. Studies utilizing this technique have revealed that only the upper eyelid moves along the surface of the eyeball during an unforced blink[36-37].

Another proposed explanation posits that eyelids function as sliding electrodes that short-circuit corneal positive charges. According to this view, the electrical charges of the eye originate from the corneal-retinal potential, where the retinal side carries a negative charge, and the corneal side holds a positive charge. During blinks, the eyelid comes into contact with the corneal surface, facilitating the conduction of positive potentials to the fronto-polar EEG electrodes. Consequently, during eyelid closure, the total electrical potential difference between the retinal and corneal sides of the eyeball becomes the sum of the retinal and corneal potentials[38].

3.1.2   Existing Algorithms for Blink Correction

Blinks are frequent and repetitive events that significantly disrupt the recording of neural activity in EEG signals. Various methods have been proposed to tackle this problem, such as removing data related to noticeable eye movements or correcting the influence of eye activity on the EEG using EOG correction. However, the high frequency of blinking in certain subjects makes it impractical to simply discard the contaminated data. From a practical perspective, it is also preferable to minimize the number of electrodes required for the setup. Therefore, algorithms that do not rely on additional sensors (e.g., EOG) are typically favored.

Numerous techniques have been developed to address blink correction in EEG signals. Among them, blind source separation (BSS) methods are widely employed. These statistical approaches assume that the multi-channel EEG recording can be viewed as a linear combination of underlying brain sources and blinks occurring instantaneously. The objective is to uncover these unknown sources s from the EEG recordings x. The process of de-mixing involves finding the inverse of the unknown matrix A, which is typically impossible unless additional constraints are provided and a square matrix is assumed:

$$\begin{cases} x = As \\ v = Wx \end{cases} \tag{3–1}$$

The W matrix is defined through the principles of decorrelation and independence. There are multiple W matrices that can be used for decorrelation, which accounts for the diversity of existing BSS algorithms. One of the most commonly employed methods, by far, is Independent Component Analysis (ICA)[39]. This approach can be conceptualized as approximating the data on lower-dimensional tangent space of a nonlinear Riemannian submanifold using nonlinear mapping functions.

ICA is a computationally intensive task that necessitates substantial computational resources. This problem can be bypassed by using the recursive technique, which employs the natural gradient and enables

online usage with only a slight decrease in efficiency. However, conducting two consecutive computations on identical data produces disparate outcomes, indicating a lack of consistency. Although the global statistical approach employed by this method aims to address errors comprehensively, it leads to a less precise model. Additionally, it applies corrections even to data that is unaffected by artifacts directly. In summary, it appears that applying ICA for blink correction is only suitable when precise results are not required or can be rectified using supplementary information. These critics pertain specifically to the utilization of ICA for blink correction. For other purposes, the use of ICA may be adequate.

Another commonly employed technique is the Artifact Subspace Reconstruction (ASR)[40]. The ASR method involves training a statistical model using clean calibration data, i.e., without artifacts, and attenuates them by decomposing short segments of EEG data and comparing them to the calibration data in the component subspace. This process operates on 500 ms chunks of data and uses principal component analysis (PCA) to identify the dominant principal components. If their standard deviation exceeds a threshold derived from the calibration data, they are removed from the window under scrutiny using a de-mixing matrix.

ASR offers the benefit of operating with low computational complexity yet corrects anything above a specified threshold. This correction process is entirely localized and doesn't depend on a statistical model. In fact, it has the potential to eventually "correct" a substantial source of interest (SOI). However, it should be noted that this system might still miss hidden or weakened blinks, which could affect its overall accuracy.

3.2 The Detection's Guide to the Blink

The Adaptive Blink Correction and De-drifting (**ABCD**) algorithm is developed with the aim of real-time implementation. The detection part, which is the first step of **ABCD**, is thus required to be as fast as possible while maintaining the lowest computational complexity achievable. Spatial, time, frequency, and amplitude parameters allowing for efficient blink detection are detailed hereafter.

3.2.1 Blink Characteristics and ABCD Parameters Choice

The effects of blinks on EEG signals are predominantly picked up by frontal electrodes but extend further[41]. Setting up one of the available frontopolar channels (FP1, FPZ, or FP2) as the reference bypasses the constraint of an additional electrode (e.g., EOG) for blink detection. In this thesis, we arbitrarily chose FP1, though any of the other frontopolar electrodes could also have been selected.

Though blink rate is highly task-dependent, its mean was found to be close to 17 blinks per minute at rest[42], meaning that a blink occurs on average every 3.5 seconds. Searching for blinks on 2s-time windows achieves faster removal of blink-free data, while ensuring that the remaining windows will contain only one blink on average.

In our experiment, the sampling recording EEG frequency has been set at 1000 Hz. Such a high frequency is usually not needed, especially to detect very distinctive blinks shapes. A down-sampling frequency of 200 Hz decreases the computational load yet safeguards its accurate time peak detection.

A human eye blink has an amplitude that can be 10 to 100 times larger than the electrical signals originating from the cerebral cortex[43]. Some blinks may have very low amplitudes, where the smallest detectable by **ABCD** may only increase by 35 µV (on FP1) compared to the averaged background activity. This amplitude threshold has been found to discriminate between eyelid movements that do not cross the pupil and 'real' blinks. Heuristically, all blinks maximum amplitudes on FP1 are included in a [40; 450] µV interval. Blinks higher than 360 µV are likely caused by a mixture of events, for example, two successive blinks or a blink accompanied by an eye movement, and might occur frequently for some subjects.

Blinks often last from 200 ms to 400 ms[43], with a shorter anterior width compared to the posterior width. These time length differences might be the result of the morphological disparities between the OO and LPS muscles involved. A total time length of 500 ms guarantees the whole coverage of the blink, where the anterior width is set at 200 ms and the posterior one at 300 ms.
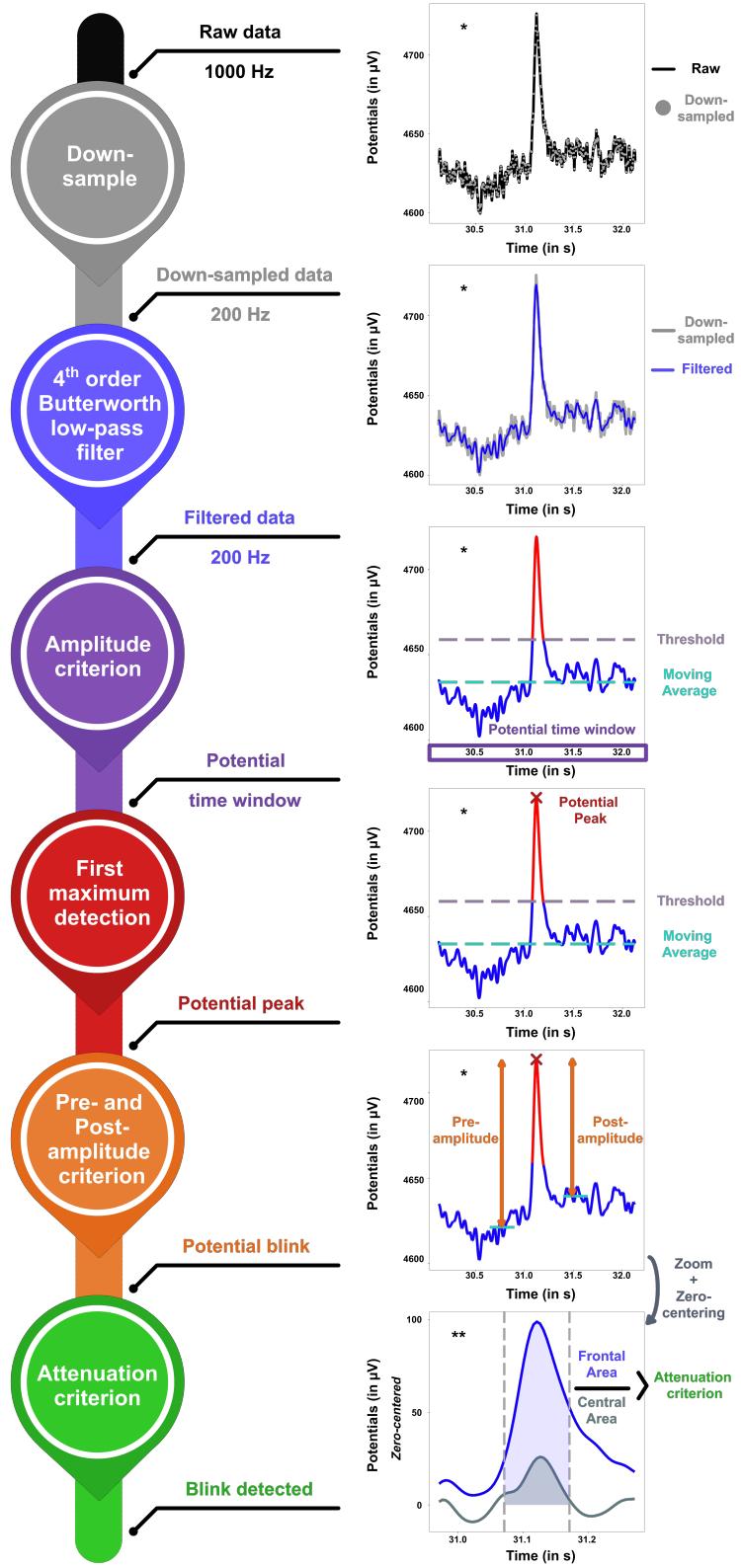
The effects of eyelid movements propagate differently across the scalp[41]. The blink's amplitude is highest at the frontopolar electrodes and decreases until nearly completely vanishing at occipital electrodes. Central channels (e.g., CZ) have been found to be close enough to the eyes to display the blink's effect and far enough to reveal the signal attenuation. Table 3–1 summarizes all the parameters choice for **ABCD**.

Table 3–1 Adaptive Blink Correction and De-drifting (ABCD) parameters choice

Parameters	Values
Reference channel	FP1
Down-sampling frequency	200 Hz
Time window	2s
Low-pass filter	10 Hz
Amplitude threshold	35 µV
Anterior width	0.2 s
Posterior width	0.3 s
Comparison channel	CZ

3.2.2 Blink Detection in EEG

The amplitude threshold γ_s is set such that a detection has a probability less than 0.003% of falsely detecting a blink (i.e., detecting a blink when it is normal neural activity). Mathematically, this is equivalent to saying $\gamma_s = 3\sigma$, where σ is the standard deviation of EEG data on the sliding 2 s-time window.

Figure 3–1 Blink detection flowchart

Algorithm 3–1 Pseudo-code for blink detection in the **ABCD** algorithm

Data: raw EEG data on reference $s_0(t)$ and central electrode $c_0(t)$, duration time window T, number of time windows N, minimum amplitude threshold γ_s, duration time for pre- and post-criterion T_p, anterior width W_a, posterior width W_p, and attenuation ratio r_a

Result: blink

1 Initialize the blink index: $b = 0$;

2 **for** $i = 1$ *to* N // Loop over all N time windows of size T

3 **do**

4     $s(t) = B_4 \circ D_5(s_0(t))$ // Data pre-processing: Down-sampling and Filtering. For simplicity, s_i is noted s

5   **if** $\overline{s} = \frac{1}{T} \sum s(t) \geq \gamma_s$ // Amplitude criterion from moving average

6   **then**

7        $\begin{cases} \widehat{A_M} = \max(s(t)) \\ t_0 = s^{-1}(\widehat{A_M}) \end{cases}$ // First maximum detection: $\widehat{A_M}$ is the peak potential and t_0 is the peak time

8   **if** $\overline{s_{Pre}} - \widehat{A_M} \leq \gamma_s$ *and*   // Pre-and post-amplitude criterion: $\overline{s_{Pre}}$

9        $\overline{s_{Post}} - \widehat{A_M} \leq \gamma_s$ (resp. $\overline{s_{Post}}$) is the average before (resp. after) the detected peak during a time interval T_p

      then

10         $c(t) = B_4 \circ D_5(c_0(t))$ // Data pre-processing for central electrode

11         $\begin{cases} S = \sum_{t=t_0-W_a}^{t_0+W_p} s(t) \\ C = \sum_{t=t_0-W_a}^{t_0+W_p} c(t) \end{cases}$ // Compute areas around blink on reference electrode and central electrode

12   **if** $\frac{S}{C} > r_a$ // Attenuation criterion

13   **then**

14            $b = [b; t_0]$ // Blink is detected

15   **end**

16   **end**

17   **end**

18 **end**

19 **return** b

Statistical analysis on the recorded data has shown that the maximum standard deviation (SD) is equal to 12 μV among all electrodes (10 μV for the maximum SD of each electrode SD, and 7 μV for the max median SD of each electrode SD). A conservative estimates thus yields $3\sigma = 36\mu V$, which justifies why this threshold has been set at $\gamma_s = 35\mu V$.

The first step consists in down-sampling to 200 Hz and low-pass filtering the data with a fourth-order 10 Hz Butterworth low-pass filter on 2 seconds-wide windows over a reference frontopolar channel (FP1).

This pre-filtering avoids unnecessary variability while keeping the potential shape of a blink if there is one to be detected. The amplitude criterion then selects all data higher than the moving average to which a heuristic threshold is added, as highlighted in Figure 3–1. If no data is greater than the threshold, **ABCD** swipes right to the next window. Most of the blink-free windows are already discarded after this first criterion.

The fourth step consists merely in calculating the maximum amplitude on the selected window. Up to this point, noisy data can be mistakenly considered as a potential blink. For example, electrode "pops" or other kinds or artifacts might exceed the aforesaid threshold. To remove them, the differences between the maximum amplitude and the average mean before the Anterior Width and after the Posterior Width are computed.

If both these values are greater than the threshold, the Pre- and Post-amplitude criterion is validated, assessing the blink shape likeliness. The data around the maximum potential is then extracted and zero-centered. Since blinks are known to propagate along the scalp, the area at the reference channel FP1 and the area at a further electrode (CZ) are compared. A ratio comparison allows eliminating brain-related (only) or eyeball movement events and consists of the last criterion (i.e., the attenuation criterion). When all three criteria have been verified, the blink and its maximum amplitude are extracted for distribution analysis and classification.

Let $s_0(t)$ be the raw EEG recording on the reference electrode (e.g., FP1), sampled at 1000 Hz. The corresponding time difference between two successive points is then $\tau_0 = 1$. Let's also define $T = 2000$ the duration of a time window in milliseconds, and γ_s the minimum amplitude threshold (on the reference electrode). Let's note s^{-1} the reciprocal function giving t_i from the amplitude $A_i = s(t_i)$. Two functions operate on the initial signal. D_5 is the down-sampling function, which has a modulo 5 action, meaning that it retains only one element on the five of the $s_0(t)$ time series. B_4 is the fourth order Butterworth low-pass filter. The whole blink detection process is summarized in the following pseudo-code (summarized in Algorithm 3–1).

3.2.3 Blink Detection in Video

As described in Section 2.5, a high-speed camera records the left eye. A simple template matching algorithm in Python is used to detect five landmarks: the inner and outer corners of the eye, the middle upper and lower eyelid, as well as the pupil (see Figure 3–2). Once the landmarks are detected, the eye width and openness can simply be deduced from the subtraction of the corresponding landmarks' positions, expressed in pixels. Blink detection is confirmed by the lack of pupil landmark identification.
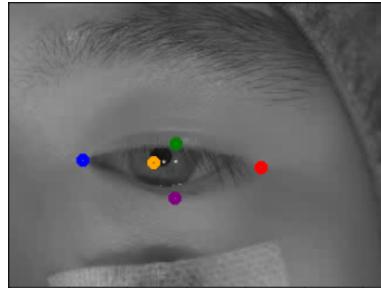
Figure 3–2 Detection of five physical landmarks

3.3 To Kill a Bad Channel

Up to this step, **ABCD** only requires (and relies on the signal quality of) two electrodes: one frontal and one central. However, non-biological artifacts (e.g., malfunctioning electrodes or power line) are a recurrent issue during BCI experiments. In line with this thesis rationale, this pre-processing step is carried out by taking into account the physiological shape of the blink propagation.

3.3.1 Blink Propagation

An extremely simplistic blink propagation model can consider the eyes as the point source and the effect of blinking as a single wave (or pulse) propagating in anisotropic media. Since the EEG electrodes are located on the skull, blink propagation may then be reduced to waves propagating on a 2D sphere (i.e., the head). The electric field produced by the blink induces a potential given by the electrostatic laws and implies a $\frac{1}{r^2}$ attenuation factor relative to the equivalent dipole situated at the reference electrode FP1:

$$V_{th} = \frac{Q}{4\pi\epsilon_r d_{FP1}^2} \tag{3–2}$$

Where Q represents the unknown dipolar moment and d the distance to the dipole center, here FP1. It

Table 3–2 Experimental and theoretical potential propagation comparison at central electrodes

Electrodes	FZ	FCZ	CZ	CPZ	PZ
d_{FP1}	86.12	118.00	148.33	173.80	192.28
V	54.35	31.11	21.22	16.35	13.83
V_{th}	57.92	32.32	21.38	16.25	13.83
w_s	0.84	0.88	0.92	0.96	1

should be noted that if the synthetic permittivity value were known, the dipolar moment project norm could be calculated within the approximations made using normalized distances. Reciprocally, given a known artificial dipole, this unknown synthetic permittivity could be deduced. Setting the skull width ratio (w_s)

at 1 for the farthest electrode, the theoretical potential at each electrode can then be computed using the distance between electrodes, which are calculated from the MNI Average Brain (305 MRI) Stereotaxic Registration Model[44].

$$V_{th} = \frac{Q w_s}{4\pi\epsilon_r d_{FP1}^2} \tag{3-3}$$

The effective attenuation can be measured from the real data and compared to the theoretical calculation corrected with the skull thickness (see Table 3–2). Figure 3–3) illustrates these theoretical and real potentials, revealing a nearly perfect match between the two curves. The amplitude is expected to decrease
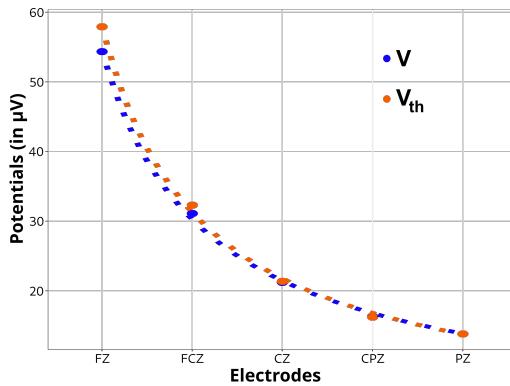
Figure 3–3 Quasi-perfect match of theoretical and real potential attenuated propagation at the central electrodes
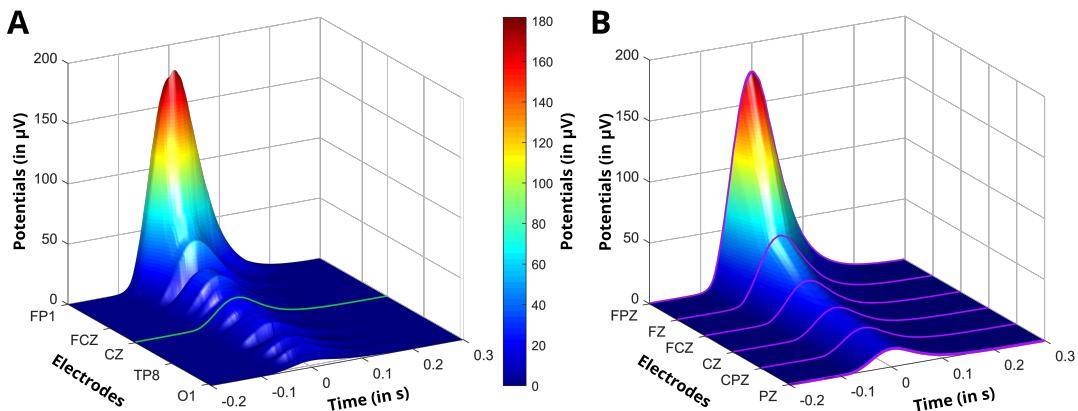
Figure 3–4 (A): Blink propagation on all electrodes (the electrode used for the propagation criterion - CZ - is in green); (B): Blink propagation on the midline electrodes

as the distance increases from the eyes. The global shape recorded by further electrodes remains similar, with a blink base width (that would correspond to a wavelength for a sinusoidal wave) also shrinking with the distance. Figure 3–4A displays the effects of blink propagation on all electrodes, with a special highlight for the central electrode CZ that is used in the attenuation criterion (Figure 3–1). Figure 3–4B

47

displays only the midline electrodes, for which the maximum blink amplitude decreases monotonically as the distance to the eyes increases.

3.3.2 Bad Channel Removal

Improper recordings by some of the EEG channels may arise from various technical reasons. Abnormal impedance or broken wire contacts can cause malfunctions, and too much gel may create a bridge between two or more electrodes. In the case of malfunctioning sensors, the corresponding EEG recordings will look very dissimilar to the signals from the neighboring good channels. In the extreme situation where the electrode is not able to record anything, the resulting signal can even be a flat line. On the contrary, bridged electrodes will display an identical signal as their neighbors.

In both scenarios, bad channels can be automatically identified using blinks. The Blink- Related Potential (BRP) is computed on each electrode and compared to the median of the BRP neighboring electrodes. The median is here favored, as it better avoids the influence of outliers. The longest common subsequence (LCSS) is then computed between these two trajectories. When this LCSS is below an arbitrarily chosen threshold, the corresponding channel is labeled as malfunctioning or "bad" (see Figure 3–5). In the event of bridged electrodes, the BRP on the problematic channels is indistinguishable. Setting a threshold on the BRP time-series difference allows for identifying the bridged or "bad" electrodes. Let's
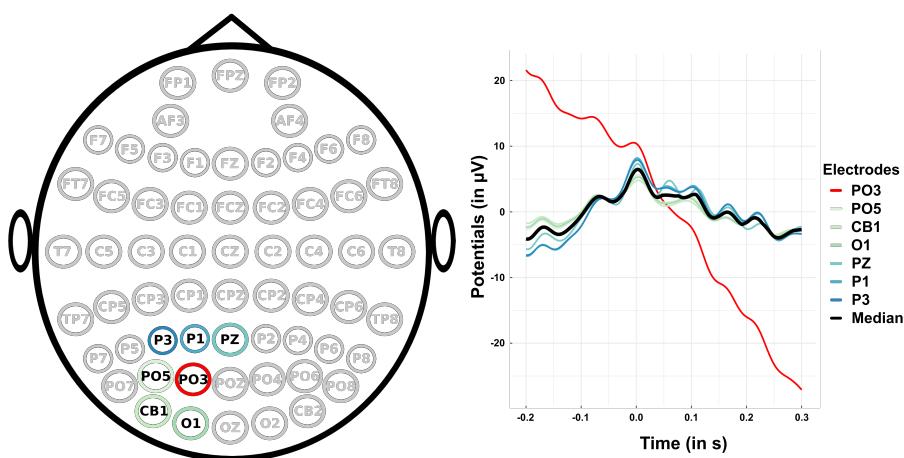
Figure 3–5 Identification of bad channels using the blink-related potential (BRP) on neighboring electrodes

call the electrode on which the BRP_i is calculated and $e_{i,n}$ the neighboring electrodes. The bad channel detection process is summarized in the following pseudo-code (Algorithm 3–2):

The experiment as described in Chapter 2 recorded data from 62 EEG electrodes. If any channel was labeled as bad in any session, it was removed from the whole dataset. After this pre-processing step,

Algorithm 3–2 Pseudo-code for bad channel detection

Data: the blink-related potential and $e_{i,n}$ at the neighboring electrodes, L threshold for LCSS bad channel detection

Result: bad and bridged channel

1   Initialize bad and bridged electrodes: $bad = []$; $bridged = []$;

2   **for** $i = 1$ *to E*   `// Loop over all E electrodes`

3   **do**

4      $BRP_{i,n} = \text{median}(BRP_j)$ with $e_j \in e_{i,n}$ `// BRP at the neighboring`
                                                   `electrodes`

5   **if** $LCSS(BRP_i, BRP_{i,n}) = 0$ **then**

6         $bridge = [bridge, i]$ `// Bridged electrode detected`

7   **else if** $LCSS(BRP_i, BRP_{i,n}) < L$ **then**

8         $bad = [bad, i]$ `// Bad electrode detected`

9   **end**

10   **return** $(bad, bridge)$

---

the number of sensors dropped to 51, with the following 10 malfunctioning channels excluded: PO3, F1, POZ, OZ, F3, O2, P8, PO7, FC3, P7 and the P4 electrode being bridged.

3.3.3   Drift Curve Removal

Another persistent non-biological artifact presents itself as a direct current drift that is usually caused by sweating or, on the contrary, by gel drying. It produces low-frequency shifts that, if left unattended, will generate biased outweighs for some of the electrodes (leading to false source localization).
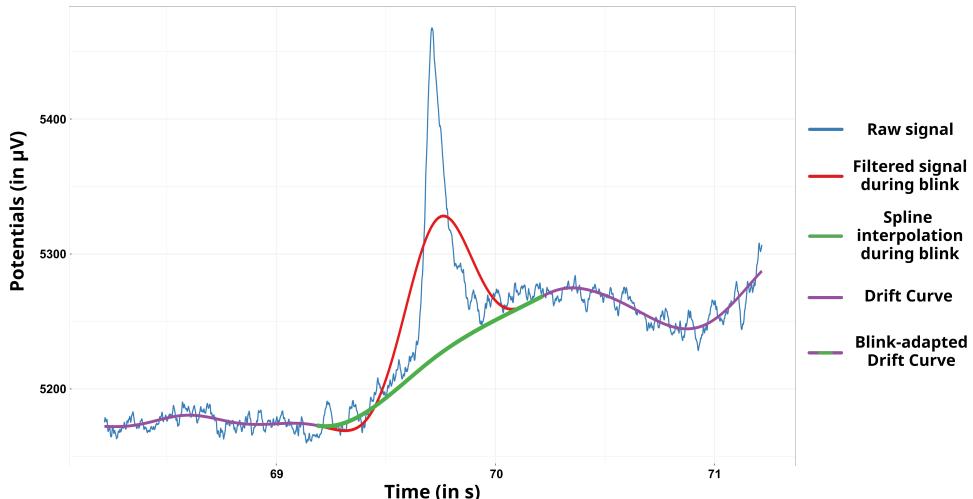
Figure 3–6 Drift curve removal in the blink's vicinity

Drift curves can be computed on each channel by applying a fourth-order 2 Hz low-pass Butterworth filter that reveals the global trend (Figure 3–6). In the vicinity of blinks, this drift curve is replaced with a spline interpolation to avoid removing part of the blinks mistakenly. Let's call the window signal, where a blink has been identified. The filtered drift of the signal window is then $D_{SB} = Spline[B_4(s_B)]$. The ensuing blink-adaptive drift curve is then simply removed on each electrode with $s_B - D_{SB}$.

The EEG signals obtained after removing blink-adapted drift curves are non-biological artifact-free. The blink-related potential (BRP) can then be computed again and its de-drifted characteristics examined and classified. Technically, the bad channel removal step and the drift curve removal step could be interchanged. However, this order proved more computationally efficient.

3.4 The Great Distribution

Distribution of the blinks' peak amplitude, recorded on the reference channel FP1 after Drift Curve Removal (DCR), displays a high intra- and inter-subject variability. How can these differences be explained?

3.4.1 Blinks Distribution

As the tear fluid evaporates, usually within 15 to 30 seconds after a blink[45], lubrication of the cornea is ensured by the eyelids' motion. Though blink rate is highly task dependent, its mean was found to be close to 17 blinks per minute at rest[42], which is several times more than required for ocular lubrication. This over-blinking has been assumed to have a role in disengaging attention during a cognitive behavior.
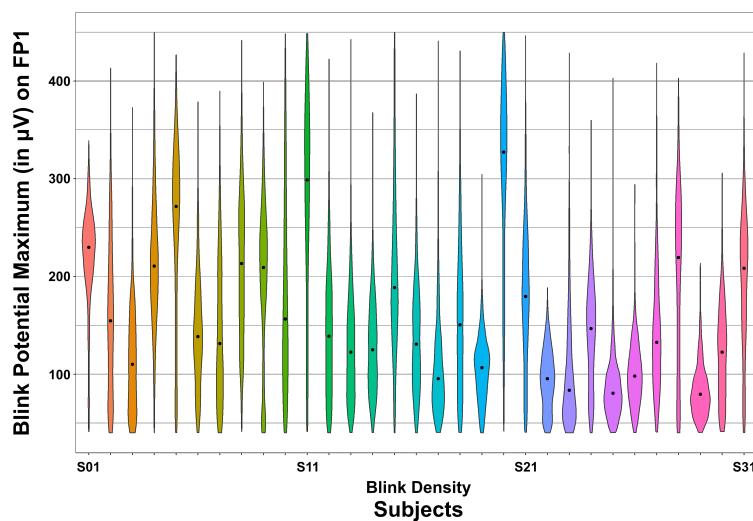
Figure 3–7 Violin/density plots showing the blink distribution on the frontopolar channel (FP1) for all subjects; the median is indicated as a black point

To counterbalance this overblinking and minimize the time that vision is obscured, most blinks in most subjects are not complete, meaning that the upper eyelid usually comes to rest before attaining the lower eyelid[45]. This creates high intra-subject variability with standard deviations ranging from ~20 µV to ~80 µV and high inter-subject variability with medians varying from ~80 µV to ~295 µV (resp. means range from ~85 µV to ~290 µV) (see Table 3–3 and Figure 3–7).

Table 3–3 Blink potential distribution parameters per subject

Subject	Peak potential Mean ± SD (in µV)	Frequency Mean ± SD (in blinks/min)	Subject	Peak potential Mean ± SD (in µV)	Frequency Mean ± SD (in blinks/min)
S01	220.4 ± 50.8	16.9 ± 5.0	S18	155.6 ± 72.8	12.2 ± 4.2
S02	158.3 ± 76.8	9.3 ± 1.9	S19	105.2 ± 33.1	9.5 ± 1.2
S03	115.0 ± 53.3	9.4 ± 2.5	S20	306.1 ± 93.8	42.8 ± 3.5
S04	214.4 ± 67.1	34.5 ± 7.8	S21	178.3 ± 71.2	18.1 ± 2.5
S05	250.3 ± 88.7	9.2 ± 4.3	S22	94.5 ± 30.4	16.8 ± 2.0
S06	142.7 ± 55.9	20.7 ± 3.4	S23	101.6 ± 59.7	7.3 ± 1.6
S07	144.3 ± 72.1	12.6 ± 2.9	S24	142.3 ± 52.7	46.1 ± 3.4
S08	211.4 ± 74.5	45.0 ± 14.4	S25	85.7 ± 32.1	15.4 ± 4.2
S09	195.5 ± 77.9	17.9 ± 5.0	S26	100.2 ± 31.9	8.2 ± 2.4
S10	178.6 ± 96.9	14.0 ± 3.5	S27	142.2 ± 61.7	27.7 ± 4.5
S11	279.0 ± 101.8	32.9 ± 5.9	S28	199.8 ± 87.2	25.5 ± 1.7
S12	142.5 ± 63.2	11.9 ± 3.3	S29	83.8 ± 27.6	15.3 ± 3.4
S13	127.7 ± 53.9	54.9 ± 10.9	S30	121.5 ± 47.5	13.4 ± 2.2
S14	127.7 ± 47.0	22.3 ± 4.3	S31	194.7 ± 69.4	13.0 ± 4.2
S15	201.4 ± 82.9	28.8 ± 6.2			
S16	135.9 ± 53.9	26.7 ± 8.5	**Mean**	**160.1 ± 56.4**	**20.8 ± 12.8**
S17	106.6 ± 50.3	7.2 ± 1.5	**SD**	**62.5 ± 20.4**	**4.3 ± 2.9**

Full closure can be defined as the lowest point the upper eyelid reaches, or, in other words, the most closed position, which does not necessarily imply complete closure.

3.4.2 Upper Eyelid Displacement

The peak amplitude of a blink in frontopolar EEG channels reveals the global eyelid displacement (closing plus opening). Figure 3–8 displays two blinks with dissimilar peak amplitudes (~120 µV and ~220 µV). For the smaller blink, the upper eyelid is lower (i.e., in a more closed position) at the beginning ($t_1 - 0.2$) and at the end ($t_1 + 0.3$), showing that the smaller blinks EEG effects happen with shorter eyelid displacements.
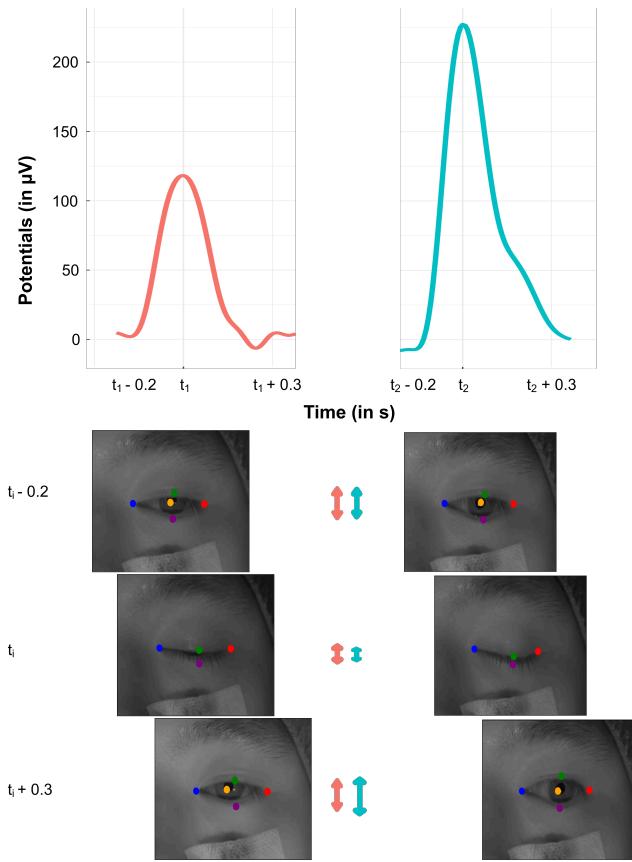
Figure 3–8 Comparison of blinks with different peak amplitudes (centered at t_1 and t_2); EEG blink signals extracted after fourth-order 10 Hz Butterworth low-pass filter on drift curve removed signals

In this example, both blinks last for the same period of time, illustrating that higher velocity of the upper eyelid is also correlated with higher EEG amplitude[46].

3.4.3 Blink-Related Potential

Three kinds of electrodes are employed to record blinks: the frontopolar EEG channels, the electrooculogram (EOG) electrode (placed above and below the left eye), and two electromyogram (EMG) sensors (one placed on the right eye upper eyelid and one on the right eye lower eyelid).

On the de-drifted signals, blinks are extracted, and their median yields the Blink-Related Potential. Figure 3–9A displays the corresponding potentials (in microvolts), where base width corresponds to the time width for which the potentials reach zero and half maximum width is the one for which the potentials reach half of the maximum amplitude. Similar averaging is computed for the distance in pixels between the upper and lower eyelids. The normalized values of potentials and distance are represented in Figure 3–9B.
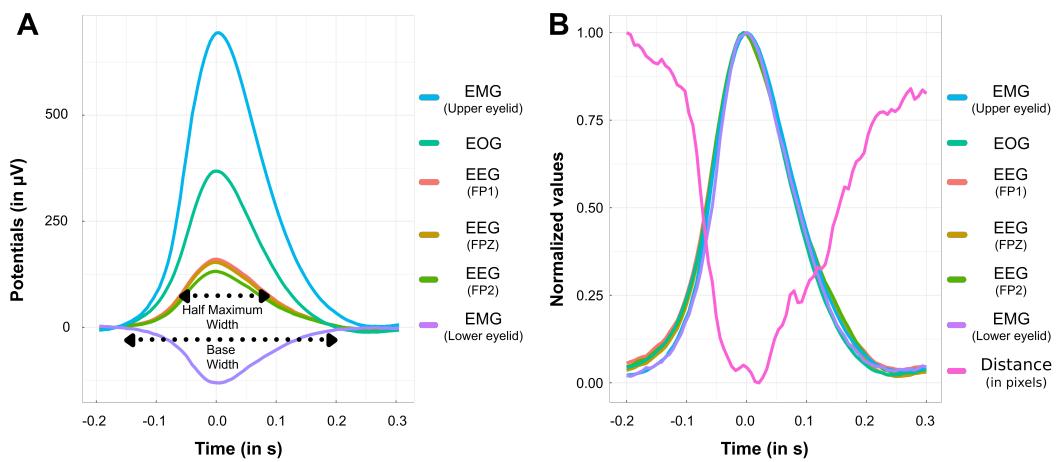
Figure 3–9 Blink-Related Potential signals comparison (A) for averaged signals from EEG, EMG, and EOG; (B) for normalized averaged electrical (EEG, EMG, EOG) and video signals

The different sensors all record the same phenomenon, i.e., the movement of the eyelids during the blinks. The peaks are reached when the eye is the most closed. Most blinks were not accompanied by a pupil movement, which is thus not represented in this figure.
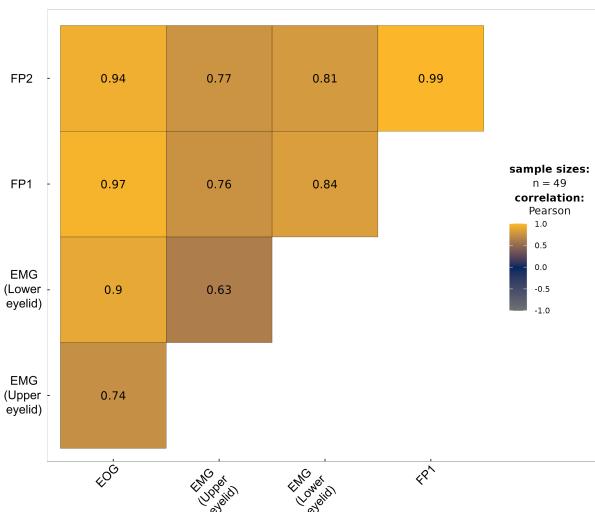
Figure 3–10 Correlogram between areas under the curve of the normalized BRP potentials (for EEG, EMG, and EOG)

The areas under the curve of the normalized BRP potentials are then computed for each subject. The correlogram[47] in Figure 3–10 highlights that all sensors are significantly correlated during blinks, or, in other words, that they all record the same signal at different levels. Hence, EMG and EOG electrodes are redundant with the EEG frontopolar channels and can thus be discarded from further analysis.

3.5   The Complete Blink Classification

The aim is to find an adaptive blink correction that will not only fit the EEG blink artifact better but will also reveal more accurately concealed brain signals once removed. The high blinks potential variability observed in the statistical analysis issued from our dataset warrants the use of a quantified error model.

3.5.1   Inter-Subject Variability

Blinks distribution is highly subject-dependent, with medians varying from ~80 μV to ~295 μV (see Figure 3–7). Can physical characteristics explain these differences?

For each subject, the area under the BRP is calculated on each electrode (the peak is deemed to not hold enough information to proceed). The goal is to find a linear relationship across subjects between one or a combination of two electrodes areas and one or a combination of no more than three explanatory variables (for computational time's sake). The usual arithmetic operations ⁻addition, subtraction, division, multiplication, as well as the square function ⁻are applied when combining the variables. This algorithm is inspired by the 21 dice game, in which all numbers from 1 to 21 must be found by combining the numbers from three dice.

Eleven variables have been investigated, specifically six derived from the facial landmarks – Jaw Width, Left Eye Width, Right Eye Width, Inner Canthi Distance, Outer Canthi Distance, Upper Nose to Lower Chin Distance ⁻as illustrated in Figure 2–4, three quantitative from the demographic, behavioral, and physical characteristics ⁻Height, Head Circumference, Nasion-Inion Distance ⁻summarized in Table 2–3, and two computed from the BRP ⁻Half Maximum Width, Base Width. The computational complexity being quite large for eleven variables, the 21 algorithm is first run only on a few combinations of electrodes.

The resulting linear relationships linking an electrode function to a physical characteristics function are then ordered by r-squared value. The best-ranking linear relationships link time values (Half Maximum Width, Base Width) and areas computed from the BRP with four recurring physical characteristics, namely: Head Circumference, Inner Canthi, Outer Canthi, and Nasion-Inion Distance ($R^2 = 0.81$). In other words, this variable importance analysis reveals that the shape of the skull impacts mostly the blink distribution, as could be expected.

3.5.2   Intra-Subject Variability

The typical parameters of a blink waveform are its peak maximum amplitude, base width, and half maximum width. For a specific peak maximum amplitude, the base width and half maximum are correlated. Higher peaks will have larger base widths but not systematically larger half maximum widths. The high intra-subject variability results in a variety of blink shapes that mostly depend on the maximum potential on the reference electrode. Consequently, our current objective is to develop an optimized

methodology that adequately models these blink waveforms, taking into consideration their inherent variability.

3.5.3   Blink Class Segmentation

The Blink-Related Potential (BRP) is defined on both subjects and classes to account for both inter- and intra-subject variability. To avoid potential outliers, the median is computed for each class and each subject. Such a class and subject-dependent median BRP will be referred to hereafter as the Grand Median (GM). How many blinks (i.e., minimum blinks sample size) should a class consist of?

The variance of the BRP can help decide on the minimum number of blinks a class should hold. The resulting median should be representative of the class and, as such, unchanged with an increased number of blinks. The smaller the class width, the easier the fitting to a single blink will be. However, it comes at the cost of longer EEG recordings to ensure the minimum number of blinks per class. Figure 3–11A illustrates histograms with different class widths (or bin widths) for a single subject in a single session, i.e., for about 1 hour of recording. For the smallest class width, the number of blinks per class rarely exceeds 20 blinks.
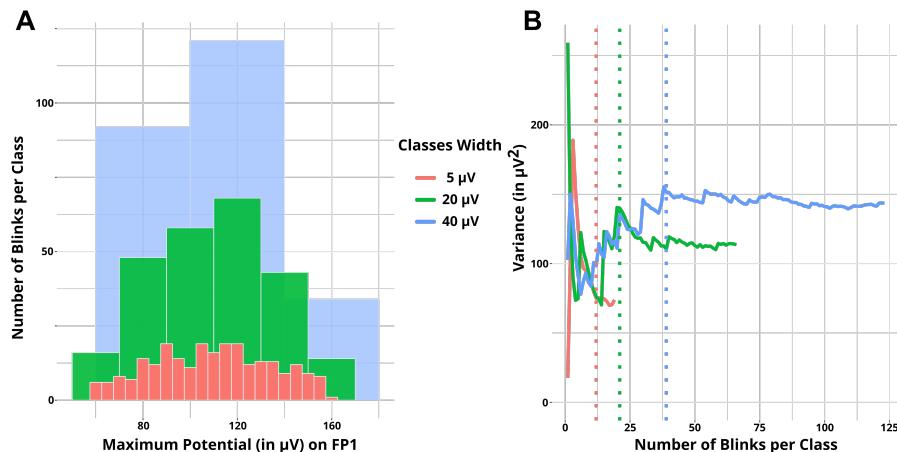
Figure 3–11 Comparison of three class widths for a subject: A. Histogram, B. Variance (dotted lines represent the minimum of blinks for each class width); smaller class widths have a better fit to single blinks and require less blinks to obtain a representative median, but their probability of reaching the required minimum sample size is lower for the same amount of data compared to medium class widths

The corresponding GM BRP are computed, and their variances are calculated according to Equation (2–5). Figure 3–11B depicts the evolution of each GM variance for an increasing number of blinks N. The minimum number of blinks per class is then computed as the value for which the absolute difference between more than five successive variances is lower than a chosen threshold.

As expected, the variance decreases faster for smaller class widths, with a minimum number of blinks

as low as ten for a class width of 5 µV (resp. twenty blinks for 20 µV and forty blinks for 40 µV). Similar small classes would, however, require extensive recordings before reaching this threshold. On the other hand, a larger class width will be less likely to fit to a single blink. To trade-off between reasonable recording times and single blink fitting, the chosen class width is 20 µV with a minimum number per class at twenty blinks. Any class with fewer blinks than that will not have its GM computed. The corresponding blinks will be attributed to their neighboring class instead. The rationale behind this decision is that it is better to tweak a more representative GM than stretch a less untypically shaped GM.

3.6 And Then There Were None: Adaptive Blink Correction and De-drifting (ABCD) Algorithm

One of the main contributions of this thesis is an algorithm that adjusts to the specific shape of any single blink (and removes it). The aim is to find an adaptive blink correction that will not only fit the EEG blink artifact better but will also reveal more accurately concealed brain signals once removed.

The resulting GM BRP per class are represented in Figure 3–12A. The remaining irregularities should disappear with an increased number of blinks per class, yielding a theoretical smooth GM. To avoid introducing variability in classes with a small number of blinks, the GM BRP per class is artificially smoothed. More specifically, the BRP is assumed to be equal to zero when it is far from the blink's peak, and the GM BRP is presumed accurate even for a number of blinks as low as twenty in the vicinity of the blink's peak. In between, the GM BRP should increase (resp. decrease) as illustrated on 3–12B.
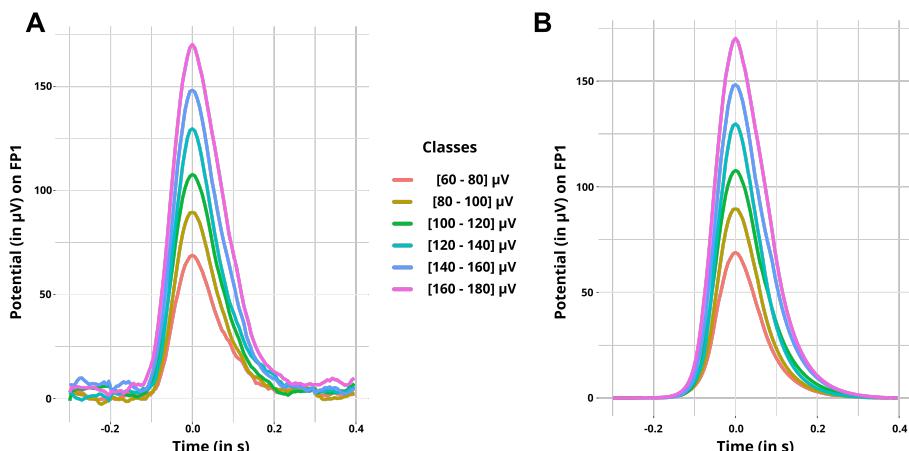
Figure 3–12 Comparison simple and smoothed GM BRP per class on the reference electrode FP1 for one specific subject

The raw EEG signals were also analyzed using EEGLAB with similar processing steps as prior published studies[48-49]. Data from the 51 remaining channels after bad channel detection were used to avoid

human error with visual inspection of artifactual electrodes. The signals were re-referenced with CAR and low-pass filtered at 0.1 Hz using a zero-phase, 4th order Butterworth filter. ICA was performed using the InfoMax algorithm and components corresponding to the eye artifacts automatically detected with ICLabel were removed[50]. The resulting data were then averaged around the blinks (detected by **ABCD**). Figure 3–13 compares the BRPs used by ABCD (3–13A) and ICA (3–13B) for blink correction.
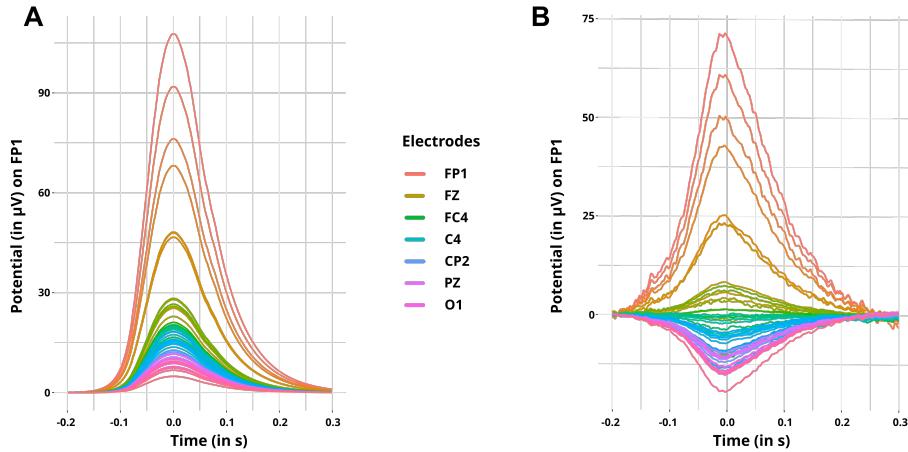
Figure 3–13 Comparison smoothed GM BRP and ICA BRP for one subject at one specific class for all electrodes

The smoothed GM BRP corresponding to a single blink, categorized according to its blink's peak value on the reference channel, is then stretched to fit the single blink as best as possible. Finally, the resulting adapted GM is removed for each single blink and compared to the same data corrected with ICA and ASR during (Figure 3–14A) and outside (Figure 3–14B) a blink.

In the presence of blinks in the raw data, the default settings of EEGLAB for both ASR and ICA do not effectively correct the signal, resulting in residual blink artifacts (albeit to a lesser extent for ICA). In contrast, **ABCD** performs well by successfully eliminating the impact of blinks on the corrected signal. When blinks are absent, **ABCD** preserves the original data, while ASR attempts to do so with less success. Conversely, ICA significantly alters the original signal.

3.7   Summary of Results and Discussion

The recorded data (see Chapter 2 for details about the experiment) is pre-processed to correct blinks with our newly developed **ABCD**. In parallel, the open-source MATLAB toolbox EEGLAB is used for the comparisons with ICA and ASR. The pre-processing recommendations provided in the online tutorial were adhered to, which involved applying a basic finite impulse response (FIR) filter at a cutoff frequency of 0.1 Hz and re-referencing the data using the average reference method.
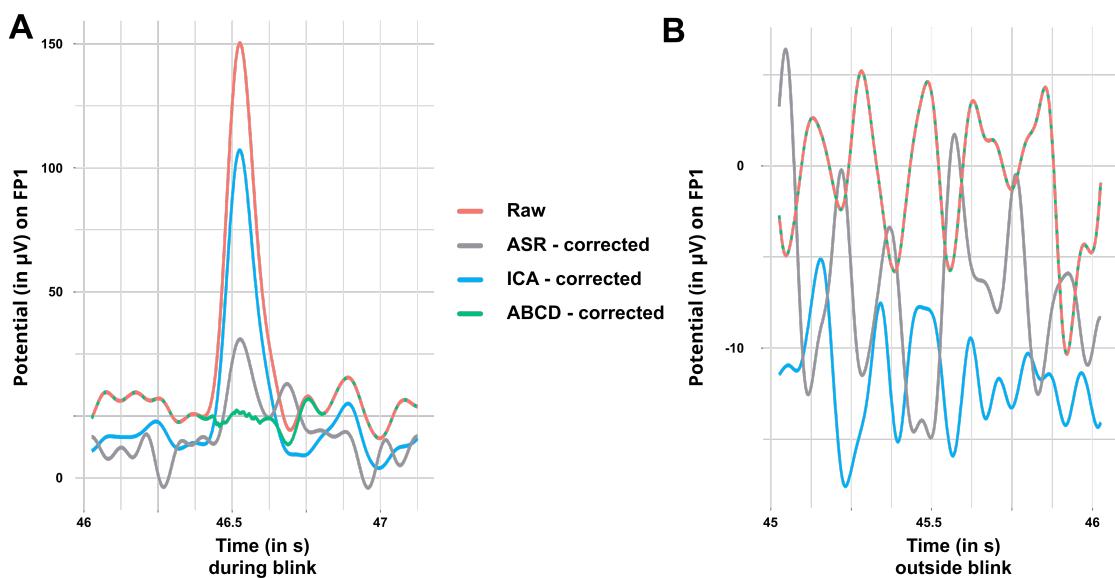
Figure 3–14 Comparison between raw and corrected data with ASR, ICA, and ABCD (A) during a blink; (B) and outside of any blink

In the case of ICA, the ICLabel plugin was employed to automatically detect and remove the labeled blink components. As for ASR, the default parameter values were retained when utilizing the Clean Rawdata plugin. It is important to note that prior to running ICA or ASR, the bad channels identified using **ABCD** were excluded. This decision was made based on the unsatisfactory results observed when using the actual raw data (i.e., all 62 original EEG electrodes).

The resulting cleaned data are then similarly processed and classified using the methods outlined in Chapter 4. The whole process is fully automated to avoid any user or choice bias and provide the best comparison as possible between these three pre-processing algorithms. The obtained classification accuracy is computed along with the confidence interval calculated from the equations described in Chapter 5. Table 3–4 presents a summary of these results for each subject and for the mean.

The confusion ellipses detailed in Chapter 5 are plotted for each pre-processing method of blink correction, revealing the superiority of **ABCD** over ICA (Figure 3–15).

A comprehensive representation of commonly employed metrics is provided in Figure 3–16, which presents a spider web visualization. This graphical depiction serves to enhance readability and facilitate further investigation. Additionally, the figure includes metric values to provide a quantitative summary of these metrics for reference and comparison purposes.

Cross-validation is a widely employed technique for evaluating the performance of machine learning algorithms and models. It involves partitioning the available data into distinct training and testing datasets to assess the model's generalization ability. However, it is important to note that conventional randomized cross-validation schemes may not always be suitable for ensuring the generalizability of processing meth-
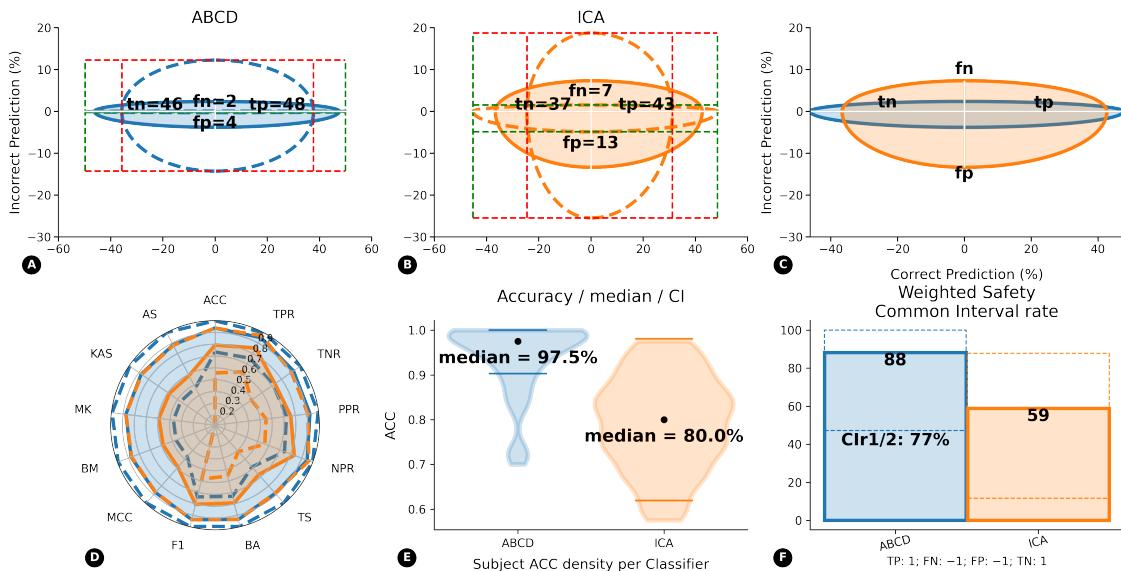
Figure 3–15 Confusion ellipses (CE) comparison between ABCD and ICA; (A) CE with corresponding confidence interval (CI) at confidence level = 95% for ABCD; (B) for ICA; (C) comparison CE without CI; (D) spider web comparison; (E) violin plots comparison for classification accuracy with median and CI; (F) weighted safety comparison
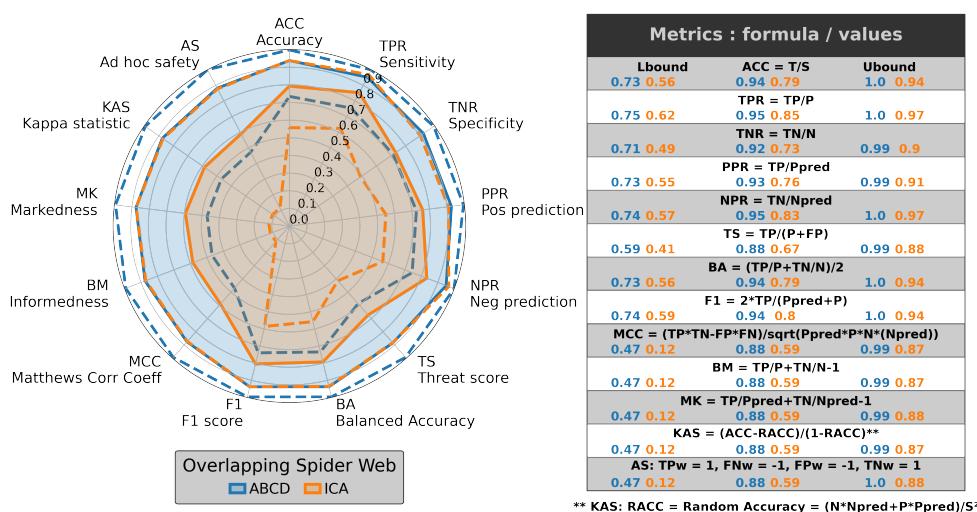
Figure 3–16 Spider web and associated metrics, comparison between ABCD and ICA

Table 3–4 Comparison of classification accuracies (fully automated pipeline) between the new ABCD and other existing methods, such as Independent Component Analysis (ICA) or Artifact Subspace Reconstruction (ASR), with corresponding confidence intervals (CI) at 95% confidence level

Subject	Classification accuracy (in %) with confidence interval (CI) at confidence level = 95%; CI approximation = Clopper-Pearson		
	ABCD	ICA	ASR
S01	98% [79%; 100%]	88% [66%; 98%]	93% [72%; 100%]
S02	89% [69%; 79%]	86% [64%; 97%]	88% [68%; 98%]
S03	100% [83%; 100%]	84% [62%; 95%]	80% [58%; 94%]
S04	89% [70%; 97%]	77% [56%; 92%]	83% [62%; 95%]
S05	99% [82%; 100%]	79% [58%; 93%]	80% [58%; 94%]
S06	95% [75%; 100%]	60% [36%; 81%]	100% [83%; 100%]
S07	97% [78%; 100%]	78% [56%; 92%]	85% [64%; 97%]
S08	96% [77%; 99%]	72% [48%; 89%]	84% [61%; 96%]
S09	87% [68%; 96%]	74% [52%; 90%]	93% [72%; 99%]
S10	91% [71%; 99%]	66% [43%; 85%]	78% [56%; 91%]
S11	85% [62%; 97%]	75% [51%; 91%]	65% [43%; 84%]
S12	98% [81%; 100%]	83% [62%; 96%]	83% [64%; 94%]
S13	92% [72%; 99%]	80% [58%; 93%]	89% [68%; 99%]
S14	98% [80%; 100%]	76% [53%; 92%]	80% [59%; 93%]
S15	88% [67%; 98%]	83% [61%; 96%]	89% [68%; 99%]
S16	95% [76%; 100%]	93% [74%; 100%]	86% [64%; 96%]
S17	95% [76%; 100%]	89% [68%; 98%]	80% [59%; 93%]
S18	97% [78%; 100%]	88% [66%; 98%]	77% [55%; 92%]
S19	90% [69%; 100%]	65% [41%; 85%]	85% [62%; 97%]
S20	95% [76%; 100%]	58% [34%; 78%]	95% [76%; 100%]
S21	70% [47%; 88%]	75% [52%; 91%]	85% [62%; 97%]
S22	100% [83%; 100%]	83% [60%; 96%]	75% [51%; 91%]
S23	94% [62%; 97%]	73% [50%; 89%]	88% [65%; 98%]
S24	100% [83%; 100%]	78% [57%; 93%]	95% [75%; 100%]
S25	93% [73%; 100%]	70% [47%; 87%]	90% [69%; 98%]
S26	73% [50%; 89%]	75% [55%; 91%]	68% [43%; 86%]
S27	95% [76%; 100%]	93% [72%; 100%]	78% [57%; 93%]
S28	100% [83%; 100%]	95% [76%; 100%]	93% [72%; 100%]
S29	100% [83%; 100%]	68% [44%; 86%]	88% [67%; 99%]
S30	100% [83%; 100%]	78% [54%; 93%]	98% [79%; 100%]
S31	100% [83%; 100%]	68% [45%; 86%]	88% [67%; 99%]
Mean	**93.81%** [74.81%; 98.76%]	**79.29%** [57.41%; 92.89%]	**84.05%** [62.88%; 95.31%]

ods. In our specific case, the utilization of such schemes is particularly ill-suited due to the training being conducted on a subset of each session, resulting in considerably smaller training and testing datasets.

As a result, we introduce a new **Consistency** metric aimed at addressing these limitations. This metric involves selecting smaller subsets of sessions through twenty successive random draws (instead of trials for the cross-validation approach). The algorithms are then executed on these subsets, and their results are averaged. This approach has the advantage of providing a more comprehensive understanding of how the compared methods would perform on a new dataset.

Table 3–5 confirms that **ABCD** consistently achieves higher classification accuracy compared to ICA and ASR. The differences between the classification accuracy of methods run on the entire dataset and the consistency can be attributed to the statistically less probable presence of outliers in the smaller subset of sessions.

Table 3–5 Consistency comparison of ABCD, ICA and ASR (20 successive random draws, corresponding to 6.3 sessions in average)

Methods	Classification accuracy (entire dataset)		Consistency accuracy (mean of 20 draws, corresponding to 6.3 sessions in average)	
	Mean	Standard Deviation (SD)	Mean	Standard Deviation (SD)
ABCD	93.81%	8.07%	93.92%	3.23%
ICA	79.29%	10.58%	81.70%	4.03%
ASR	84.05%	10.19%	84.71%	2.85%

Conceptually, ICA carries out the computations on a Riemannian submanifold and is thus subjected to its intrinsic practical limitations. Specifically, this requires finding a trade- off between a scale large enough to be above the noise level yet small enough to avoid curvature effects that would disprove the tangent plane approach. The partial solution brought by ICA is to use numerous samples. A commonly given rule of thumb is to use at least the number of channels squared times 20 (for less than 64 electrodes) or times 30 (for more than 64 sensors).

In the official tutorial of EEGLAB, ICA is clearly stated as working best for large amount of basically similar and mostly clean data. Yet, it is often used for blink correction, assuming that the resulting increased variability (during and outside of the blinks) in the resulting ICA-corrected data is of no consequence. Our results clearly indicate, on the contrary, that ICA should be avoided for blink correction.

Furthermore, the inherent variability that characterizes ICA can also be visualized in the results of the **Consistency** (in Table 3–5). Both the Standard Deviation (SD) of the classification accuracy for the entire dataset and for a smaller subset of sessions are higher for ICA compared to **ABCD**. This reveals the

larger sensitivity of ICA to any additional variability.

Similarly, ASR as a general artifact reduction method does not seem to be well equipped to deal with blink correction. By default, ASR rejects data regions if they exceed 20 times the standard deviation of the calibration data. An optimized choice of this parameter would probably have yielded better results, but to provide a fair comparison with ICA (for which all default values have been kept), we decided not to change the default settings. Therefore, our results indicate that, at least with the default values, ASR should be avoided for blink correction.

3.8   Conclusion and Highlights

We develop a novel method for accurate blink correction in EEG signals. The Adaptive Blink Correction and De-drifting (**ABCD**) algorithm is designed for lower computation complexity (real-time application), low hardware complexity (no need for additional sensors), and without systematic correction (only corrects when an artifact is detected).

The diverse features and specificities of this method can be outlined in a few essential ideas:

- **ABCD** integrates both stationary and non-stationary characteristics of the signal.
- **ABCD** detects blinks using three criteria: (1) moving average amplitude, (2) pre- and post-amplitude, (2) propagation.
- **ABCD** only requires two reference electrodes: a frontopolar one (e.g., FP1) for the first two criteria and a central one (e.g., CZ) for the third criterion.
- **ABCD** automatically identifies bad channels using blink-related potential (BRP).
- The analysis of BRP shows that frontopolar EEG, EOG, EMG, and upper eyelid displacement from video recordings all represent the same blinking signal at different levels and are thus redundant.
- Differences in inter-subject variability can be explained by physical characteristics, and intra-subject variability is linked to changes in the global eyelid displacement (the more the eyes close, the higher the amplitude).
- **ABCD** categorizes blinks into subject-specific 20 μV-width classes, with the smallest class starting at 35 μV and the highest one ending at 360 μV.
- **ABCD** reaches a mean classification accuracy of 93.81% [74.81%; 98.76%] (confidence interval at 95% confidence level), which is significantly better than ICA 79.29% [57.41%; 92.89%] and ASR 84.05% [62.88%; 95.31%].

Most studies devoted to pre-processing comparison try to find some characteristics either related to the channel or to the artifact(s) to estimate which algorithm works better. We argue that for most BCI systems, the primary goal is to accurately distinguish between tasks (e.g., Left vs. Right). Consequently, we opted to use classification accuracy as a metric to assess its efficacy. In this regard, **ABCD** largely outperformed both ICA and ASR.

To ascertain the generalizability of **ABCD**'s superiority, further benchmarking across diverse datasets is warranted. This evaluation should encompass not only MI tasks but also other cognitive tasks. It is worth noting that while **ABCD** has been specifically designed for real-time implementation, its efficacy in this context has yet to be empirically tested. Therefore, future research endeavors will be dedicated to achieving online blink correction using **ABCD**.

Chapter 4   Dual Classifier (DC) based on Source Localized Spatio-Temporal (SLST) Features Using Core Channel Selection (CCS)

This chapter's objective is to identify optimal algorithms for feature extraction and classification. A combination of various methods is implemented to extract the most prominent signal of interest (SOI) in the temporal, spatial, and frequential domains. The resulting covariance matrices characteristics are optimized by combining different strategies for re-referencing and band-pass filtering to enhance the quality and relevance of the extracted features.

Our feature extraction algorithm is based on the results of source localization to draw out the spatio-temporal characteristics of the SOI. The Source Localized Spatio-Temporal (**SLST**) features are identified through a similarity analysis across trials. A dual classifier (**DC**) is then applied to the values derived from the **SLST**, using both its spatial locations and time-derived covariance matrices on the Riemannian manifold. The class-related Fréchet mean as well as all other covariance matrices are computed at the Core Channel Selection (**CCS**) that is derived from a meta-analysis. Comparisons across different frequency bands (delta, theta, alpha, and beta), filter types (Butterworth, Chebyshev, and Elliptic), and re- referencing, e.g., Common Average Reference (CAR), Large and Small Laplace, are also carried out to reveal the optimized combination of all these processing steps.

Common Spatial Pattern (CSP) or covariance matrices are widely used for feature extraction. They are usually coupled with Linear Discriminant Analysis (LDA) or Support Vector Machine (SVM). More recently, methods such as the Minimum Distance to Riemannian Mean (MDRM) or Tangent Space (TS) have taken advantage of Riemannian geometry for covariance matrices distance computation. A comparison of these four already existing combinations of methods reveals the superiority of our method presented in this chapter using the newly introduced **Consistency** measure (with twenty successive random draws, corresponding to 6.3 sessions on average). Using similar pre-preprocessing (**ABCD**) and processing (Large Laplace for beta Butterworth filtering) steps, CSP+LDA obtains a mean classification accuracy of 89.16% (SD = 1.76%), CSP+SVM 88.99% (SD = 1.60%), MDRM 81.13% (SD = 4.64%), and TS 86.09% (SD = 4.48%). The method presented in this chapter **SLST+DC** shows a significant improvement, yielding a mean classification accuracy of 95.03% (SD = 3.41%). Furthermore, the associated CI's lower bound for our method is consistently higher than all the other methods CI's upper bounds. This confirms that our method performs significantly better than all the other tested algorithms.

Our **SLST+DC** method is however limited in its real-time applications because of its intrinsic dependency on source localization. Such methods being very slow to compute, a proxy will need to be found to obtain similar temporal and spatial features as **SLST** or derived results. Yet, these results are of great

interest since more information can be learned about MI that either might be applicable for later online use or could be used for analysis that does not require real-time use (e.g., diagnosis).

4.1   Introduction

The primary goal of a BCI pipeline is to accurately identify the Signal of Interest (SOI) by optimizing the three inherent features of EEG signals in the temporal, spatial, and frequential domains. Various strategies have been developed to achieve this objective. Some approaches focus on utilizing statistical measures directly from the raw EEG signal within a specific time window, allowing for the computation of relevant metrics. Alternative strategies involve extracting more intricate and sophisticated representations from the EEG signal to capture the underlying patterns and characteristics.

4.1.1   Common Algorithms for Feature Extraction and Classification

A widely utilized technique for feature extraction in EEG analysis is the Common Spatial Patterns (CSP) method. This approach can be used to derive spatial patterns or topographical information from the EEG data. The fundamental principle of CSP is rooted in the assumption that different mental states or tasks are linked to unique patterns of brain activity, which can be effectively captured by specific spatial filters. These filters are designed to enhance the discriminative information present in the EEG signals by amplifying the differences between the classes of interest (e.g., Left vs. Right MI).

On a training dataset, the covariance matrices of the EEG signals are computed for each class and decomposed into their eigenvalues. The eigenvectors represent the spatial filters, while the eigenvalues indicate the amount of variance captured by each spatial filter. To identify the most discriminative spatial filters, the spatial filters are sorted based on their corresponding eigenvalues. During testing, new EEG signals are projected onto the selected top-ranked spatial filters, usually a subset of the most informative filters (typically four filters). The extracted features can then be used as input to a classifier, such as linear discriminant analysis (LDA) or support vector machines (SVM), to classify the new EEG signals into their respective classes.

Newer algorithms in EEG analysis have embraced the principles of Riemannian geometry, which offers a mathematical framework for studying the properties of covariance matrices. These algorithms operate on the premise that covariance matrices can effectively capture the statistical relationships among various electrode locations. The Riemannian mean, which represents the central tendency of a collection of covariance matrices, is computed for each class labeled in the training data. The minimum distance to the Riemannian mean (MDM) approach aims to classify EEG signals by calculating the distance between a test covariance matrix and the Riemannian mean of each class.

An analogous algorithm seeks to exploit the geometric properties of these covariance matrices by mapping them onto a tangent space (TS), enabling more efficient linear operations. The tangent space

mapping can be carried out for each class's Riemannian mean. Subsequently, a new test covariance matrix is projected onto each TS, and its distance to the corresponding mean is then computed.

4.1.2 Source Localization

We posit that the complex, intricate characteristics of EEG signals necessitate more than a simple conversion to frequency or covariance representations to unveil their underlying trends and, consequently, meaningful SOI. We further assume that source localization techniques have the ability to reveal such SOI, although different techniques all have their inherent limitations.

Source localization can estimate the spatial locations of neural sources within the brain that generate the recorded EEG signals. Its primary objective is to reconstruct the underlying neural activity and gain insights into the spatial distribution of brain activation. The process of source localization involves solving the inverse problem (i.e., estimating the neural sources given the measured EEG signals). This is a challenging task due to the ill-posed nature of the problem, where multiple possible source configurations can explain the same EEG data.

The eLORETA (exact Low-Resolution Brain Electromagnetic Tomography) method is based on the premise that the electrical activity originating from neural sources can be represented as a distribution of current density throughout the brain. This approach employs a distributed source model, which assumes that the neural sources are spread throughout the brain rather than being localized at specific points. One notable advantage of eLORETA is its ability to handle multiple simultaneously active sources, which is often the case in EEG recordings. By accurately estimating the locations and strengths of these sources, eLORETA provides valuable insights into the underlying dynamics of neural activity. Implementations of the eLORETA method can be found in MNE[51] and LORETA-KEY[5].

4.2 Full Meta-Analysis: Core Channel Selection (CCS)

One of the primary purposes of a meta-analysis is to draw conclusions from past studies, often reporting disparate results. Pooling independent effect size estimates can inform us about the direction and magnitude of an effect even more accurately than any of the individual estimates. Non-systematic biases that may arise, for instance, from pre-processing steps, are assumed to become insignificant using this approach[52].

4.2.1 PRISMA Flowchart

The review is based on the PRISMA statement (preferred reporting items for systematic reviews and meta-analyses) and the recommendations of the Cochrane Collaboration[53-54]. Fully published articles were drawn from searches on Google Scholar in September 2020. English-language studies in which the

subjects were asked to perform either motor imagery (MI), motor execution (ME), or to focus on SSVEP or P300 were included. This literature search yielded 117 articles.
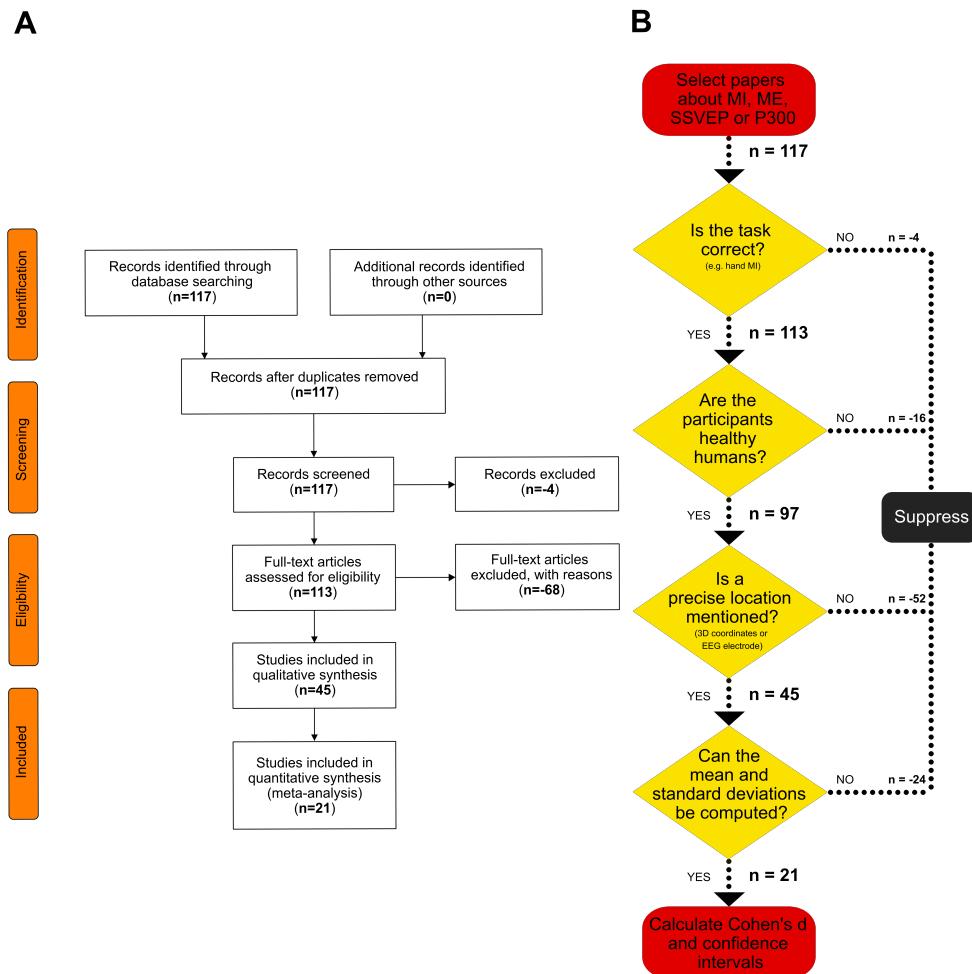
Figure 4–1 (A): Standard PRISMA flowchart diagram; (B): details of the exclusion criteria used with the associated number of studies (n)

All neuroimaging techniques were incorporated as long as the subjects were healthy human participants performing the correct task (e.g., only hand movement was considered for MI and ME, foot motion was rejected). 97 articles met these further requirements. The quality of the data was the most exclusive criteria. Studies that did not specify the exact location (3D coordinates from any atlas or EEG electrodes) of the brain structures involved in the task or did not provide enough data to compute the mean and standard error or deviation of these locations were excluded from further analysis. After applying these criteria, 21 articles remained. Figure 4–1A reproduces the standard PRISMA flowchart diagram, while Figure 4–1B highlights the precise number of papers excluded for each exclusion criterion.

The main purpose of this review is to create a knowledge-based channel selection framework with

sufficient statistical power and practical significance. The latter can be established by quantifying the effect size. To compute all these measures, either the Z-score, the t-test, the Cohen's r, or the Cohen's q should be available, along with the corresponding standard error and deviation when necessary. The lack of reporting of such measures (or access to the original data) is a well-known issue in neuroscience [8] and might explain the generally small numbers of included studies in meta-analyses ($n = 9$)[55], ($n = 30$)[56], ($n = 9$)[57]).

The metadata of the twenty-one articles ($n = 21$) included in this meta-analysis is detailed in Table 1 in Appendix A. In total, 311 healthy subjects participated in these studies, with a highly variable sample size ranging from 5 to 50 participants ($mean = 14.8 \pm 11.1$). Every corresponding author of the twenty-one studies has been contacted and solicited to provide the necessary information for computing the effect size. Most effects are reported as group differences, thus Cohen's d has been favored in this review[21]. For all studies, the mean across subjects (computed from Z-score, t-test, percentage of signal change, or raw data) was calculated on every selected electrode. For each task and each electrode, the mean effect size was computed and weighted by the number of subjects in the individual studies.

4.2.2 Sample Size Estimation: Number of Studies

How many studies are necessary to obtain sufficient statistical power? Conventionally, in meta-analysis methods, fixed effects and random effects models are used. In the former case, the included studies are assumed to represent the entire universe of studies of interest, and hence, there is no possibility of sampling error[58]. In view of the high variability of the phenomena under study, the included studies are considered to only represent a small portion of all the available data, and thus, the slightly more sophisticated random effects model is preferred.

As usual, the Type I error rate is set at $\alpha = 0.05$ and the relative seriousness of Type I to Type II error is 4 to 1 ($\beta = 0.2$ and thus $power = 0.8$). The standard normal cumulative distribution function for $\beta = 0.2$ yields a value of -0.842, and a mean of the Z statistic of $1.64 - (-0.842) = 2.482$ for a one-tailed test.

Under the random effects hypothesis, the Z statistic has a normal distribution with a mean equal to $\lambda = \frac{\overline{d} - 0}{\sqrt{v^*/k}}$, where k is the number of studies (random effects estimates are usually noted with an asterisk[59]). The "typical" sampling variance of the random effects estimate is calculated with $v^* = v + \tau^2 = \frac{2}{n} + \frac{\overline{d}^2}{4} * n + \tau^2$[60]. Given a random effects analysis, the number of studies should then be $k = \frac{v^*}{(\overline{d}/\lambda^*)^2}$, where $\lambda^* = 2.482$.

A scoping review on different electrodes led to the belief that the estimated population effect size may vary from $\overline{d} = 1.5$ to $\overline{d} = 3$ with a variance of $\tau^2 = 1$ and for an average number of subjects per study of $n = 15$. For an estimated population effect size of $\overline{d} = 1.5$, approximately four studies ($k = 3.2$) are needed under the random effects assumptions to obtain an approximate power of 0.80. For a greater estimated effect size of $\overline{d} = 3$, only one study ($k = 0.88$) is needed to obtain sufficient power, with all

other hypotheses remaining equal.

4.2.3 EEG Channel Determination

Multichannel EEG measures signals using either the 10-20, 10-10, or 10-5 international systems for standard electrode placement[61]. The goal of the BCI paradigms is to produce distinct patterns of cortical activation to discriminate between various classes (e.g., left- or right-hand movement)[62]. The various studies reviewed here use source localization to identify significant cortical generators. The closest of the 62 selected EEG electrodes are then estimated for each of the results obtained from fMRI or PET. Our goal is to provide a knowledge-based channel selection framework that could be used either as a verification step or as part of the classification procedures.

The Koessler 3D anatomical atlas[63] was used to correlate the 3D Talairach coordinates of the EEG electrode of the 10-10 system to their cortical projection points. For convenience, the Koessler 3D anatomical atlas is available as a csv file at https://github.com/QinXinlan/review-effect-size. This 10-10 resolution combines the advantages of higher precision (yet still providing background compatibility with the 10-20) and reproducibility (by avoiding overlapping that may arise with the 10-5). The likely Brodmann area (BA) (respectively, macro-anatomical structure) beneath each sensor is then calculated as the most frequently found in the population.

Using the mean 3D Talairach coordinates and standard deviations of each sensor, it is possible to compute the Euclidian distance to each cortical generator reported in the reviewed articles. Standard deviations in x, y, and z play asymmetric roles, so the most likely electrode is not necessarily the closest one. When two EEG electrodes are nearly equidistant from a cortical source, both are recorded in the following tables, with the electrode at the minimal Euclidian distance in the first position.

This review and its potential applications rely on three assumptions: (1) despite the substantial inherent variability in the exact source localization of neural generators, the sensorimotor and visual networks are postulated to be consistent across subjects and sessions, and the other cortical networks relevant to the four common BCI paradigms should probably be considered more subject-dependent; (2) the estimated EEG electrodes or their nearest neighbors are likely to reflect most of the relevant cortical information; (3) the classification accuracy is presumed to be optimized by the choice of an adequate subset of electrodes.

4.2.4 Now You See It: Motor Execution

A voluntary movement is comprised of three phases: planning, execution, and recovery. During the first phase, similar functional circuits, located in the fronto-parietal, subcortical, and cerebellar areas, are activated in both ME and MI, suggesting they share a common pattern in the planning and preparation tasks. However, MI also activates distinct regions that can be predominantly found in the left hemisphere, namely the middle temporal gyrus and the fusiform gyrus (BA 21 and 37, electrodes TP7, P7, and T7).

Table 4–1 3D Talairach coordinates of significant cluster maxima for motor execution of left and right hand Movement; ++ specific to ME; left underline for left-handed subjects; right underline for right-handed subjects; NP: not pertinent; BA: Brodmann area; Cohen's d mean effect size with a 95% confidence interval and the number of relevant studies in superscript

Lobe	Anatomical Area	Area	Macro-anatomical structure	Laterality	BA	Hem	Talairach			Closest electrode(s)	Cohen's d Weighted Effect size
							X	Y	Z		
Frontal	Prefrontal cortex	Inferior frontal	Inferior frontal gyrus	Unilat.	44, 45	R	57	18	24	F8	3.05 [1.93; 4.15][2]
						L	-49	5	16	FC5	2.47 [1.51; 3.41][1]
	Premotor and motor cortex	Lateral pre-motor	Superior frontal gyrus	Contralat.	6	R	42	-3	60	FC4, FC2	2.44 [1.84; 3.03][4]
						L	-36	-3	66	FC1, FC3	2.52 [1.77; 3.27][3]
		Medial premotor Pre-SMA		Bilat.	6	R	6	6	57	FCZ	3.7 [2.32; 5.09][1]
						L	-9	9	51	FCZ	3.7 [2.32; 5.09][1]
		Post-SMA		Bilat.	6	R	3	-3	66	FCZ	3.7 [2.32; 5.09][1]
						L	-3	-3	66	FCZ	3.7 [2.32; 5.09][1]
		Primary motor (M1)	Precentral gyrus	Contralat.	4	R	39	-15	63	C2	1.64 [1.2; 2.27][5]
						L	-42	-18	66	C1	1.75 [1.29; 2.54][6]
Parietal	Parietal cortex	Primary sensory (S1)	Postcentral gyrus	Contralat.	1, 2, 3	R	42	-33	60	C4	1.6 [1.15; 2.26][4]
						L	-48	-24	57	C3	1.62 [1.16; 2.42][5]
				Bilat.	7, 4	R	4	-40	38	CPZ, CZ	0.78 [0.59; 0.97][1]
		Superior parietal		Bilat.	7	R	36	-36	66	CP2	1.15 [0.59; 1.72][1]
						L	-30	-42	66	CP1	1.06 [0.49; 1.62][1]
		Inferior parietal	Anterior intraparietal sulcus (AIP)	Bilat.	40	R	42	-45	48	CP4	0.9 [0.62; 1.18][1]
						L	-45	-36	60	CP3	1.12 [0.72; 1.53][2]
		Secondary sensory (S2) ++	Parietal operculum	Bilat.	40, 43	R	60	-9	15	CP6	1.14 [0.57; 1.71][1]
						L	-60	-15	21	CP5	1.13 [0.57; 1.70][1]
Subcortical regions	Basal ganglia	Putamen Anterior part		Bilat.		R	24	18	6	NP	
						L	-30	6	6	NP	
		Posterior part ++		Bilat.		R	27	-9	12	NP	
						L	-33	-3	6	NP	
	Thalamus			Bilat.		R	15	-12	9	NP	
						L	-18	-15	9	NP	
Cerebellum	Rostral part			Bilat.		R	21	-45	-18	NP	
						L	-27	-45	-18	NP	

The predicted visual consequences of an action are believed to be represented in the former gyrus[64], whereas lesions in the latter impair pantomime recognition[65]. Thus, these regions could be hypothesized to participate in the planning of an imaginary movement. Furthermore, lesions in the superior and inferior parietal cortex (BA 7 and 40; electrodes CP4, P1, and P2) lead to impairment in the ability to imagine movements. Therefore, these cortices could be responsible for ensuring that the action being executed matches the intended action[66].

During ME, the secondary sensory area (S2, BA 43; electrode CP6) seems to display heightened activation and is assumed to be related to the processing of proprioceptive input[67]. More specifically, action planning has been associated with an increase in activity in the left-lateralized supramarginal gyrus[68] (BA 40; electrode CP5). A possible explanation of these differences between ME and MI relies on the evolutionary theory of motor task learning. Mentally rehearsing the action would allow for multiple attempts without the risk of causing any harm and, hence, would help find the best strategy for a difficult motor task[69]. Some studies[69-72] compared the somatotopic mapping during ME and MI, while others[73-74] examined the effects of handedness. Their results are showed in Table 4–1.

Figure 4–2 Motor execution paradigm: Activated electrodes with their Brodmann Areas and handedness-linked brain laterality

For the ME task, only two clusters have been identified, as illustrated in Figure 4–2 – either non-contralateral (bilateral or unilateral) or contralateral. The ME topographical distribution, as expected, largely overlaps the MI one. However, the assumption that MI is a more complex cognitive task might explain why the number of cortical generators activated with ME is much smaller for a simple task[75]. Interestingly, mental chronometry studies have shown that the time course of MI positively correlates with ME, illustrating the parallelism between these two tasks[76].

For right-handed subjects (an estimated 90% of the worldwide population), right hand ME and MI activate a restricted cluster compared to left hand ME and MI[72]. Right- handed subjects also show a predominant activation of motor and visuomotor control in the contralateral (i.e., left) hemisphere, while left-handed subjects have been characterized by a more bi-hemispheric recruitment of neuronal circuits when performing a precision grasping task. This might indicate a left hemispheric location of visuomotor control[73]. In general, the more repetitive the task is, the more specialized and smaller the neural circuit will be. Therefore, the nondominant hand, being used less often, would require additional control provided by a more widespread neural activation. Finally, it should be noted that the complexity of the task, implying other functions, influences the lateralization, and the level of activation in some regions is age-dependent[77].

For the ME task, the **CCS** subset consists of electrodes FC4, FC2, FC1, FC3, C2, C1, C4, C3.

4.2.5 Now You Don't: Motor Imagery

MI-based BCI is a popular interaction paradigm because it relies on voluntary movement and can be used, in theory, by healthy and impaired subjects alike in synchronous or asynchronous paradigms. In particular, the movement of the right and left hands is often selected because of its easily distinguishable activity in the contralateral cortical regions responsible for the movement of the limbs.

MI is usually defined as the mental rehearsal of a motor act[78]. This cognitive process appears to correspond to the activation of the neural correlates of motor representations[79]. It includes the planning and preparation of movements but not the motor output or somatosensory feedback. The neural network involved during MI activates several cortical areas depending on the task and on the participant's handedness and familiarity with BCI and the task. Several studies[69-72][80-82] have described the mapping of cerebral networks by localizing the main sources in each functional area across participants. The Table 4–2 compiles these results with similar inclusion criteria (i.e., for healthy subjects).

Table 4–2 3D Talairach coordinates of significant cluster maxima for motor imagery of left and right hand movements with corresponding international 10-10 EEG system closest electrode(s); BA: Brodmann area; Hem: hemisphere; L: left; R: right; NP: not pertinent; * for BCI-trained subjects, ** for BCI-naïve subjects; ++ specific to MI; left underline for left-handed subject's dominant hand; right underline for right- handed subject's dominant hand; and Cohen's d mean effect size with a 95% confidence interval and the number of relevant studies in superscript

Lobe	Anat. Area	Area	Macro-anatomical structure	Laterality	BA	Hem	Talairach X	Y	Z	Closest elec.	Cohen's d Weighted Effect size
Frontal	Prefrontal cortex	Rostral prefrontal	Superior frontal gyrus	Unilat.	10, 11	R	30	51	24	FP2	3.24 [1.85; 4.62][1]
		Dorsolateral prefrontal **	Middle frontal gyrus	Contralat.	6	L	-20	18	59	F1	1.25 [0.34; 2.15][2]
					9,8, 46	L	-54	42	15	F5, F3	2.79 [1.28; 4.28][2]
		Inferior frontal	Inferior frontal gyrus	Bilat.	44, 45	R	48	24	6	F8	3.36 [1.97; 4.74][1]
						L	-54	12	12	F7	3.05 [1.66; 4.43][1]
	Premotor and motor cortex	Lateral premotor	Middle frontal gyrus	Contralat.	6	R	42	6	57	FC4, FC2	1.28 [0.6; 1.95][5]
						L	-42	3	51	FC3	3.31 [1.82; 4.76][2]
		Medial premotor	Pre-SMA	Superior frontal gyrus	Bilat.	6	R	3	6	69	FCZ	4.8 [2.91; 6.66][2]
						L	-3	6	69	FCZ	4.8 [2.91; 6.66][2]
			Post-SMA		Bilat.	6	R	20	-7	44	FC2	1.57 [0.68; 2.45][2]
						L	-6	-3	69	FC1	1.37 [0.68; 2.06][5]
		Primary motor (M1) *	Precentral gyrus	Contralat.	4	R	18	-22	54	C2	1.66 [0.69; 3.07][2]
						L	-19	-15	-58	C1	3.39 [1.61; 5.11][1]

Continued on next page

Table 4–2 (continued)

Lobe	Anat. Area	Area	Macro-anatomical structure	Laterality	BA	Hem	Talairach			Closest elec.	Cohen's d Weighted Effect size
							X	Y	Z		
Parietal	Parietal cortex	Primary sensory (S1)	Postcentral gyrus	Contralat.	1, 2, 3	R	54	-21	39	C4, C6	1.26 [0.36; 2.28][4]
						L	-60	-21	39	C5, C3	2.29 [1.28; 3.48][3]
		Superior parietal ++	Precuneus	Bilat.	7	R	24	-54	51	P2, CP2	2.55 [1.16; 3.93][1]
						L	-27	-60	54	P1	3.03 [1.54; 4.49][2]
		Inferior parietal ** ++	Inferior parietal gyrus	Bilat.	40	R	36	-30	42	P2, CP4	1.47 [0.58; 2.35][2]
						L	-36	-42	48	CP3, CP1	1.38 [0.37; 2.39][4]
Temporal		Middle temporal	Middle temporal gyrus	Unilat.	21	L	-60	-57	6	TP7, T7	3.29 [1.9; 4.67][1]
		Inferior temporal	Inferior temporal gyrus	Bilat.	19	R	56	-60	5	P8	1.55 [0.73; 2.36][1]
					19, 37	L	-55	-65	6	P7	0.73 [−0.01; 1.46][1]
Occipital		Superior occipital	Occipito-parietal sulcus	Unilat.	19	L	-25	-61	33	PO3	2.61 [1.05; 4.11][1]
		Middle occipital	Calcarine gyrus	Bilat.	17, 18	R	12	-84	6	O2	2.01 [0.7; 3.33][1]
						L	-12	-88	-2	O1	2.54 [1.08; 3.99][1]
Sub-cortical regions	Caudate nucleus			Bilat.		R	15	9	15	NP	
						L	-18	9	21	NP	
	Basal ganglia	Putamen Anterior		Bilat.		R	27	9	12	NP	
						L	-27	6	12	NP	
	Insula **			Bilat.	13						

Table 4–2 (continued)

Lobe	Anat. Area	Area	Macro-anatomical structure	Laterality	BA	Hem	Talairach			Closest elec.	Cohen's d Weighted Effect size
							X	Y	Z		
	Thalamus	Ventral lateral nucleus		Contralat.							
Cerebellum				Ipsilat.							

The estimated EEG electrodes are then classified into three main clusters, as shown in Figure 4–3. The first one gathers all electrodes that are not contralateral (i.e., either bilateral or unilateral). The second cluster represents all contralateral sources that depend on the user experience (familiarity with BCI). The third one corresponds to all contralateral sources consistently present among subjects. Ipsilateral clusters can also be activated alongside contralateral sources to a lesser extent[83]. If an electrode may be classified as both contralateral and unilateral, the bilaterality may be more influential, and the electrode is labeled as such. Finally, the effects of handedness for the dominant hand on electrode activation are emphasized with an upper black point. For instance, FC2 is mostly activated for left-handed participants when moving their left dominant hand.
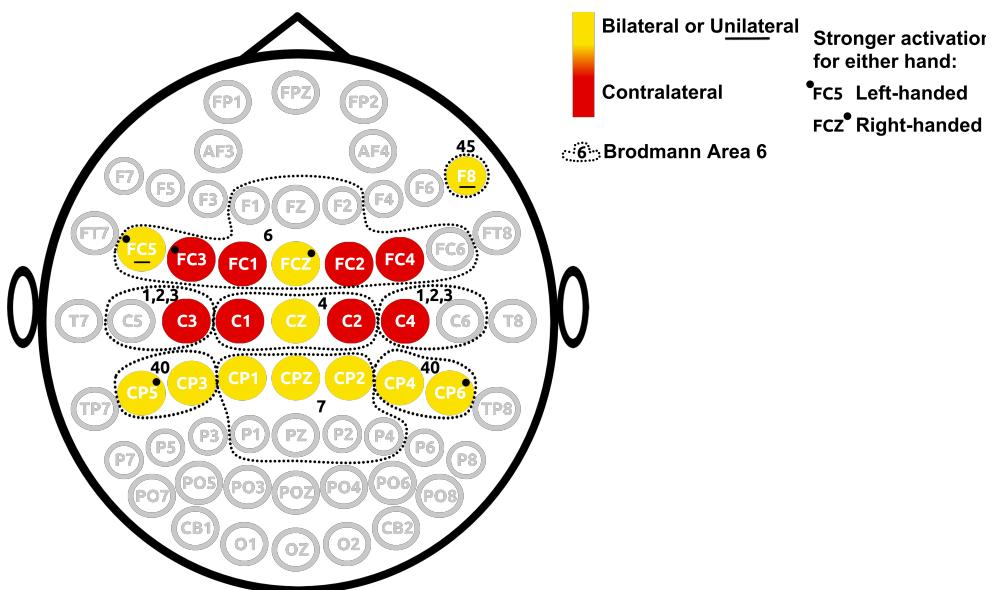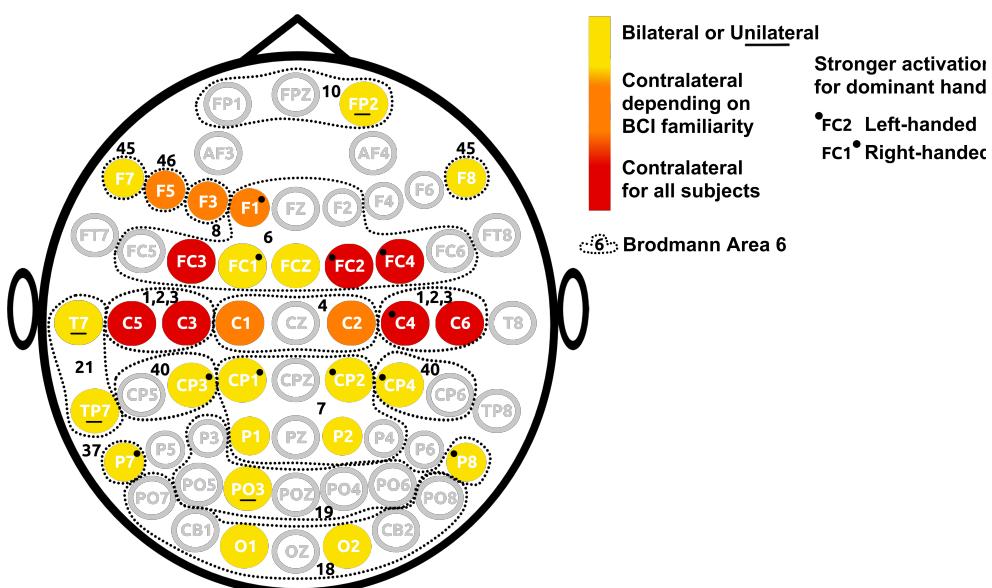

Figure 4–3 Motor imagery paradigm: Activated electrodes with their Brodmann Areas and handedness-linked brain laterality

For the MI task, the **CCS** subset consists of electrodes FC4, FC2, FC1, C4, C6, C5, C3, CP3, CP1. A common **CCS** for all motor-related tasks (MI and ME) could also be defined with the electrodes FC1, FC2, FC4, C3, C4.

4.2.6 The Way of the SSVEP

Changes in the visual field are known to impact EEG activity. In particular, the periodic contrast or luminance modulation of a fixed frequency (usually at a range of 6-40 Hz) elicits an SSVEP that can be detected at the same fundamental frequency as the flickering stimulus. The cortical oscillations are phase-locked to the periodic stimulus and appear to be predominantly present in posterior occipital areas as well as the lateral geniculate nucleus and optic radiation[84]. The two major neural generators seem to be localized in the primary visual cortex (V1/V2) and the motion-sensitive (MT/V5) areas, respectively. However, the two minor contributors seem to be located in the mid-occipital (V3A) and ventral occipital (V4/V8) areas[85].

Table 4–3 Activated electrodes depending on the SSVEP stimulus frequency

10 Hz	11 Hz	12 Hz	13 Hz
CZ	P1	F1	F6
CPZ	PZ	FZ	F8
P1	P2	P3	FC6
PZ	PO3	P1	FT8
P2	POZ	PZ	P3
PO3	PO4	P2	P1
POZ	O1	P4	PZ
PO4	OZ	PO3	P2
CB1	O2	POZ	PO5
O1		PO4	PO3
OZ		PO5	POZ
O2		PO6	PO4
		CB1	PO6
		O1	PO8
		OZ	CB1
		O2	O1
			OZ
			O2
			CB2

In addition to the occipital areas, steady-state responses may also flicker over the frontal and prefrontal

areas depending on the stimulus frequency[86]. Furthermore, the narrow frequency bands surrounding each flicker frequency appear to synchronize different patterns of cortical functional networks. Table 4–3 shows the EEG electrodes calculated from the topographical maps of the potential SSVEP power at the frequencies relevant to one of our previous experiment[87-88].

SSVEP-based BCIs often have a high temporal and spectral resolution (usually less than 0.1 Hz) as well as a high accuracy and information transfer rate (ITR)[89] but can produce visual fatigue or discomfort[90]. The temporal frequency, spatial frequency, contrast, luminance, color, and hue of the driving flickering stimulus all influence the amplitude and phase of the SSVEP[90-91]. This dependency on the input frequency partly reflects the delay between the retina and primary visual cortex as well as the delays between areas of the visual system[92]. Moreover, the area of on-screen stimulus also has an impact on cortical modulation. For instance, a 6 Hz frequency sinusoidal waveform displayed in the upper left quadrant of the screen will elicit a different SSVEP waveform than a similar stimulation displayed in the lower left quadrant (see Table 4–4)[85,93].

The review of five previous studies[85-87,94-95] is summarized in Figure 4–4. Three different types of information are illustrated using distinct systems. The color clusters illustrate the differences between contralateral and unilateral or bilateral sources. The corner symbols depict the activations that depend on the on-screen quadrant location (upper or lower, left or right). Finally, the text colors highlight whether the EEG electrodes are activated for some or all frequencies relevant to our experiment.

Table 4–4 3D Talairach coordinates of the two significant cluster maxima for steady-state visual potentials depending on the visual stimulus location on the screen. The available data did not allow calculations of the Cohen's d effect size per electrode.

Quadrant Location	Lobe	Macro-anatomical structure	Laterality	BA	Hem	Talairach			Closest electrode(s)
						X	Y	Z	
Upper left	Right Occipital	Middle occipital gyrus	Bilateral	18	R	13	-92	-9	O2, OZ
		Superior occipital gyrus	Contralateral	19	R	36	-71	-4	PO4
Lower left	Median Parieto-Occipital	Superior parietal Lobule, Cuneus	Bilateral	7, 19	R	5	-74	2	PZ, POZ
	Right Parieto-Occipital	Inferior parietal lobule, Superior occipital gyrus	Contralateral	7, 19	R	38	-65	0	P4, PO4
Upper right	Median Occipital	Cuneus	Bilateral	19	L	-9	-86	-7	POZ, O1
	Left Occipital	Superior occipital gyrus	Contralateral	19	L	-35	-73	-5	PO3
Lower right	Left Occipital	Cuneus	Bilateral	19	L	-6	-80	3	POZ
		Superior occipital gyrus	Contralateral	19	L	-31	-64	8	PO3, PO5

For the SSVEP task, the **CCS** subset consists of electrodes P1, PZ, P2, PO3, POZ, PO4, O1, OZ, O2.
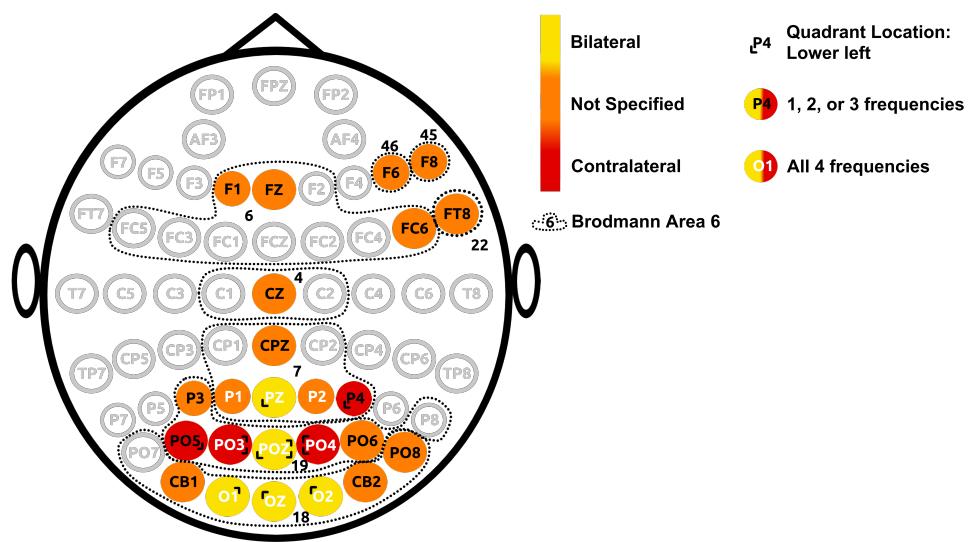
Figure 4–4 Steady-state visual evoked potential paradigm: Activated electrodes with their Brodmann areas and brain laterality linked to quadrant location or full-screen frequency

4.2.7   Raise the P300 Lantern

Attention to a change in the environment elicits a burst of activity, peaking at about 300 ms after the sensory stimulus. This P300 event-related potential (ERP) is an umbrella term encompassing two separate attentional processes.
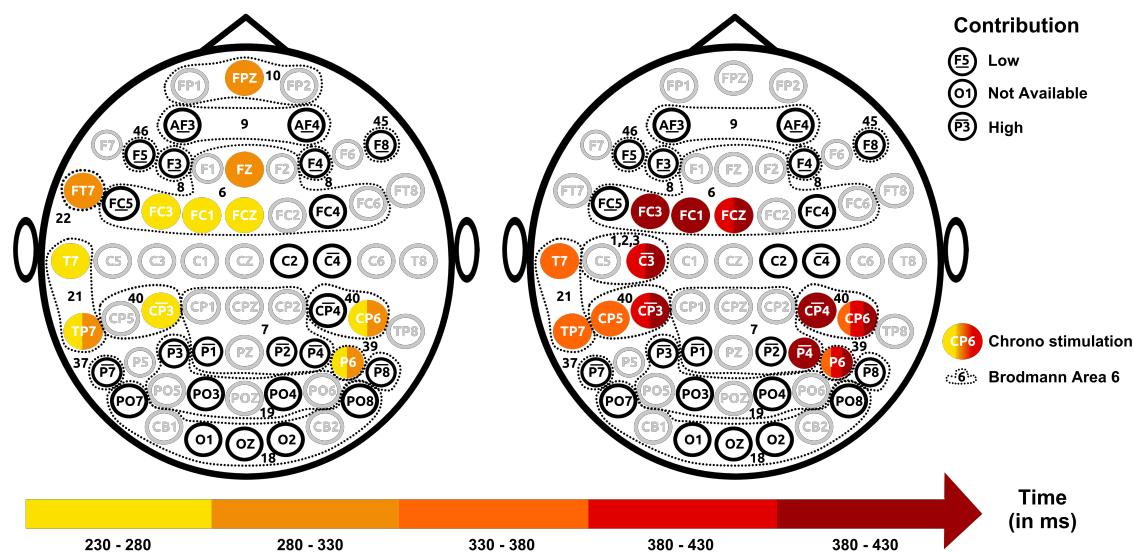
Figure 4–5 P3b component of the P300 paradigm: Electrodes' level and chronology of activation with corresponding Brodmann areas

Table 4–5 3D Talairach coordinates of significant cluster maxima for the P3b component of the P300 event-related potential; NP: not pertinent; BA: Brodmann area; Cohen's d mean effect size with a 95% confidence interval and the number of relevant studies in superscript

Time (in ms)	Contri-bution	Lobe	Anat. Area	Area	Macro-anat. struct.	BA	Talairach X	Talairach Y	Talairach Z	Closest elec.	Cohen's d Weighted Effect size
	Low	Front.	Pre-front. cort.	Inf. front.	Inf. front. gyr.	9	-64	19	15	F5, FC5	3.27 [1.67; 4.87][1]
				Sup. front.	Sup. front. gyr.	8, 9	-39	42	39	F3, AF3	3.48 [1.65; 5.57][1]
							40	43	36	F4, AF4	4.4 [2.48; 6.4][1]
					Pre-cent. gyr.	44	44	17	6	F8	3.27 [1.67; 4.87][1]
230 - 280	430 - 480				Mid. front. gyr.	6	-24	-4	57	FC1, FC3	NA
	NA		Pre-motor and motor cort.	Dorso-lat. pre-front. gyr.		6, 9	32	0	56	FC4	2.86 [1.26; 4.46][1]
280 - 330						11	-3	38	-13	FZ, FPZ	NA
230 - 280	380 - 430			Medial pre-motor SMA	Medial frontal gyrus	32, 6	-3	10	46	FCZ	3.45 [1.82; 5.47][2]
	430 - 480			Cing. cort.	Cing. gyr.	6, 24	-1	3	56	FCZ	3.45 [1.82; 5.47][2]

Table 4–5 (continued)

Time (in ms)	Contribution	Lobe	Anat. Area	Area	Macro-anat. struct.	BA	Talairach			Closest elec.	Cohen's d Weighted Effect size
							X	Y	Z		
230 - 280				Prim. motor (M1)	Pre-central gyrus	6	24	-16	52	C2	2.44 [1.13; 3.75][1]
	380 - 430			Prim. sens. (S1)	Post-cent. gyr.	1,2 3	-47	-24	53	C3	3.13 [1.8; 4.46][2]
	430 - 480			Inf. pariet. lob.	Supra-marg. gyr.	40	-47	-45	43	CP3	3.3 [1.82; 4.95][3]
	430 - 480						48	-39	38	CP4, P4	4.78 [3.31; 6.44][3]
				Post. pariet. cort.	Sup. pariet.	Sup. pariet. lob.	7, 19	-30	-73	44	P1, P3	4.28 [2.85; 5.97][3]
						7	30	-74	44	P2, P4	3.98 [2.52; 5.51][4]
230 - 280	High			Inf. temp.	Fusif. gyr.	19, 37	-38	-60	-16	P7	4.97 [3.53; 6.83][2]
330 - 380							29	-52	-14	P8	9.68 [7.88; 11.8][1]
380 - 430		Temp.		Mid. temp.	Mid. temp. gyr.	21	-59	-18	-13	T7, TP7	NA
380 - 430	NA					37	-44	-63	-1	P7, PO7	4.97 [3.53; 6.83][2]
							41	-58	-4	PO8	2.85 [1.54; 4.15][1]
				Sup. temp.	Sup. temp. sulc.	40, 39	52	-44	14	CP4, P6	4.87 [3.44; 6.56][3]

Table 4–5 (continued)

Time (in ms)						Contri-bution	Lobe	Anat. Area	Area	Macro-anat. struct.	BA	Talairach			Closest elec.	Cohen's d Weighted Effect size
												X	Y	Z		
	280 - 330									Sup. temp. gyr.	42	-59	-32	8	FT7, TP7	NA
		330 - 380									22	-61	-45	21	CP5	3.13 [1.8; 4.46]²
230 - 280	280 - 330	330 - 380	380 - 430	430 - 480								60	-39	15	CP6, P6	4.66 [3.3; 4.4]²
										Cuneus	19	20	-82	36	PO4, O2	3.79 [2.075.49]²
						NA	Occip.	Inf. occip.		Ling. gyr.	19	-29	-76	-3	PO3, O1	4.31 [2.37; 6.23]²
								Med. occip.			18	-1	-76	2	OZ	4.01 [2.09; 5.9]²
						NA	Sub-cort. reg.	Thal.				-1	-12	11	NP	
								Ins.			13	-45	11	5	NP	

The P3a, or Novelty P3, occurs in response to all rare sounds or images ⁻designated as deviant stimuli ⁻regardless of whether they are targets[96-97] . The P3b appears after the appearance of a low-probability target item embedded in a train of high-probability nontarget (or standard) items. Both the amplitudes of the P3a and P3b, as well as their peak latencies, can be characterized as functions increasing (respectively decreasing) with age[98]. Additionally, the topographical distribution of the two P3 components seems to also be age-dependent[99].

Typically elicited between 250 ms and 500 ms, the P3b is believed to reflect the process of directed attention leading to conscious awareness of salient stimuli[100]. Detecting this cognitive component can reflect user intention and provide useful biomarkers for normal aging or several brain or mental diseases[101-102]. Table 4–5 summarizes the locations and time-courses (when available) of its neural generators from seven studies[100][103-108]. Furthermore, the source contributions calculated from the functional magnetic resonance imaging (fMRI) activation clusters are also indicated. However, as the authors warn, the intrinsic technical and theoretical differences between ERP and fMRI might cause some discrepan-

cies. For instance, the temporally dispersed activity reflected by fMRI (which is typically integrated over 1 second and then averaged) might get lost when compared to ERP (which is only averaged)[107].

Two different types of information are depicted in Figure 4–5. One highlights the contribution to the generation of the P300 for each electrode (low or high), while the other indicates the time-dependent activation for several EEG electrodes, using five 50-ms time intervals[105]. The chrono stimulation is illustrated on two consecutive images for simplicity's sake.

For the P300 task, the **CCS** subset consists of electrodes CP3, CP4, P1, P3, P2, P4, P6.

4.3   A Romance of Source Localized Spatio-Temporal (SLST) Features

The overarching objective is to identify recurrent brain sources across multiple trials for specific tasks, such as Left or Right Motor Imagery (MI). The signal of interest (SOI) in this context refers to the neural activity patterns consistently present when an individual performs an action. Identifying and analyzing these recurrent brain sources results in a better understanding of the underlying neural mechanisms and patterns associated with specific mental tasks, ultimately leading to improved classification and interpretation of brain activity.

4.3.1   The Source Localization Hypothesis

Source localization, a technique used to identify the specific brain regions generating cortical activity, provides valuable insights into the neural mechanisms underlying these tasks. We assume that emitted signals X from any brain source(s) are easier to discriminate than their scalp signals Y counterparts, that are mixed with noises. This means that the intensity of the brain source signals surpasses the intensity of the EEG potentials at the correct location: $||X|| >> ||Y||$. It should be emphasized that this inequality is valid for the specific spatial coordinates of the SOI, but does not hold true for all locations of brain sources. This source localized information is thus presumed essential for understanding how different mental processes are represented in the brain and for distinguishing the neural signatures of distinct cognitive activities.

However, knowing the most probable locations of the SOI has proven insufficient for effective classification. The timings during which these SOI are the most prominent also need to be extracted. To achieve this, we conjecture that EEG scalp potentials alone lack adequacy in providing such information. Previous studies have demonstrated a correlation between source localization results obtained from EEG and functional Magnetic Resonance Imaging (fMRI)[109-110]. That is why our main hypothesis is that these transformed spatio-temporal signals possess sufficient significance for robust classification.

Various challenges have been identified that can compromise the accuracy of source localization outcomes. Notably, the inherent variability in brain anatomy and function across individuals introduces model mismatch issues. Source localization methods, built upon generic head models, may struggle to precisely capture individual differences. This observation leads us to extend our hypothesis, suggesting

that the characteristics derived from source localization may not universally hold true across sessions and subjects. We posit that these features exhibit accuracy on a local timescale, i.e., within a single session.

To accommodate the inherent variabilities across trials, sessions, and subjects, we deliberately refrain from introducing additional assumptions about the characteristics detected through source localization. This approach avoids imposing requirements, such as the sequential activation of brain sources, and allows for a more flexible and inclusive analysis that acknowledges the complexities inherent in neuroimaging data across diverse experimental conditions. Although this approach may yield a less precise spatial source location for a specific subject, we hypothesize that the recurrent activation of (a) cortical source(s) holds significance, especially in terms of temporal dynamics.

4.3.2 Spatio-Temporal Patterns Detection

Identifying and extracting relevant feature(s) from the SOI(s) is a non-trivial question. The effectiveness of a chosen feature can be assessed based on its impact on classification accuracy. To choose such a feature, we further assume that the SOI, represented by the estimated source data matrix $\widehat{X}$ (in Talairach space), should be considered a sparse matrix, meaning that most of its elements are zero. This assumption arises from the consideration that the SOI is mixed with noise or background brain activity, and retaining all the values in $\widehat{X}$ would actually hinder classification accuracy.

The cleaned EEG potentials (after **ABCD**) $Y \in \mathbb{R}^{52 \times T}$, with T being the total number of data points, are transformed into source localization signals (cortical signals) $X \in \mathbb{R}^{20,484 \times T}$ at any given time point t. This conversion is achieved by calculating the inverse solution using a template brain derived from a combination of 40 MRI scans of real brains. This computation is performed using the MNE library with the eLORETA method in Python[51]. Adopting the default parameters of MNE results in the generation of one source signal per cortical vertex (i.e., 20,484 source signals).

eLORETA serves as a mathematical solution to a regularized, weighted minimum-norm problem. Its objective is to minimize the norm discrepancy between signals detected at scalp electrodes and the estimated values provided by the eLORETA method. Considering the inherent challenges of the ill-posed source localization problem, the inclusion of all voxels (or vertices) in the brain as potential sources introduces greater (and unwanted) complexity. This means that the process of source estimation is greatly underspecified.

In this context, the activity estimated in a particular voxel represents a weighted sum of activities from neighboring voxels, resulting in a blurred representation of the actual activity within the brain. Consequently, pinpointing the location with the highest activation (i.e., the maximally activated brain source $M \in \mathbb{R}^{3 \times T}$) serves as an approximation of the most probable source, offering a simplified yet informative perspective.

The maximum brain source activation $M \in \mathbb{R}^{3 \times T}$ is assumed to represent the feature of interest (FOI) and its intensity exceeds the intensity of the entire matrix of brain source activations $X \in \mathbb{R}^{20,484 \times T}$ (i.e., the

SOI) at the correct location, and by extension, the scalp potentials $Y \in \mathbb{R}^{52xT}$. Only statistically discriminant maxima are kept for further analysis as FOI. The obtained sparse matrix allows for an approximate solution to the source localization problem:

$$||M|| >= ||\widehat{X}|| >> ||Y|| \tag{4–1}$$

We also hypothesize that the SOI, and consequently the maximum brain source activation (or FOI), should exhibit stability both in time and space to accurately reflect genuine brain activity. This assumption ensures that the detected brain activity is not merely a transient signal caused by noise. To capture the consistent and reliable characteristics of the brain source activations, we integrate over time Δt on a volume $\Delta v = \Delta x \Delta y \Delta z$, which can be heuristically expressed as Equation (4–1), represented by:

$$\int^{\Delta t} ||M||_{\Delta v} >= \int^{\Delta t} ||X||_{\Delta v} >> \int^{\Delta t} ||Y||_{\Delta v} \tag{4–2}$$

Sliding time windows are swept across the 3D Talairach points, increasing as long as the consecutive distances are under a defined threshold (X=10; Y=10; Z=20) $\Delta v \subseteq v_0$ for a fixed minimum time interval $\Delta t \geq \tau_0$. This spatio-temporal threshold is linked to the available resolution of the template brain and the inter- and intra-subject variability. The sliding averaged 3D coordinates (computed for each sliding time window) are clustered according to the same distance threshold. Figure 4–6 illustrates the recurrent brain source activated across trials for a similar spatial cluster. The recurrent brain sources are retained for further analysis only when they occur at least with a probability p on all N trials.
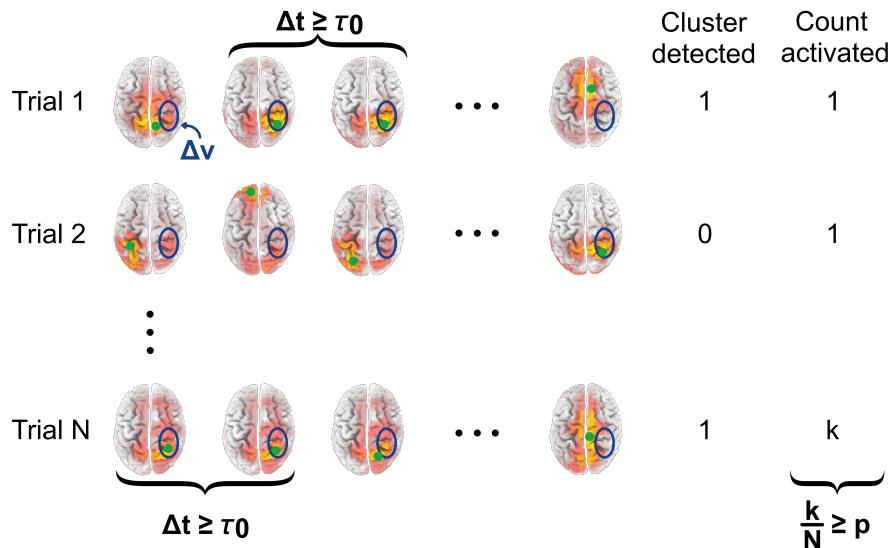
Figure 4–6 Identification of spatio-temporal recurrent clusters (visualization with LORETA-KEY software)

The number of repetitively activated sliding averaged 3D coordinates is counted per trial and across trials, as detailed in Algorithm 4–1. The probability that a spatial cluster is task-related depends on its

occurrence and the timings of its activation. This source localization-based spatio-temporal similarities during and across trials consist of a novel feature extraction technique in this thesis.

Algorithm 4–1 Pseudo-code for identification of Source-Localized Spatio-Temporal (SLST) features

Data: 3D source localized data S at each time point t, minimum time of activation τ_0, maximum volume Δv, total number of trials N, probability of occurrence p

Result: recurrent spatio-temporal features

1 Initialize features and occurrence of features: $Med = []; count = []$

2 **for** $i = 1$ *to* N // Distance between two consecutive 3D source localized data

3 **do**

4      $t_1 = t$ // Loop over N trials

5      $d(t) = S(t+1) - S(t)$ // Time start

6   **while** $d(t) < d_{x,y,z}$ // Same zone

7   **do**

8        $t_2 = t$ // Time stop

9   **if** $\Delta t = t_2 - t_1 \geq \tau_0$ **then**

10   **return** $(Med, count)$ // Same time

11         $Med = \text{median}_{[t_1,t_2]}(S(t))$ // Median on same consecutive time-zone

12         $count \mathrel{+}= 1$ // Count the number of trials on which the same consecutive time-zone median is identified

13   **end**

14 **end**

15 **if** $count > p$ **then**

16   **return** $(Med, count)$ // Spatio-temporal feature identified on at least p trials

---

4.4 The Poetics of Dual Classifier (DC)

The underlying rationale is grounded in the notion that brain activation in task-related regions occurs only for brief durations. Consequently, our dual classifier (**DC**) exclusively operates on the temporal information extracted from the **SLST** features. For reference, most conventional classifiers commonly employ features derived from extended windows, typically spanning several seconds.

4.4.1 Exclusive Electrodes

The **SLST** analysis yields two types of outcomes: the spatial location, which can be associated with the nearest EEG electrode, referred to as the electrode for simplicity, and the corresponding activation times of this electrode.

During the MI training phase, the electrodes can be categorized as either exclusively associated with Left MI, exclusively associated with Right MI, or responsive to both. When analyzing a new trial, if the **SLST** analysis identifies an exclusive Left (resp. Right) electrode, the trial is straightforwardly classified accordingly. Across all sessions, the average number of exclusive electrodes is found to be 2 ± 1. However, there are instances where no exclusive electrodes are detected during certain sessions or trials that lack the presence of exclusive electrodes. In such cases, an alternative strategy is developed for classification that relies on the covariance matrices calculated at the **CCS**.

4.4.2 Voting System on the Riemannian Manifold Using Covariance Fréchet at CCS

The scalp EEG potentials are extracted at the **SLST** timings and the corresponding covariance matrices are computed. These resulting 52x52 covariance matrices serve as a means to capture and represent the spatial relationships among the various electrode locations.

Covariance matrices, being semi-definite positive matrices, possess a fixed-dimension Riemannian manifold structure. This manifold represents a convex smooth submanifold within Euclidean space. It is equipped with an inner product inherited from the Euclidean metric defined on its tangent space, enabling the calculation of distances between its elements. Riemannian-based classification uses the geodesic distance, which refers to the shortest length curve connecting two points on the manifold. This facilitates the formation of clusters based on their proximity to a mean value. The Fréchet mean (resp. median), known as the natural mean value for such clusters, minimizes the mean distance to the elements within the cluster. The convexity of the manifold ensures the existence of this solution.

To illustrate this method, let's consider a simple example in R^2, where a convex Riemannian submanifold, such as a circle, serves as an illustration. In this example, two clusters, each containing two elements, are examined. It is important to note that the same principle applies to scenarios involving multiple clusters, elements, and higher dimensions.

To measure the similarities among a cluster of covariance matrices (e.g., Left or Right MI), the Fréchet median can be used to represent the central tendency of a set of covariance matrices. The distances between each covariance matrix and the Fréchet median of each label are compared. The predicted label corresponds to the smallest distance to the geometric median. For each trial, if there are more than one timing extracted from the spatio-temporal similarities, a voting system is set with the predicted label corresponding to the label that is the most picked by the different covariance matrices.
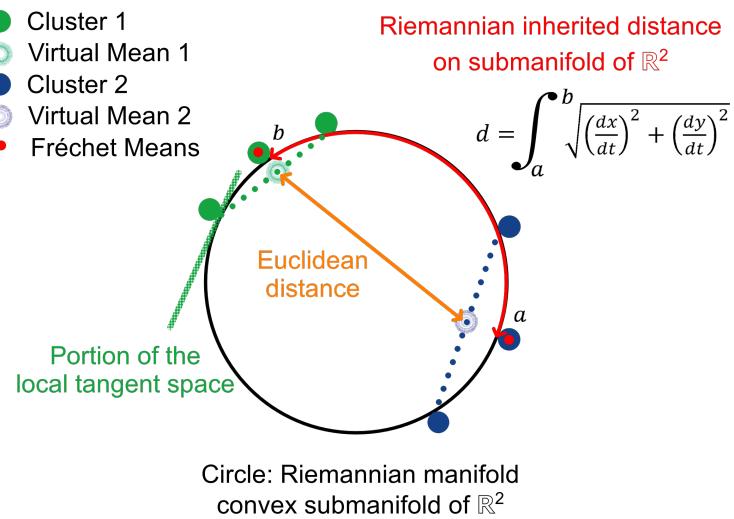
Figure 4–7 Illustration of the Riemannian manifold on a simple example

The figure contains the following labels:

- Cluster 1
- Virtual Mean 1
- Cluster 2
- Virtual Mean 2
- Fréchet Means

Riemannian inherited distance
on submanifold of $\mathbb{R}^2$

$$d = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}$$

Euclidean distance

Portion of the local tangent space

Circle: Riemannian manifold
convex submanifold of $\mathbb{R}^2$

4.5 Classification Under the Influence of Filtering

Filters are designed to optimize the frequency response, with the primary objective of favoring signals within designated frequency bands while simultaneously maximizing rejection and minimizing any undesirable ripple. Band-pass filtering is commonly employed in the BCI processing pipeline to emphasize the waveform of interest (e.g., beta band) under the assumption that it reveals the SOI neuronal synchronization.

4.5.1 shinyVizFilter

Thanks to its interactive design, shinyVizFilter facilitates the visualization of signals in both the temporal and spectral domains, enabling an examination of the filtering effects based on the selected characteristics. Many software and programming languages are available for analyzing EEG signals. This diversity might sometimes become a burden when trying to use the available online code to perform exploratory data analysis. To alleviate this issue and enhance accessibility for both novice and experienced programmers, shinyVizFilter has been crafted as a user-friendly Web-page tool, accessible at https://eva-guttmann-flury.shinyapps.io/shinyVizFilter/.

This tool is implemented in the R programming language, renowned for its open-source nature and extensive ecosystem. Developed as a Shiny app hosted in the cloud, shinyVizFilter allows users to freely and interactively modify filter characteristics while simultaneously visualizing the effects on any EEG data. Additionally, shinyVizFilter is also available as an R-package, facilitating seamless integration into other statistical R-packages or larger workflows. Its design aims to facilitate a comprehensive understanding of the impact of filtering on the EEG data at hand, with a particular emphasis on intuitive visualization of

band filtering.

4.5.2 Comparison Across Filter Types and Bands

Within the realm of linear Fourier filters, the three most extensively investigated types are the Butterworth, Chebyshev, and Elliptic filters. Butterworth filters are commonly employed due to their characteristic of providing a maximally flat response in the passband. This implies that they exhibit a uniform gain across the desired frequency range, making them a suitable choice when maintaining passband flatness is of utmost importance. In contrast, Chebyshev filters introduce ripple in the passband but compensate for this by offering steeper attenuation slopes compared to Butterworth filters. Consequently, they are preferred when a sharper roll-off is desired, even at the expense of passband ripple.

Table 4–6 Comparison classification accuracy for four frequency bands and three type of filters; data from the MI paradigm of the first ten subjects (25 sessions or 1000 trials) in our experiment (detailed in Chapter 2)

Filter band	Filter type	Classification accuracy (with CI at 95% confidence level)
Unfiltered		85% [64%; 96%]
Delta	Butterworth	87% [67%; 97%]
	Chebyshev	89% [68%; 98%]
	Elliptic	89% [68%; 97%]
Theta	Butterworth	89% [69%; 97%]
	Chebyshev	88% [68%; 97%]
	Elliptic	83% [62%; 94%]
Alpha	Butterworth	91% [71%; 99%]
	Chebyshev	92% [72%; 99%]
	Elliptic	92% [72%; 99%]
Beta	**Butterworth**	**94% [75%; 98%]**
	Chebyshev	92% [72%; 99%]
	Elliptic	54% [62%; 95%]

On the other hand, elliptic filters exhibit ripple in both the passband and stopband, yet they excel in providing the most pronounced attenuation slopes among the three filter types. As a result, elliptic filters are selected when there is a requirement for both a sharp roll-off and a high degree of stopband attenuation. Even though the phase response of the Butterworth filter has been found to become more non-linear with increasing order, it remains the most favorable compromise between attenuation and phase response[111]. Butterworth filters have also been found to have the lowest effects on time shifts, thereby minimizing

distortion of neuronal timing information[112].

The influence of these three types of filters is evaluated by comparing the classification accuracy obtained with the three rithms developed in this thesis, namely **ABCD** for the pre-processing, **SLST** for feature classification, and our dual classifier. Table 4–6 summarizes the classification accuracies obtained for the first ten subjects (for a total of 25 sessions or 1000 trials).

4.5.3 A Spatial Re-Referencing Odyssey

Common physical reference channels during EEG recordings are positioned at the cephalic electrode, such as the vertex (Cz), the tip of the nose, or the mastoid (behind the ear). The re-referencing goal is to correct these temporally non-neutral references by calculating a theoretically neutral one. Popular techniques include the Common Average Reference (CAR) and the Laplacian references. The CAR method relies on the assumption that the head can be approximated as a sphere and that the sum of all potentials recorded due to current sources inside this sphere would be equal to zero. However, since EEG electrodes cover mostly the upper part of the head, these hypotheses are quite unlikely[113].

The Laplacian reference aims at attenuating low spatial frequency signals that are broadly distributed across the scalp while preserving high spatial frequency signals that are localized in a small volume below each electrode. This reference can be estimated by calculating the difference between the potential at each electrode and the averaged potential of its four neighboring sensors. In the case of the small Laplacian, the nearest four neighbors are chosen, while for the large Laplacian, further neighbors are selected (see Figure 4–8).
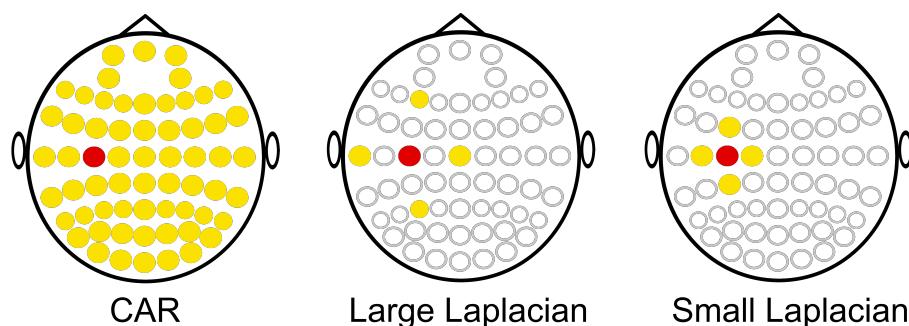
Figure 4–8 Common spatial re-referencing: comparison between CAR, large and small Laplacian

The impact of these three types of re-referencing, along with no re-referencing (simply called None hereafter), is evaluated by comparing the classification accuracy obtained with our three algorithms (**ABCD**, **SLST**, and **DC**) after filtering on the beta band using the Butterworth filter. Table 4–7 summarizes the classification accuracies obtained for the first five subjects (for a total of 13 sessions), revealing the superiority of the Laplacian re- referencing (with very similar results for the Large or the Small variant).

Table 4–7 Comparison classification accuracy for four re-referencing; data from the MI paradigm of the first ten subjects (25 sessions or 1000 trials) in our experiment (detailed in Chapter 2)

Classification accuracy (with CI at 95% confidence level)	Classification accuracy (with CI at 95% confidence level)
None	90% [70%; 88%]
CAR	93% [74%; 99%]
Small Laplace	93% [74%; 99%]
Large Laplace	**94% [75%; 98%]**

4.6   Summary of Results and Discussion

The recorded data (see Chapter 2 for details about the experiment) is pre-processed to correct blinks with our newly-developed **ABCD**. The cleaned EEG data is re-referenced with the Large Laplacian spatial filter, followed by the application of a Butterworth beta (12.5-30 Hz) pass-band filter. The resulting EEG data is converted with eLORETA using the MNE library in Python. The brain source exhibiting maximum activation is extracted, and its spatio-temporal characteristics are analyzed using **SLST**. Finally, the extracted data is subjected to classification using the dual classifier (**DC**).

The ensuing results are compared to four renowned feature extraction and classification algorithms. The first, and perhaps most widely recognized, is Common Spatial Patterns (CSP) combined with Linear Discriminant Analysis (LDA). Another well-known variant involves using both CSP and Support Vector Machine (SVM) for classification purposes. Given that our Dual Classifier (DC) incorporates Riemannian geometry as a partial basis for trial classification, a comparative analysis is also conducted with other established methods that utilize Riemannian geometry. Among these methods, the Minimum Distance to the Riemannian Mean (MDRM) and the Tangent Space (TS) are the most widely recognized.

Similar to the **Consistency** and comparison procedure carried out between the three pre- processing algorithms, these four well-known classification methods are compared to our **SLST+DC** on twenty successive random draws. The resulting averaged accuracy presented in Table 4–8 confirms that **SLST+DC** consistently achieves higher classification accuracy compared to CSP+LDA, CSP+SVM, MDRM, and TS.

The algorithms for feature extraction and classification developed in this thesis are founded upon the outcomes obtained through source localization. This methodology allows us to derive task-related characteristics, explaining the huge improvement of our methods compared to the others. Nevertheless, this advantage is accompanied by a considerable computational burden. The computational time is contingent on each choice of re-reference, filter type, and filter band, introducing a multitude of permutations. Additionally, given that the optimized pipeline is not inherently known, years of computation on multiple computers and virtual machines (VM) have been required to achieve these results. Table 4–9 summa-

Table 4–8 Consistency and comparison between four common feature extraction and classification algorithms and our SLST+DC, ICA and ASR (20 successive random draws, corresponding to 6.3 sessions in average); all data is first pre-processed with our ABCD algorithm for a fair comparison

Methods	Consistency (mean of 20 draws, corresponding to 6.3 sessions in average)		
	Accuracy	Standard Deviation (SD)	Standard Error (SE)
CSP + LDA	89.16%	1.76%	0.39%
CSP + SVM	88.99%	1.60%	0.36%
MDRM	81.13%	4.64%	1.04%
TS	86.09%	4.48%	1.00%
SLST + DC	**95.03%**	**3.41%**	**0.76%**

rizes typical computational time to obtain the source localization results from MNE of our dataset. The exhaustive exploration of these configurations, especially for subject-specific filter bands, would require exponential computational time and pose a trade-off between computational intensity and practical feasibility.

Table 4–9 Benchmark of source localization computing times

Machine	Specifications	Computational time per session (in hours)
Laptop 1	11th Gen Intel(R) Core(TM) i9-11900K CPU @ 3.50GHz	~2.7
Laptop 2	Intel(R) Core(TM) i7-6700HQ CPU @ 2.60GHz	~9
Server with 30 VM	Intel(R) Xeon(R) e5-2698B V3 CPU @ 2.00GHz	~3.5

Our feature extraction and classification are obviously not designed for real-time applications. How-

ever, we anticipate the potential to derive certain features from their outcomes, thereby potentially circumventing the need for the source localization computation. Meanwhile, supplementary analyses can be conducted using the identical pipeline on alternative recorded paradigms, as detailed in Chapter 2. This research endeavor, for instance, can aid in ascertaining the consistent superiority of the Butterworth filter across paradigms or determining the optimal frequency band (1-15Hz) for P300 signals, as frequently reported.

4.7   Conclusion and Highlights

This chapter centers its attention on mitigating the important inter- and intra-subject variability, which predominantly causes low classification accuracy. The primary conjecture posits that EEG signals are not suitable for direct BCI classification. This is analogous to attempting to solve a system with three unknowns, namely, time, location, and frequency.

Hence, the development of the Source Localized Spatio-Temporal (**SLST**) feature extraction method serves the purpose of identifying two of the aforementioned unknowns through the utilization of similarity analysis across multiple trials. The Dual Classifier (**DC**) is then employed to process the extracted features. The concepts and main results presented in this chapter can be summarized into a few main ideas:

The diverse features and specificities of this method can be outlined in a few essential ideas:

- Cortical sources activations differ depending on the BCI task and their modulations.
- By combining the findings of twenty-one studies, an optimized subset of EEG electrodes is determined for each BCI task. This knowledge-based core channel selection (**CCS**) framework can serve as a reference for EEG researchers.
- The eLORETA algorithm is used to compute the source localization of **ABCD**-cleaned EEG signals, while the features are extracted using **SLST**. The resulting spatial locations and time-derived features are utilized as inputs for the **DC** method.
- **DC** incorporates time-derived covariance matrices and calculates the Fréchet medians associated with each class. When classifying a new distance, the distance between the covariance matrices of the new trial and the class-specific Fréchet medians is computed. A voting system then helps to determine whether the user intends to move the left or right hand.
- The shinyVizFilter tool is an interactive online resource specifically developed to facilitate the visualization of EEG signals across temporal, spectral, and spatial domains. The impacts of filtering are studied on **SLST+DC** for different frequency bands (delta, theta, alpha, and beta) and filter types (Butterworth, Chebyshev, and Elliptic), revealing that the best filter for MI is the beta passband Butterworth filter.
- The effects of various spatial re-referencing techniques, namely the Common Average Reference (CAR), Large and Small Laplace, are also investigated, showing the superiority of the Laplacian

filter (with both the Large and Small Laplace filters yielding comparable outcomes).

- With twenty successive random draws, corresponding to 6.3 sessions on average, **SLST+DC** achieves a Consistency mean classification accuracy of 95.03% (SD = 3.41%), versus 89.16% (SD = 1.76%) for CSP+LDA, 88.99% (SD = 1.60%) for CSP+SVM, 81.13% (SD = 4.64%) for MDRM, and 86.09% (SD = 4.48%) for TS.

The efficacy of the **SLST+DC** approach in classification substantiates its potential for wider implementation in EEG applications. For instance, it would be of considerable interest to ascertain the applicability of these methodologies in diagnostic scenarios or the identification of novel brain signals.

Chapter 5   Standard Deviation-based Confidence Interval Estimations (SDCI) and Innovative Elliptical Representations for Confusion Matrices (CE)

This chapter focuses on the development of sound approximate calculations for confidence intervals (CI) associated with a confusion matrix (CM) and the subsequent classification accuracy. The CI for the four main performance metrics of the CM are estimated using a novel induced standard deviation for accuracy. An alternative approach to estimating the minimum sample size is built using a newly-established distance separation (**DS**) method relying on the chosen CI approximation. An original method for computing the accuracy's CI estimated through standard deviation (**SDCI**) is also presented and applied to innovative visual representations of confusion matrices called confusion ellipses (**CE**).

While several CI approximations already exist, the most common being the Wald- Laplace method, their application to classification accuracy is uncommon, except for PyCM, a Python library dedicated to CM evaluation and comparison. Similarly, there is a notable scarcity of easily comprehensible CM representations. The majority of existing visualizations typically involve matrices of colored squares, which can pose challenges in terms of understanding and interpretation.

Simulated data generated from a Binomial distribution (with a sample size of 30) is employed to assess the differences between six CI approximations. The minimum sample size is similarly estimated using a Binomial distribution (with a sample size of 1000). Finally, a comparison is carried out between the accuracy's CI obtained from our **SDCI** vs. PyCM, using simulated data from 35 CMs. The results show that **SDCI** yields a CI mean widths of 15% against 9.54% for PyCM, indicating that PyCM underestimates by 36.38%. **CE** plots are also generated for simulated data from two CM and for the previous results comparing **ABCD** vs. ICA, showing a qualitative improvement over more conventional visualizations.

The newly developed **DS** method can serve as an alternative approach for estimating sample sizes, particularly when the signal of interest is mostly unknown and only competitive accuracies from different pipelines are reported (the most widely reported metrics). The innovative **CE** representations can also be plotted as a summary of results when researchers wish to illustrate the differences between various CM.

5.1   Introduction

5.1.1   Binomial Distributions

Confusion matrices (CM) serve as the conventional means to evaluate the performance of a classifier. These matrices are derived from a supervised learning process and are designed to provide predictions

regarding correct (True) or incorrect (False) assessments of two distinct samples, typically labeled as Positive and Negative sets. These two samples can be represented as follows:

- The positive sample (P) consists of instances classified as True Positive (TP) and False Negative (FN) instances, with the corresponding probability of TP/(TP+FN);
- The negative sample (N) comprises instances classified as True Negative (TN) and False Positive (FP), with a probability of TN/(TN+FP).

The statistical distributions for each sample are discrete Binomial distributions, with only two possible classes: True or False. These distributions are denoted as $B(n, p)$, where n represents the sample size and p represents the associated probability. The Binomial distributions constitute a family of distributions that depend on two parameters: n and p. The cumulative distribution function associated with these distributions calculates the probability of obtaining at least k correct assessments:

$$B(n, p, k) = \mathbb{P}(X \leq k) = \sum_{i=0}^{\lfloor k \rfloor} \binom{n}{i} p^i (1 - p)^{n-i} \tag{5--1}$$

Where $\lfloor k \rfloor$ is the "floor" under k, i.e., the largest integer less than or equal to k.

5.1.2 Confidence Intervals (CI): Analytic Approximations and PyCM

The simple nominal probabilities alone are not sufficient and should be presented within confidence intervals (CI). Experimental probabilities, denoted as $\widehat{p}$, are associated with a confidence level typically set at 95%, resulting in a two-sided tail of $\alpha_2 = \frac{\alpha}{2} = 2.5\%$.

Traditionally, these CI have been calculated using approximations of $B(n, p)$ with continuous distributions that have straightforward analytic expressions for their CI. These approximations have been useful in situations where inexpensive and effective calculation methods are not available. However, significant differences arise near the extreme probabilities of 0 and 1, leading to frequent misunderstandings and questions about their validity. Therefore, the choice to use any of these methods must be justified and supported with sound arguments.

Recent advancements in this field have been made through the online-available PyCM[114]. This Python library specifically caters to data scientists who require a wide range of metrics for predictive models and comprehensive evaluation of various classifiers. However, some of the hypotheses used in the computations are questionable. For instance, the CI estimation for classification accuracy assumes that the sum of two Binomial distributions is a Normal one, despite this being known to be inaccurate.

5.2 The Color of Binomial Approximations

The Binomial approximation is a specific case of approximating a discrete distribution with a continuous one. The choice of this estimation depends on its intended application. The primary objective is

to obtain approximate confidence intervals (CI), which are calculated by summing the probability mass function (PMF) until the probability α_2 is reached. This iterative process starts with 0 for the lower bound or 1 for the upper bound, representing the Binomial tails. It is crucial to emphasize that only the tails of the distribution are estimated, not the entire distribution itself. This will be referred to as the Binomial approximation, as it is the common term used to describe either the distributions or their CI, the fundamental difference being that the CI themselves are not probability distributions.

5.2.1 Binomial Tails

An asymptotic confidence interval (**ACI**) is a statistical tool used to estimate an unknown population parameter based on a sample. It provides a range of values where the true parameter is likely to be, with a certain confidence level. The term "asymptotic" refers to the interval's behavior as the bootstrap size increases indefinitely. It should be mentioned that this term may be ambiguous due to its frequent association with the Normal distribution, which asymptotically tends to the Binomial ones for large sample sizes. As the sample size grows, the interval becomes more accurate and approaches the true parameter value. The ACI is issued from the Binomial formula mentioned in Equation 5–1 and the pseudo-code, as outlined in Algorithm 5–1, to calculate them for any given probability is:

Algorithm 5–1 Pseudo-code for asymptotic confidence interval (**ACI**) construction

---

    Data: the sample size n, the probability p, the confidence level α_2

    Result: confidence interval for the lower (resp. upper with $1 - \alpha_2$) bound

1 Initialize the confidence level probability: $p\alpha = 0$

2 **for** $i = 1$ *to n* **do**

3   **if** $p\alpha = 0 \leq \alpha_2$ **then**

4         $p\alpha + \binom{n}{i} p^i (1-p)^{n-i}$ // Sum of the PMF, corresponding to the area
                                              under the Binomial distribution function

5   **else**

6   **break**   // Break when confidence level probability is
                                         greater than critical value α_2

7   **end**

8 **end**

9 **return** $\frac{i-1}{n}$

The lower and upper bounds distributions obtained for any and/or are represented by two symmetric step functions that are continuous almost everywhere. Figure 5–1A shows the n steps of 1/n height (excluding zero) and the central symmetry along the nominal probability, although not around every nominal value. The upper and lower bounds usually differ, except in the vicinity of $p = 0.5$, which serves as the center of symmetry. The left (resp. right) Binomial tail corresponds to the lower (res. upper) bound.
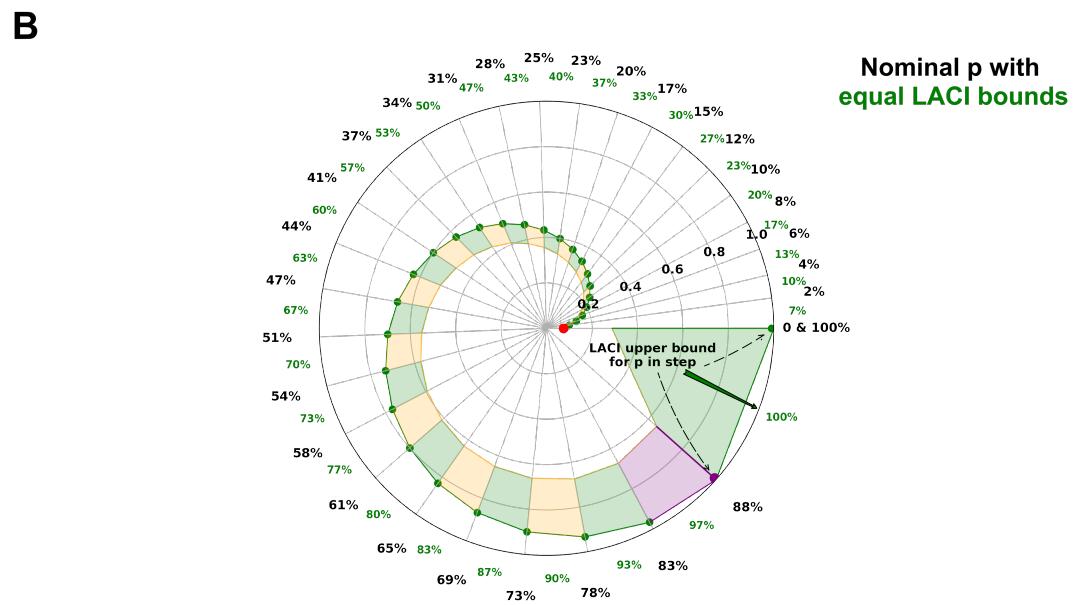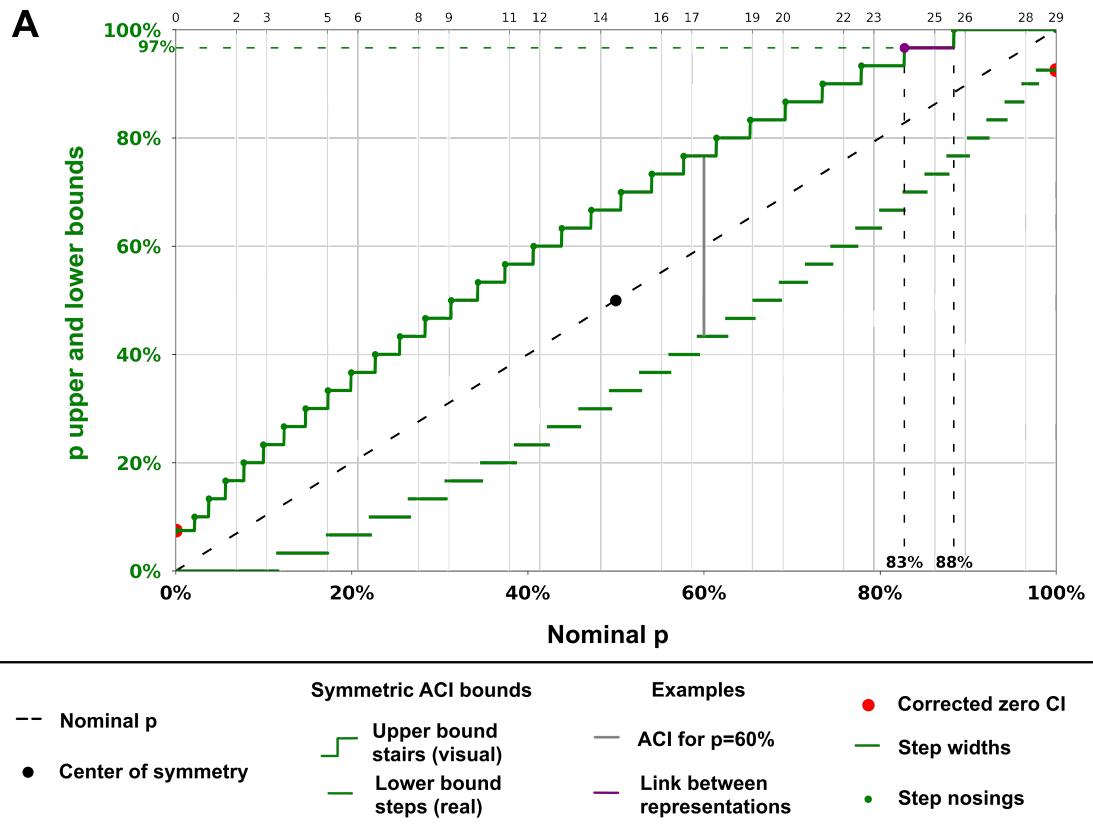
Figure 5–1 Asymptotic Confidence Intervals (ACI) upper and lower bounds for a Binomial distribution (n=30) at 95% confidence level; dual representation showing the standard (A) and snailed LACI (B). The purple example illustrates how the snailed LACI is constructed from the standard.

For simplicity, both bounds will be referred to as Lateral Asymptotic Confidence Intervals (LACI). To enhance the readability of the ACI and visualize the step values, a new representation called snailed LACI has been introduced, specifically designed for sample sizes in the medium range, as illustrated in Figure 5–1B.

The equality of bounds for ACI does not imply the equality of the ACI themselves. It simply indicates that a slight increase in the nominal probability will not change the ACI if the opposite lower or upper bounds are also locally equal. These neutral probabilities ACI do not accurately represent practical CI, which are derived from long-run simulations or real experiments that allow for a proper assessment of their coverage.

Coverage refers to the probability that a CI will contain the true value of a population parameter. It is defined as the percentage of trials for which the estimated value remains within the chosen CI approximation. However, this definition poses a problem as it is based on a frequentist approach, while most methods are mathematically grounded in Bayesian reasoning, either directly or indirectly.

Discrepancies between these estimated results are typically quantified through their performance in terms of apparent coverage. However, this approach can be misleading and often lacks comprehensive explanations justifying the selection of a particular method. Before delving further, it is important to explain why the preceding formula yields values close to zero when the probability is near 0 or 1 (by symmetry). Zero is a singular point that is addressed in the Bayesian approach, but not in the frequentist one.

5.2.2 Comparison of CI Approximations

All CI approximations are fundamentally derived from an original Laplace-Bayes estimator, also known as the rule of succession. This rule states that when no additional information is available about random independent variables with binary values (0, 1), the probability of success is given by:

$$\mathbb{P}(X_{n+1} = 1 | X_1 + ... + X_n = s) = \frac{s + 1}{n + 2} \tag{5–2}$$

Continuity corrections use different strategies based on this equation to adjust the approximation of a discrete probability distribution to a continuous distribution. This involves shifting the k successes to simulate a probability density function around the value of the probability mass function (PMF), while simultaneously shifting the sample size to compensate for the imbalance.

The Clopper-Pearson method is directly derived from this formula. By using a non- informative probability distribution (i.e., a uniform distribution), the posterior distribution can be modeled as a Beta distribution. This Beta distribution yields the desired expected value: $\frac{s+1}{n+2}$. This is why the Clopper-Pearson approximation intervals are considered "exact". Other approximations, on the other hand, are generally not mathematically justified.

As the sample size increases, the effect of shifting by 2 in the formula becomes negligible, as does the impact of the numerator (number of successes). According to the Central Limit Theorem, as the sample size increases, the distribution of the sample mean approaches a normal distribution. In this case, $\frac{1}{n}$ represents the scaling factor for the standard deviation of the sample mean. This explains why the basic normal approximation, known as the Wald-Laplace interval, becomes the asymptotic value for all CI approximation methods when the sample size is sufficiently large. This limit approximation is however not accurate for very small (close to 0) or very high (close to 1) probabilities.

A general expression of the CI issued from the CDF of the normal distribution $N(0, 1)$ is obtained without any continuity correction (i.e., for Wald-Laplace):

$$p = \widehat{p} \pm Z \sqrt{\widehat{p} \, \frac{1-p}{n}} \tag{5–3}$$

Tens of CI approximations have been developed using different assumptions and properties. A few standard references for CI approximations are listed below:

- Clopper-Pearson

$$\begin{cases} p_{upper} = 1 - Beta.Inverse(\frac{\alpha}{2}, n - k, k + 1) \\ p_{lower} = 1 - Beta.Inverse(1 - \frac{\alpha}{2}, n - k + 1, k) \end{cases} \tag{5–4}$$

- Jeffreys

$$\begin{cases} p_{upper} = 1 - Beta.Inverse(\frac{\alpha}{2}, n - k + 0.5, k + 0.5) \\ p_{lower} = 1 - Beta.Inverse(1 - \frac{\alpha}{2}, n - k + 0.5, k + 0.5) \end{cases} \tag{5–5}$$

- Laplace with continuity correction

$$p = \tilde{p} \pm Z \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{\tilde{n}}} \quad with \quad \begin{cases} \tilde{p} = \frac{k+1}{\tilde{n}} \\ \tilde{n} = n + 2 \end{cases} \tag{5–6}$$

- Wilson

$$p = \tilde{p} \pm Z \frac{\sqrt{\tilde{p}(1 - \tilde{p}) + Z^2/4}}{\tilde{n}} \quad with \quad \begin{cases} \tilde{p} = \frac{k + Z^2/2}{\tilde{n}} \\ \tilde{n} = n + Z^2 \end{cases} \tag{5–7}$$

- Agresti-Coull

$$p = \tilde{p} \pm Z \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{\tilde{n}}} \quad with \quad \begin{cases} \tilde{p} = \frac{k + Z^2/2}{\tilde{n}} \\ \tilde{n} = n + Z^2 \end{cases} \tag{5–8}$$

A Confidence level at 95% (n=30)

B Confidence level at 99% (n=30)
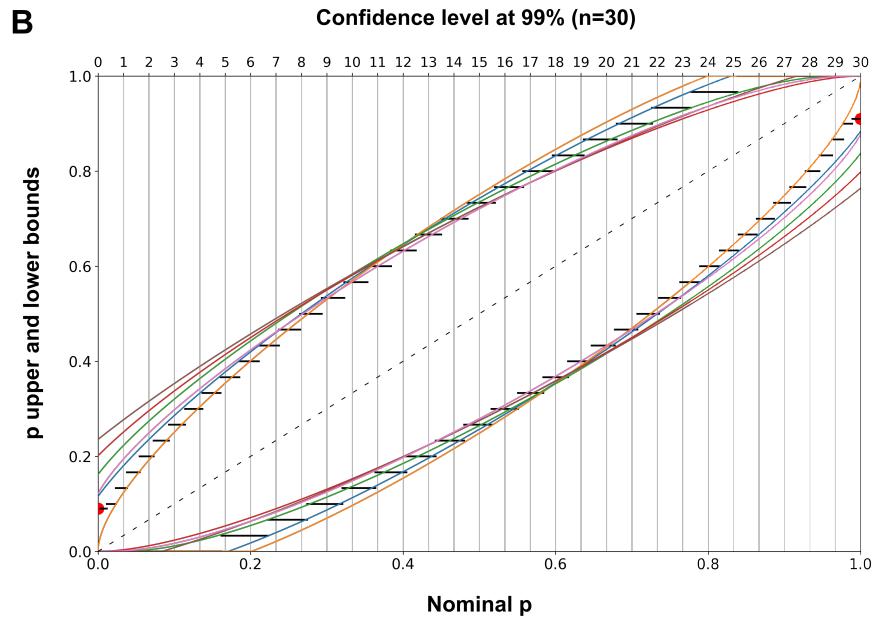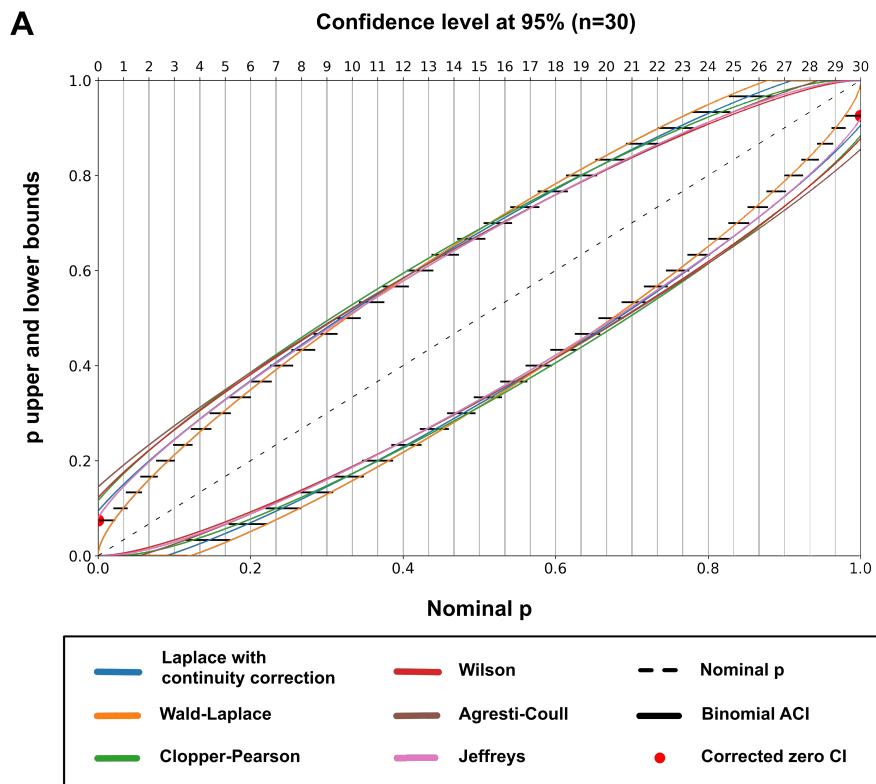
Figure 5–2 Seven different CI approximations with the Asymptotic Confidence Interval (ACI) upper and lower bounds for a Binomial distribution (n=30) at 95% (A) and 99.9% (B) confidence levels

The here so-called Laplace with continuity correction is simply a mix of the normal approximation with the "rule of succession" from Laplace. These methods all tend towards the Normal approximation for a large sample size, effectively addressing the limitations of the Wald-Laplace approximation in the vicinity of 0, as depicted in Figure 5–2.

5.3 The Book of CI Approximation

The primary goal is to understand the nature of the sought-after confidence intervals (CI) and assess the suitability of the coverage probability procedure used for comparing various approximation methods. With a "naïve" frequentist approach, the aim is to find the probability of obtaining the correct parameter value and, assuming its distribution is binomial, to anticipate whether 95% (the usual confidence level) of the draws will be included in the calculated CI (i.e., 5% of the draws are expected not to be included). However, this naïve assumption is incorrect, leading to ongoing debates regarding the appropriate approximation method(s) and their evaluation through the coverage probability procedure.

5.3.1 Uncertainty Factors

Essentially, there are two fundamental sources of uncertainty that must be addressed:

1. Is the binomial distribution $B(n, p)$ the correct one? In other words, is the observed sample probability accurate (i.e., similar to the population probability)?
2. What is the probability of a draw from the chosen distribution to fall within a given CI?

Methods for obtaining approximate intervals, excluding Clopper-Pearson, tend to address the second question. They indirectly assume that the observed sample probability p is correct, which is generally a false assumption. This hypothesis has minimal consequences when the sample size is large. This is, however, not true for critical cases where the probability is very low, yielding significant inconsistencies.

Consequently, two scenarios can be encountered: either additional information is available that confirms the accuracy of (a large amount can be assimilated to this case), or there are no other clues about its value. In the former situation, the coverage procedure can appropriately be used and holds significance. In the latter case, it becomes meaningless, except for the Clopper-Pearson interval, which demonstrates its correctness by achieving a probability coverage of 100%.

5.3.2 Illustrating Example

Let's assume an experiment was conducted and resulted in 9 zeros from 10 Bernoulli draws (a small sample size of $n = 10$ is here chosen to emphasize the phenomenon). This unlikely succession of draws has only a 10% chance of happening ($p = 0.1$). Yet, it would be very conceivable to either have obtained 8 or 10 zeros in a similar unlikely event. The corresponding approximate intervals upper bound for different

102

CI approximations are given in Table 5–1.

Table 5–1 Approximation of the Confidence Interval's (CI) upper bound for the illustrative example; the asterix * shows the nominal probability p (i.e., the probability in this particular experiment)

Number of zeros (#)	Probability	Clopper-Pearson	Jeffreys	Wald-Laplace
8	0.2	0.556	0.503	0.482
9	0.1*	0.445	0.381	0.311
10	0.0	0.308	0.217	0.0

These results reveal the inadequacy of the Wald-Laplace interval, which produces an upper bound of 0 when $p = 0$. More importantly, the Clopper-Pearson bound for # = 9 falls within the middle range of the Jeffreys upper bound for # = 8 and # = 9, as well as for # = 9 and # = 10. This indicates that the Clopper-Pearson interval, calculated with a nominal probability of $p = 0.1$, covers half the misses of the Jeffreys one when calculated on $p = 0.0$ and $p = 0.2$. That is because the Clopper-Pearson method takes into account the uncertainty surrounding the accuracy of the model distribution. It does not assume that the true distribution precisely follows a $B(10, 0.1)$ distribution, but rather considers it to be between the $B(10, 0.05)$ and $B(10, 0.15)$ distributions.

To sum up, in the absence of additional information, more rigorous methods will use the Clopper-Pearson approximate interval, which is referred to as the credible interval in the Bayesian approach. Labeling it as conservative is a misinterpretation of its significance, as it fulfills its intended purpose precisely. Similarly, the Jeffreys method is specifically devised to eliminate the underlying Bayesian interpretation. This analysis also clarifies why the naive approach fails to meet the 95% CI probability assumption. Consequently, confidence intervals do not warrant their name in such instances, as they cannot be associated to a confidence level. These considerations are of the upmost importance, yet largely unstated, for justifying the selection of interval approximation methods, which in turn ensure accurate comparison of classification results derived from the confusion matrix.

5.4   Through the Distance Separation (DS) Method

The calculation of the significant difference between two classification results, with a 95% confidence level, follows the same method as classical significance tests. This approach assumes that in 95% of cases, one result is significantly better than the other. In other words, the confidence intervals of each result do not overlap. For instance, when comparing the success rates (p_1 and p_2) of two classifiers (C_1 and C_2), it is necessary for their respective confidence intervals to not interfere.

However, this principle lacks informativeness when the limit is not reached. To address this, we

A

90% and 95% classification results
CI type: Clopper-Pearson

$D_{459} = 0$

Lb_{M250} Ub_{m250}

Lb_{m250}

$D_{250} = \frac{Lb_{M250} - Ub_{m250}}{Ub_{m250} - Lb_{m250}} = -0.22$

— CI for min p = 90%
— CI for Max p = 95%

- - Nominal p
······ Student Normal (30)

● Intersection of min upper and max lower bounds

CI interference

CI separation

82% and 93.5% classification results
CI type: Clopper-Pearson

B

$D_{139} = 0$ Lb_{M200}

Ub_{m200}

$D_{200} = \frac{Lb_{M200} - Ub_{m200}}{Ub_{m200} - Lb_{m200}} = 0.24$

Lb_{m200}

— CI for min p = 82% (ICA)
— CI for Max p = 93.5% (ABCD)

Figure 5–3 Sample size estimation using the Distance Separation (DS) method: (A) illustrative example for 90% and 95% classification results; (B) real data from ABCD vs ICA comparison

introduce a more indicative quantification that measures the level of interference between the respective confidence intervals at each sample size. The corresponding distance separation (**DS**) is calculated using:

$$D_n = \frac{Lb_{Mn} - Ub_{mn}}{Ub_{mn} - Lb_{Mn}} \tag{5-9}$$

Where L_b and U_b represent the lower and upper bounds, M and m the maximum and minimal probabilities, and n the sample size where it is evaluated.

The ratio is negative when the discriminating sample size is not reached, zero at the critical sample size, and then becomes positive. This pattern is depicted in Figure 5–3, which can be plotted for any probabilities once the method for calculating the CI is selected.

This means that the C_1 classifier characterized by a 95% classification accuracy will be considered statistically significantly better than the C_2 classifier with 90% accuracy if the results are computed from a minimum of 459 trials. With 200 trials, it still shows a 29% interference ratio with C_2. After 459 trials, a separation ratio is observed. This **DS**-based discriminating ratio is calculated based on five parameters:

- The type of confidence interval approximation (here Clopper-Pearson)
- The required confidence level (here 95%)
- The two classification accuracies being compared (here 90% and 95%)
- The sample size for which it is evaluated (in this example, 200 trials)

This ratio is calculated using the larger of the two CIs, which is always associated with the smaller success rate. Alternatively, it could have been calculated using the global CI bounds (maximum upper and minimum lower), resulting in a narrower range of values in the $]-1, 1[$ interval. However, this approach would lack intuitive meaning for the denominator.

It should be emphasized that there are always two ratios when comparing between classifiers, one for the positive sample and one for the negative sample, and that they are independent. A significantly better classifier at the confidence level can be determined when both ratios become positive. Therefore, it is necessary to consider both positive and negative distributions and ensure that they intersect (e.g., at 459 trials in the previous example).

If only one of the two samples reaches the minimum size, it can be considered partially or semi-positive (or negative) significant. This may be sufficient for certain applications, yet other metrics have been developed to accommodate different interests and requirements.

5.5 The Complete Deviation-Based Confidence Interval (SDCI) Estimations

From the four values in the confusion matrix, $CM = \begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}$, several cumulative variables (e.g., Positive $P + TP + FN$, Negative $N + TN + FP$, Sample $S = P + N$) and commonly used performance

metrics can be derived (e.g., Accuracy $ACC = T/S$, Sensitivity $tp = TP/P$, Specificity $tn = TN/N$). These metrics are consistently mentioned with a cautionary note, highlighting their qualitative nature. Some of the factors to consider include balanced samples, large sample sizes, and more. However, it is ultimately the user's responsibility to interpret their significance. This can often result in confusing recommendations when it comes to selecting the appropriate metrics.

5.5.1   Analytical Approximation of the Accuracy's CI

The convolution of two single distributions, $B(P, p_0)$ and $B(N, p_1)$, does not have a simple formula unless $p_0 = p_1 = p$. In this case, it results in a known distribution called $B(P + N, p)$, which is often used for accuracy. However, this scenario is only a special case of independent identical distributions (IID) and is rarely encountered in practice. Despite the lack of a straightforward convolution method for two different probability Binomials, current software often approximates it using the Wald-Laplace methods applied to the mean probability value.

To address this vexing issue, we devise a new simple approximation using the CI obtained with the Binomial approximations:

$$\begin{cases} tp = \frac{TP}{P} = \widehat{\frac{TP}{P}} \pm CI_{tp} = \widehat{tp} \pm CI_{tp} \\ tn = \frac{TN}{N} = \widehat{\frac{TN}{N}} \pm CI_{tn} = \widehat{tn} \pm CI_{tn} \end{cases} \tag{5–10}$$

A simple analytic method to compute the accuracy's CI (the most widely used performance metric) is subsequently developed. The assumption that the CI come from a normal distribution yield an induced standard deviation that can be expressed as:

$$\begin{cases} tp = \widehat{tp} \pm \frac{Z\sigma_{tp}}{\sqrt{P}} \\ tn = \widehat{tn} \pm \frac{Z\sigma_{tn}}{\sqrt{N}} \end{cases} \tag{5–11}$$

Thus, yielding the following equivalence between CI and standard deviation:

$$\begin{cases} \sigma_{tp} = \frac{CI_{tp}\sqrt{P}}{Z} \\ \sigma_{tn} = \frac{CI_{tn}\sqrt{N}}{Z} \end{cases} \tag{5–12}$$

And since the variance is characterized by $\mathbb{V}(kX) = k^2 \mathbb{V}(X)$, the corresponding standard variances for TP and TN are: $\sigma_{TP} = P\sigma_{tp}$ and $\sigma_{TN} = N\sigma_{tn}$. Similar approximations with the normal distribution for the accuracy yield its estimation as:

$$ACC = \widehat{ACC} \pm Z_{CL}\frac{\sigma_{ACC}}{\sqrt{S}} \tag{5–13}$$

The value of Z_{CL} is typically dependent on the confidence level (CL), but in most cases is implicitly represented as Z for simplicity. Additionally, Z is influenced by the model distribution, which can be either Normal or Student. The threshold of the sample size to use either distribution has been set to 30. When P and N are on different sides of the threshold, the Student model is retained for both. P and N being independent distributions, the accuracy's variance can be rewritten as:

$$\sigma_{ACC}^2 = \sigma_{TP}^2 + \sigma_{TN}^2 = \frac{CI_{tp}^2 P}{Z^2} + \frac{CI_{tn}^2 N}{Z^2} = \frac{CI_{tp}^2 P + CI_{tn}^2 N}{Z^2} \tag{5–14}$$

These computations all lead to the standard deviation-based confidence interval (**SDCI**) for accuracy estimation:

$$ACC = \frac{\widehat{TP} + \widehat{TN}}{S} \pm \sqrt{\frac{CI_{tp}^2 P + CI_{tn}^2 N}{S}} \tag{5–15}$$

It should be emphasized again that neither CI_{tp} nor CI_{tn} are symmetric around $\widehat{tp}$ and $\widehat{tn}$. For the sake of simplicity, this asymmetry is kept implicit, and therefore only one global expression is required. Let's also highlight that neither CI_{tp} nor CI_{tn} has distinct values when used for the upper and lower bounds.

5.5.2 Comparison of Accuracy's CI for Two Methods

Simulated data is generated for small and medium sample sizes considering different probability cases, including mesial ($p = 0.5$), astride ($p = 0.2$) and distal ($p = 0$, and $p = 0.01$) scenarios. 35 confusion matrices (CM) have automatically been constructed by mixing different total sample sizes, prevalence and probabilities, with the following characteristics:

$$\begin{cases} n = [40, 200, 1000] & samples \\ b = [0.1, 0.5, 0.7, 1] & balances \\ p = [0, 0.01, 0.2, 0.5] & probabilities \end{cases}$$

With $TP = nbp$, $FN = nb - TP$, $TN = np(1 - b)$, and $FP = n(1 - b) - TN$, where all numbers are rounded. The confidence intervals (CI) for the accuracy are computed with our analytical approximation and compared with the one obtained from PyCM in Table 5–2.

Table 5–2 Comparison confidence intervals (CI) obtained from PyCM and our newly developed standard deviation-based confidence interval (SDCI) for accuracy estimation using data simulated from 35 confusion matrices (CM)

P & N Sample size and prob.	TP	FN	FP	TN	ACC	PyCM CI			SDCI	
						SE	Lb	Ub	Lb	Ub
0 P=4; p1=0 and N=36; p2=0	0	4	36	0	0.0	0	0	0.088	0	0.227
1 P=20; p1=0 and N=20; p2=0	0	20	20	0	0.0	0	0	0.088	0	0.179
2 P=28; p1=0 nd N=12; p2=0	0	28	12	0	0.0	0	0	0.088	0	0.19
3 P=4; p1=0 and N=36; p2=0	0	4	36	0	0.0	0	0	0.088	0	0.227
4 P=20; p1=0 and N=20; p2=0	0	20	20	0	0.0	0	0	0.088	0	0.19
5 P=28; p1=0 and N=12; p2=0	0	28	12	0	0.0	0	0	0.088	0	0.19
6 P=4; p1=0.25 and N=36; p2=0.19	1	3	29	7	0.2	0.063	0.105	0.348	0.084	0.429
7 P=20; p1=0.2 and N=20; p2=0.2	4	16	16	4	0.2	0.063	0.105	0.348	0.076	0.431
8 P=28; p1=0.21 and N=12; p2=0.17	6	22	10	2	0.2	0.063	0.105	0.348	0.081	0.434
9 P=4; p1=0.5 and N=36; p2=0.5	2	2	18	18	0.5	0.079	0.352	0.648	0.305	0.695
10 P=20; p1=0.5 and N=20; p2=0.5	10	10	10	10	0.5	0.079	0.352	0.648	0.289	0.711
11 P=28; p1=0.5 and N=12; p2=0.5	14	14	6	6	0.5	0.079	0.352	0.648	0.29	0.71
12 P=20; p1=0 and N=180; p2=0	0	20	180	0	0.0	0	0	0.019	0	0.06
13 P=100; p1=0 and N=100; p2=0	0	100	100	0	0.0	0	0	0.019	0	0.037

Continued on next page

Table 5–2 (continued)

P & N Sample size and prob.		TP	FN	FP	TN	ACC	PyCM CI			SDCI	
							SE	Lb	Ub	Lb	Ub
14	P=140; p1=0 and N=60; p2=0	0	140	60	0	0.0	0	0	0.019	0	0.01
15	P=20; p1=0 and N=180; p2=0.01	0	20	178	2	0.01	0.007	0.003	0.036	0.002	0.073
16	P=100; p1=0.01 and N=100; p2=0.01	1	99	99	1	0.01	0.007	0.003	0.036	0.002	0.054
17	P=140; p1=0.01 and N=60; p2=0.02	0	20	178	2	0.01	0.007	0.003	0.036	0.001	0.058
18	P=20; p1=0.2 and N=180; p2=0.2	4	16	144	36	0.2	0.028	0.15	0.261	0.137	0.295
19	P=100; p1=0.2 and N=100; p2=0.2	20	80	80	20	0.2	0.028	0.15	0.261	0.133	0.289
20	P=140; p1=0.2 and N=60; p2=0.2	28	112	48	12	0.2	0.028	0.15	0.261	0.134	0.289
21	P=20; p1=0.5 and N=180; p2=0.5	10	10	90	90	0.5	0.035	0.431	0.569	0.404	0.596
22	P=100; p1=0.5 and N=100; p2=0.5	50	50	50	50	0.5	0.035	0.431	0.569	0.404	0.596
23	P=140; p1=0.5 and N=60; p2=0.5	70	70	30	30	0.5	0.035	0.431	0.569	0.404	0.596
24	P=100; p1=0 and N=900; p2=0.2	0	100	900	0	0.0	0	0	0.004	0	0.012
25	P=500; p1=0 and N=500; p2=0	0	500	500	0	0.0	0	0	0.004	0	0.008
26	P=700; p1=0 and N=300; p2=0	0	700	300	0	0.0	0	0	0.004	0	0.008
27	P=100; p1=0.01 and N=900; p2=0.01	1	99	891	9	0.1	0.003	0.005	0.018	0.005	0.026
28	P=500; p1=0.01 and N=500; p2=0.01	5	495	495	5	0.1	0.003	0.005	0.018	0.004	0.023
29	P=700; p1=0.01 and N=300; p2=0.01	7	693	297	3	0.1	0.003	0.005	0.018	0.004	0.024

Table 5–2 (continued)

P & N Sample size and prob.		TP	FN	FP	TN	ACC	PyCM CI			SDCI	
							SE	Lb	Ub	Lb	Ub
30	P=100; p1=0.2 and N=900; p2=0.2	20	80	720	180	0.2	0.013	0.176	0.226	0.168	0.238
31	P=500; p1=0.2 and N=500; p2=0.2	100	400	400	100	0.2	0.013	0.176	0.226	0.167	0.237
32	P=700; p1=0.2 and N=300; p2=0.2	140	560	240	60	0.2	0.013	0.176	0.226	0.167	0.237
33	P=100; p1=0.5 and N=900; p2=0.5	50	50	450	450	0.5	0.016	0.469	0.531	0.457	0.543
34	P=500; p1=0.5 and N=500; p2=0.5	250	250	250	250	0.5	0.016	0.469	0.531	0.456	0.544
35	P=700; p1=0.5 and N=300; p2=0.5	350	350	150	150	0.5	0.016	0.469	0.531	0.456	0.544
							2.03%	14.09%	23.63%	12.86%	27.86%

The percentage difference between the PyCM Upper bound (Ub) and Lower bound (Lb) is equal to 9.54%, and to 15% for SDCI. This means that there is a difference of CI widths equal to 5.46%, and thus a mean difference of width difference of 38.38% with PyCM. These differences are mainly concentrated around the low sample size and low probabilities. It is in these areas that the new approximation should be used. Table 5–2 shows results from PyCM based on the Wilson CI.

Similar computations can be carried out with the other presented CI and are summarized in Table 5–3.

Table 5–3 Mean underestimation of CI width difference for six CI approximation methods

CI Approximation method	CI width difference	Underestimation
Clopper	5.89%	38.15%
Wilson	5.46%	36.38%
Jeffrey	3.82%	28.60%
Laplace	5.38%	36.05%
Wald	2.48%	20.60%

5.6 Dream of the Confusion Ellipses (CE)

5.6.1 The Importance of Visual Representations

Figures play a crucial role in research by not only presenting data but also illustrating relationships, methods, and synthesis. Although humans and some animals possess the ability to comprehend complex patterns, their capacity to process more than three or four values simultaneously is often limited. This innate discriminative ability allows for a holistic understanding that associates surfaces, contiguity, and other properties without explicit measurements, such as the distances between eyes or from chin to hairline.

Various types of graphs, including line graphs, histograms, box plots, scatter plots, bar graphs, and pie or donut charts, rely on this inherent ability by replacing puzzling single data values with surfaces and arranging them contiguously to visually assess their relationships. Therefore, determining the most suitable representation could be theoretically achieved by appraising the number of values and relationships conveyed by a given graph using the least amount of quantified information. The aesthetic aspect is also of crucial importance and can be partially addressed through the use of colors and general shapes. For example, donut charts, pie charts, and stacked histograms can convey the same amount of information.

In addition to data comprehension and communication, visualizations enable data- driven decision-making. Validating scientific models or identifying potential biases is facilitated, leading to more robust conclusions or insightful discoveries. The saying "seeing is believing" holds true in this context. However, when it comes to confusion matrices (CM), they are often merely represented with four rectangles and, at best, a color gradient to indicate their relative values. This overly simplistic representation denies researchers the tools to fully comprehend the significance conveyed by its illustration.

5.6.2 Confusion Ellipses (CE)

The CM's four values are typically not immediately useful on their own. Instead, it is the combination of these values that holds interest, such as unbalance, accuracy, or other similar performance metrics. However, these measurements are never visually represented, not even in an intuitive manner. To address this absence, we have developed a new representation called confusion ellipses (**CE**).

The fundamental idea behind **CE** is to replace the equal rectangles by quarter-filled ellipses. The horizontal axis represents the True values based on their sign (positive and negative), while the vertical axis represents fn for the positive side (as it represents real positives) and fp for the negative half-axis. This follows the conventional CM in trigonometric order. The values are normalized by dividing by S, resulting in the total sum $tp + fn + fp + tn = 1$, meaning that all values represent percentages of the total sample.

To clarify how a **CE** is built, let's take a simple example showcasing two CM with the following characteristics: $CM_1 = \begin{bmatrix} 100 & 20 \\ 10 & 30 \end{bmatrix}$ and $CM_2 = \begin{bmatrix} 90 & 30 \\ 15 & 25 \end{bmatrix}$, i.e., with a positive sample $P = 120$ and

a negative sample $N = 40$. To normalize these classifiers, all values should be divided by $S + P + N$, resulting in the four normalized values tp, fn, fp, and tn displayed on the **CE** in Figure 5–4.
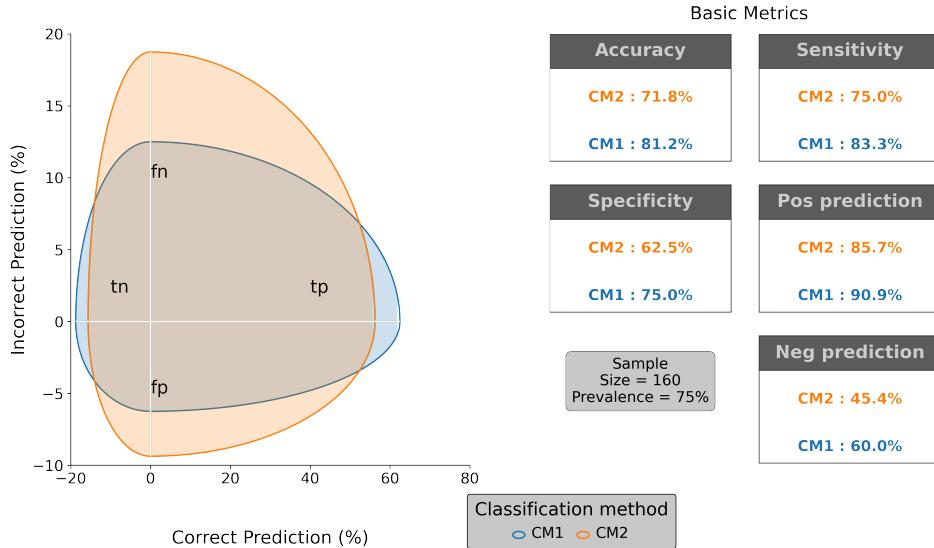
Figure 5–4 Two confusion ellipses (CEs) and associated main metrics

Figure 5–5 Six complementary representations illustrating how CM_1 can be visually identified as better than CM_2 yet not significantly with the current weighted safety; data from illustrative example (except for the violin plots in E that necessitate real data)

Even though the horizontal scale is different from the vertical one, intuitive visual information can be derived from this representation, namely:

- tp, fn, fp, and tn values and relative values for both CM.
- The large unbalanced samples: the lower semi-ellipses are much smaller than the upper ones.
- The prevalence: the first quadrants are much bigger than the third ones.
- The accuracy of CM_1 is higher than that of CM_2 as the major axis is larger.
- More generally, all the metrics given on the right side are visually accessible.

The **CE**, along with the Ad hoc safety, and their corresponding CI, can be represented individually. The here defined Confidence Interval ratio (CIr), which is calculated by including the metric in the CI interference calculations, provides a way to visually and intuitively compare the two classifiers. This also allows for determining the statistical significance of their difference, as shown in Figure 5–5.

Different geometrical shapes, such as rectangles or diamonds, can serve as alternative representations. Nevertheless, they are generally less visually appealing to the human eye, which usually prefers smoother forms over sharp ones, and have thus not been retained for further utilization.

5.7 Summary of Results and Discussion

The seven different CI approximations yield different results for the classification accuracies obtained from the **SDCI**. Their results are compared using the data obtained after **ABCD** and ICA and summarized 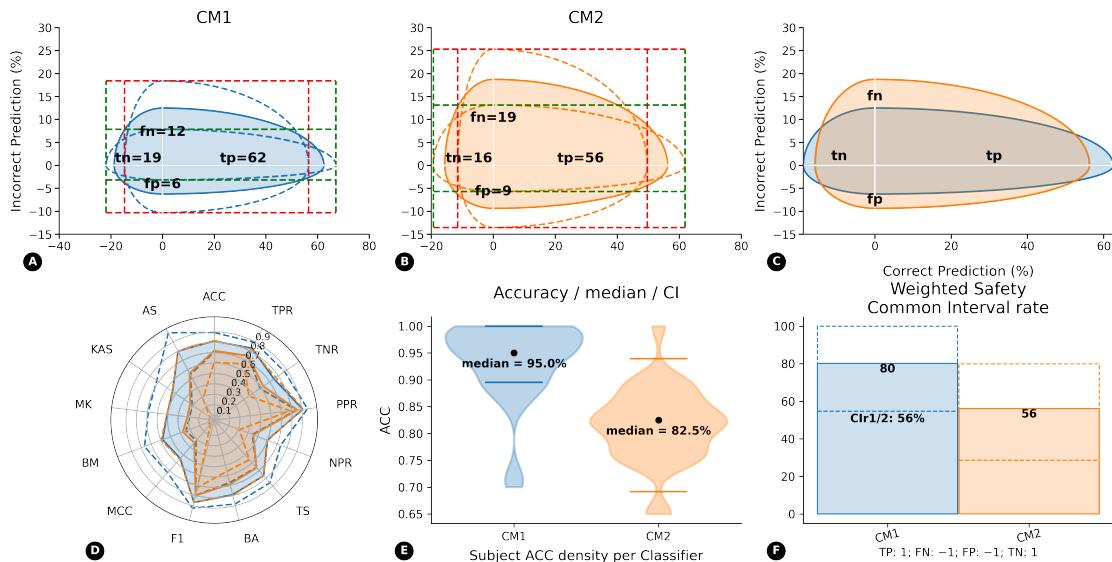in Table 5–4. Notably, the comparison between the CI widths resulting from these two pre-processing algorithms demonstrates yet again the superiority of **ABCD** over ICA.

Table 5–4 Comparison of seven CI approximation influence on CI widths

CI approximation	ABCD accuracy (with CI at 95% confidence level)	ICA accuracy (with CI at 95% confidence level)	ABCD CI width	ICA CI width
Nominal accuracies	93.81%	81.90%		
Laplace with continuity correction	[81.83%; 100%]	[66.48%; 92.64%]	18.17%	26.47%
Wald-Laplace	[87.01%; 99.95%]	[68.67%; 94.18%]	12.94%	25.51%
Clopper-Pearson	[79.77%; 98.32%]	[64.79%; 92.23%]	18.55%	27.44%
Wilson	[79.78%; 97.71%]	[65.27%; 91.00%]	17.93%	25.73%
Agresti-Coull	[78.40%; 98.88%]	[64.67%; 91.45%]	20.48%	26.79%
Jeffreys	[82.49%; 97.91%]	[66.76%; 91.31%]	15.42%	24.55%
Binomial ACI	[86.79%; 98.68%]	[68.87%; 92.65%]	11.89%	23.78%

ABCD not only yields a higher averaged accuracy but also demonstrates a lower CI width with any of the chosen approximations. This results from the lower variance of **ABCD** in individual results. Thus,

instead of detailing the classification accuracy for each subject as was done in Table 5–4, simply displaying the CI width provides the same meaningful information.

These different CI approximations can also be used with the **DS** method to estimate the a priori sample size to obtain a statistically significant difference between two known pipelines. Table 5–5 shows two computations corresponding to different classification accuracies. For example, if one were interested in designing an experiment to compare two algorithms yielding 90% and 95%, the minimum sample size would be an average of around 420 trials. Assuming a study similar to the one described in Chapter 2, a total of 21 sessions, with at least 20 trials per session, would be needed. And only 6 sessions would be required to investigate again **ABCD** vs ICA.

Table 5–5 Comparison of sample size estimation using different Binomial approximations

CI approximation	Sample size for comparison between 90% and 95% accuracies	Sample size for comparison between 81.90% and 93.81% accuracies
Laplace with continuity correction	426	124
Wald-Laplace	413	116
Clopper-Pearson	459	139
Wilson	427	125
Agresti-Coull	437	130
Jeffreys	420	121
Binomial ACI	378	105

5.8   Conclusion and Highlights

This chapter delves into the statistical aspects of classification evaluation by meticulously examining the fundamental mathematical principles that underlie it. The methodologies devised in this chapter possess the capacity to be employed in domains beyond BCI.

Particularly noteworthy is the applicability of the Standard Deviation of Confidence Intervals (**SDCI**) in assessing the superiority of one approach over another in any comparative analysis. The Distance Separation (**DS**) method also emerges as a valuable tool for a priori sample size estimation. Furthermore, the Confusion Ellipses (**CE**) offer promising prospects for evaluating confusion matrices in various comparative scenarios.

Several aspects outlined in this chapter warrant emphasis:

- The snailed Lateral Asymptotic Confidence Intervals (LACI) is a novel representation of the Binomial ACI, offering an improved visualization of the relative constant step of the upper or lower CI

bound.

- Due to its intricate relationship with the Laplace-Bayes estimator, the Clopper-Pearson CI approximation has been selected as the default method for all computations involving confidence intervals.

- Clopper-Pearson provides the most comprehensive explanation for the inherent dual uncertainty associated with determining the probable Binomial distribution and its corresponding CI. An illustrative example has been presented for easier comprehension of its impacts.

- In the domain of experimental design, **DS** emerges as a feasible substitute for **FDMC** (detailed in Chapter 2), especially when the goal is to evaluate the statistical differentiation between two classifiers with predetermined accuracies.

- **SDCI** offers the capability to estimate the CI accuracy from any given confusion matrix. Its utility becomes particularly evident when investigating whether one algorithm exhibits significant superiority over another.

- The **CE** representation facilitates a rapid visual comparison of classifiers, providing a valuable complement to the main performance metrics.

The tools and methodologies designed in this chapter demonstrate the capacity to assess and rigorously compare the performance of classifiers.

Chapter 6   Conclusion and Perspectives

This doctoral dissertation focuses on the enhancement of performance in EEG-based Brain-Computer Interfaces (BCI). The primary objective is to attain the highest possible classification accuracy of user intent, specifically for motor imagery (MI) of the right and left hands. Through the development of subject-specific pre-processing and source localization-derived processing methods, our algorithms are tailored to signals characterized by a low signal-to-noise ratio (SNR) and high inter- and intra-subject variability. Particular attention has been devoted to the systematic utilization of reliable scientific methodologies during the experimental design phase and subsequent analysis of reproducible outcomes, employing consistent statistical approaches. A summary of the contributions made in this thesis is presented in Section 6.1, followed by a comprehensive conclusion in Section 6.2, and a review of the perspectives and future avenues in Section 6.3.

6.1   Summary of the Contributions

The fundamental goal of BCI is to facilitate device control by harnessing diverse brain activity patterns that can be identified and translated into actionable commands. A comprehensive summary of the contributions made in this thesis sheds light on the key advancements and novel insights attained during the course of this research endeavor.

These contributions encompass several significant aspects, including a strong emphasis on reproducibility to strengthen the credibility of the presented findings. Additionally, novel pre-processing techniques tailored to individual subjects are introduced. Given the inherent low SNR of EEG potentials, methods based on source localization are employed to enhance signal extraction and classification. Multiple comparisons are also conducted to determine the optimal pipeline (in terms of filtering and spatial re-referencing), taking into account careful considerations of confidence interval approximations. To simplify the notations, Table 6–1 summarizes the acronyms used for the original methods developed in this work. To easily distinguish between existing algorithms and the new ones introduced in this research, the acronyms are in bold for the original ones.

Contribution 1: Record a unique online-available multimodal dataset with the corresponding code for reproducibility.

The spontaneous blink represents a major physiological disturbance, occurring simultaneously with the neural signal of interest (SOI) with an average probability of 10Given the lack of available online datasets capable of facilitating the development of an accurate blink model, an experiment is designed to simultaneously record eyelid movements using EEG/EMG/EOG, an eye-tracker, and a high-speed camera. The study incorporates four widely employed brain-computer interface (BCI) paradigms, namely Motor

Table 6–1 Nomenclature of the algorithms created in this thesis

ABCD	Adaptive Blink Correction and De-drifting
CCS	Core Channel Selection
CE	Confusion Ellipse
DC	Dual Classifier
DS	Distance Separation
FDMC	Fitted Distribution Monte Carlo
SDCI	Standard Deviation of Confidence Intervals
SLST	Source Localized Spatio-Temporal

Execution (ME), Motor Imagery (MI), Steady-State Visual Evoked Potentials (SSVEP), and P300 signals. To ensure an adequate amount of data collection, **FDMC** is conducted for a priori sample size estimation during a prospective power analysis.

Contribution 2: Design a novel fast ABCD pre-processing algorithm.

The recently devised **ABCD** algorithm demonstrates a remarkable capability to detect and correct individual blinks while preserving the integrity of the signal of interest (SOI). It effectively accommodates inter- and intra-subject variations by customizing its blink models to suit the unique characteristics of each blink. Furthermore, **ABCD** exhibits the ability to eliminate faulty channels (either bridged or malfunctioning). Its low computational load makes it particularly suitable for real-time implementation. A comparative analysis of its efficacy against the prevailing techniques for blink correction, namely Independent Component Analysis (ICA) and Artifact Subspace Reconstruction (ASR), demonstrates that **ABCD** significantly outperforms both ICA and ASR.

Contribution 3: Implement an innovative SLST feature extraction method coupled with a DC.

The **SLST** technique is able to reveal spatio-temporal characteristics of the SOI by analyzing the similarities across multiple trials. The ensuing spatial locations and time- derived covariance matrices are used as input to the DC algorithm. The class-specific Fréchet means, along with all other covariance matrices, are computed at the **CCS**, which is determined based on a comprehensive meta-analysis. Comparisons across different frequency bands, filter types, and spatial re-referencing are also carried out to reveal the optimized combination of all these processing steps. The newly developed **SLST+DC** demonstrates superior performance compared to over tested combinations of feature extraction and classification algorithms. Specifically, it outperforms Common Spatial Patterns (CSP) combined with Linear Discriminant Analysis (LDA) or with Support Vector Machine (SVM), as well as more recent algorithms also based on Riemannian geometry, such as Minimum Distance to Riemannian Mean (MDRM) or Tangent Space (TS).

Contribution 4: Develop a new accuracy's estimation using SDCI applied to CE. An original method for estimating the accuracy through **SDCI** is developed using the fact that the two independent

samples comprising a confusion matrix (CM) follow a Binomial distribution. **DS** is also developed to estimate the sample size when the purpose of a new experiment is to compare different algorithms with known classification accuracies. The results from **SDCI** are applied to innovative visual representations of CM, called **CE**.

Our contributions, summarized in Table 6–2, aim to advance the field of BCI research by addressing challenges in signal processing, reproducibility, and accuracy estimation.

Table 6–2 Thesis summary and contribution. Contemporary EEG-based BCIs are afflicted by five primary drawbacks. To address these issues, we introduce a range of novel methodologies, which are categorized into four chapters encompassing experimental design, pre-processing, processing, and evaluation domains

Chapter	Methodological contribution	Scientific contribution
2	A quantitative method to determine the a priori sample size based on a prospective power analysis and using fitted distribution	The large online-available multi-modal dataset can be used to analyze eye-related movements or evaluate pre-processing and processing algorithms. The **FDMC** simulation can be applied to a priori sample size estimation
3	An automated model-based approach for accurate detection and correction of blinks, along with the identification of bad channels	**ABCD** significantly outperforms both ICA and ASR and can be used in real-time fashion
4	Automated source localization-derived methods for feature extraction and classification	**SLST**-derived features coupled with **DC** give consistently superior results compared to other existing methods, such as CSP+LDA, CSP+SVM, MDRM, or TS
5	Quantitative and visual methods to estimate and represent the classification accuracy's confidence intervals, as well as complementary a priori sample size estimation	**SDCI** can estimate the classification accuracy and be applied to **CE**. The **DS** method can also be used for a priori sample size estimation when classification accuracies of pipelines to be tested are known

6.2 Conclusions

This thesis journey towards achieving a rigorous, reproducible, and efficient pipeline has been a remarkable voyage of discovery, comprehension, and occasional challenges to establish procedures and methodologies. Each stage has necessitated a meticulous examination of existing approaches, which, while not exhaustive, have frequently furnished the necessary evidence to validate our methods and their underlying principles.

Notable examples include the application of ICA or ASR within the domain of blind source separation (BSS), the remarkable yet logical results obtained through source localization, and the various approximations of the binomial confidence intervals (CIs).

Our overarching objective throughout this endeavor has been to maintain an open- minded and innovative approach, seeking alternative solutions or tools whenever existing options proved unsatisfactory at each step. For instance, this has led us to develop novel concepts for representation and devise new tools for measuring the distance between two CIs.

The most significant conclusion drawn from this expedition is the boundless potential for progress in each respective field. It underscores the imperative to justify choices not solely based on convention or consensus but rather through a profound and comprehensive analysis, even when confronted with obstacles and uncertainties. In essence, this thesis "odyssey" has been both stimulating and fruitful, and it is this spirit of continuous improvement that should guide future endeavors.

6.3 Perspectives

In light of our findings, it is clear that significant prospects for enhancement exist within each of the domains addressed. Future objectives will be to investigate these opportunities with a comprehensive approach while concurrently pursuing the overarching global objective of advancing BCI technology. These granular evolutions can subsequently be applied within their respective domains, contributing to the comprehension and development of a pathway towards achieving the "BCI revolution".

Several of these endeavors, though not exhaustively enumerated, are intended to be fulfilled. First and foremost, **ABCD** should be tested in a real-time fashion to evaluate its rapidity and efficiency for online applications. Statistical bias correction in each class should also be improved, and a quantification of the bias and variance offered by this method should be designed. More distributions, e.g., the truncated Normal one, are also expected to be integrated into the **FDMC** method.

The **ABCD+SLST+DC** pipeline is currently being tested on the remaining paradigms presented in this thesis. Similar comparisons across filters and other related parameters will confirm the superiority of the specific choices outlined in this thesis across different tasks.

SLST should also be adapted to integrate not only the source localization results from the maxi-

mally activated brain source. More research is needed to determine whether a couple (or more) of locally maximally activated brain sources would be enough to improve the resulting classification results. Mathematical solutions using block-localized matrices with border constraints could also be tested to improve the calculation times. **CCS** could also be investigated for more useful paradigms, for example, the error-related negativity.

A double bootstrap procedure could also be developed to emulate the accuracy and compare it to the **SDCI** and other existing Binomial CI approximations. It would also be interesting to develop a quantified tool for hierarchizing the quality of a representation according to some of the basic human ergonomic capabilities. Finally, the development of a coverage adaptable low-degree polynomial approximation of the CI is also planned.

More generally, an essential aspect of future research for EEG-based BCIs involves the incorporation of multimodal comparisons (e.g., with invasive recordings from ECoG or non- invasive ones from MEG). While EEG has undeniable advantages, the interpretation of the detected signals can be influenced by various factors, including signal attenuation or artifacts, to only cite a few.

Therefore, it is crucial to confirm through a multimodal approach whether the observed changes in EEG signals are genuinely reflective of the task-induced neural activity or if they are confounded by unrelated factors. Additionally, multimodal comparison can facilitate the refinement and validation of signal processing techniques by confirming or adapting the different hypotheses postulated (e.g., the minimum time interval to detect spatio-temporal similarities).

Lastly, the integration of BCI technology into daily life activities, such as controlling assistive devices or interacting with virtual environments, holds immense potential and breaks free from classical study designs or traditional paradigms. This requires addressing challenges related to system portability, user comfort, and real-time adaptability. Exploring ways to improve the usability, user experience, and robustness of BCI systems in real-world scenarios will be crucial to unlocking their full potential.

Bibliography

[1] GRAIMANN B, ALLISON B Z, PFURTSCHELLER G. Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction[M]. Springer Science & Business Media, 2010.

[2] PURKAYASTHA S S, JAIN V K, SARDANA H K. Topical Review: A Review of Various Techniques Used for Measuring Brain Activity in Brain Computer Interfaces [J]. Advances in Electronic and Electric Engineering, 2014, 4(5): 513-522.

[3] LORACH H, GALVEZ A, SPAGNOLO V, et al. Walking naturally after spinal cord injury using a brain‑spine interface[J]. Nature, 2023, 618(7963): 126-133. DOI: 10 .1038/s41586-023-06094-5.

[4] SINGH A, HUSSAIN A A, LAL S, et al. A Comprehensive Review on Critical Issues and Possible Solutions of Motor Imagery Based Electroencephalography Brain-Computer Interface[J]. Sensors, 2021, 21(6): 2173. DOI: 10.3390/s21062173.

[5] PASCUAL-MARQUI R D, ESSLEN M, KOCHI K, et al. Functional imaging with low resolution brain electromagnetic tomography (LORETA): Review, new comparisons, and new validation[J]. Lehmann, 2002, 30: 81-94.

[6] PONTIFEX M B, GWIZDALA K L, PARKS A C, et al. Variability of ICA decomposition may impact EEG signals when used to remove eyeblink artifacts[J]. Psychophysiology, 2017, 54(3): 386-398. DOI: 10.1111/psyp.12804.

[7] ELLIS P D. The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results[M/OL]. Cambridge: Cambridge University Press, 2010. https://www.cambridge.org/core/books/essential-guide-to-effect-s izes/72C26CA99366A19CAC4EF5B16AE3297F.

[8] POLDRACK R A, BAKER C I, DURNEZ J, et al. Scanning the horizon: towards transparent and reproducible neuroimaging research[J]. Nature Reviews Neuroscience, 2017, 18(2): 115-126. DOI: 10.1038/nrn.2016.167.

[9] LARSON M J, CARBINE K A. Sample size calculations in human electrophysiology (EEG and ERP) studies: A systematic review and recommendations for increased rigor[J]. International Journal of Psychophysiology. Rigor and Replication: Towards Improved Best Practices in Psychophysiological Research 2017, 111: 33-41. DOI:

10.1016/j.ijpsycho.2016.06.015.

[10] PASCUAL-MARQUI R D, LEHMANN D, KOUKKOU M, et al. Assessing interactions in the brain with exact low-resolution electromagnetic tomography[J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2011, 369(1952): 3768-3784. DOI: 10.1098/rsta.2011.0081.

[11] SERAJ E, SAMENI R. Robust electroencephalogram phase estimation with applications in brain-computer interface systems[J]. Physiological Measurement, 2017, 38(3): 501. DOI: 10.1088/1361-6579/aa5bba.

[12] BOONYAKITANONT P, LEK-UTHAI A, CHOMTHO K, et al. A review of feature extraction and performance evaluation in epileptic seizure detection using EEG[J]. Biomedical Signal Processing and Control, 2020, 57: 101702. DOI: 10.1016/j.bspc.2019.101702.

[13] SAEIDI M, KARWOWSKI W, FARAHANI F V, et al. Neural Decoding of EEG Signals with Machine Learning: A Systematic Review[J]. Brain Sciences, 2021, 11(11): 1525. DOI: 10.3390/brainsci11111525.

[14] DALY I, MATRAN-FERNANDEZ A, VALERIANI D, et al. Editorial: Datasets for Brain-Computer Interface Applications[J/OL]. Frontiers in Neuroscience, 2021, 15. https://www.frontiersin.org/articles/10.3389/fnins.2021.732165.

[15] COHEN J. The statistical power of abnormal-social psychological research: a review [J]. Journal of Abnormal and Social Psychology, 1962, 65: 145-153. DOI: 10.1037/h0045186.

[16] Van BELLE G. Statistical Rules of Thumb: vol. 699[M]. John Wiley & Sons, 2011.

[17] KRUSE A, SUICA Z, TAEYMANS J, et al. Effect of brain-computer interface training based on non-invasive electroencephalography using motor imagery on functional recovery after stroke - a systematic review and meta-analysis[J]. BMC Neurology, 2020, 20(1): 385. DOI: 10.1186/s12883-020-01960-5.

[18] MANSOUR S, ANG K K, NAIR K P, et al. Efficacy of Brain–Computer Interface and the Impact of Its Design Characteristics on Poststroke Upper-limb Rehabilitation: A Systematic Review and Meta-analysis of Randomized Controlled Trials[J]. Clinical EEG and Neuroscience, 2022, 53(1): 79-90. DOI: 10.1177/15500594211009065.

[19] WIERZGAŁA P, ZAPAŁA D, WOJCIK G M, et al. Most Popular Signal Processing

Methods in Motor-Imagery BCI: A Review and Meta-Analysis[J / OL]. Frontiers in Neuroinformatics, 2018, 12. https://www.frontiersin.org/articles/10.3389/fninf.201 8.00078.

[20] SONG J, DAVEY C, POULSEN C, et al. EEG source localization: Sensor density and head surface coverage[J]. Journal of Neuroscience Methods, 2015, 256: 9-21. DOI: 10.1016/j.jneumeth.2015.08.015.

[21] COHEN J. Statistical Power Analysis for the Behavioral Sciences[M]. 2nd ed. New York: Routledge, 1988.

[22] FAUL F, ERDFELDER E, BUCHNER A, et al. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses[J]. Behavior Research Methods, 2009, 41(4): 1149-1160. DOI: 10.3758/BRM.41.4.1149.

[23] MOSS J. univariateML: An R package for maximum likelihood estimation of uni-variate densities[J]. Journal of Open Source Software, 2019, 4(44): 1863. DOI: 10.2 1105/joss.01863.

[24] TASHPULATOV S N. Modeling Electricity Price Dynamics Using Flexible Distri-butions[J]. Mathematics, 2022, 10(10): 1757. DOI: 10.3390/math10101757.

[25] REEVES R R, STRUVE F A, PATRICK G. A Comprehensive Questionnaire for Sub-jects Undergoing Quantitative Research EEGs[J]. Clinical Electroencephalography, 1998, 29(2): 67-72. DOI: 10.1177/155005949802900204.

[26] OLDFIELD R C. The assessment and analysis of handedness: The Edinburgh inven-tory[J]. Neuropsychologia, 1971, 9(1): 97-113. DOI: 10.1016/0028-3932(71)90067 -4.

[27] SAGONAS C, ANTONAKOS E, TZIMIROPOULOS G, et al. 300 Faces In-The-Wild Challenge: database and results[J]. Image and Vision Computing. 300-W, the First Automatic Facial Landmark Detection in-the-Wild Challenge 2016, 47: 3-18. DOI: 10.1016/j.imavis.2016.01.002.

[28] FARWELL L A, DONCHIN E. Talking off the top of your head: toward a men-tal prosthesis utilizing event-related brain potentials[J]. Electroencephalography and Clinical Neurophysiology, 1988, 70(6): 510-523. DOI: 10.1016/0013-4694(88)901 49-6.

[29] TOWNSEND G, LAPALLO B K, BOULAY C B, et al. A novel P300-based

brain-computer interface stimulus presentation paradigm: Moving beyond rows and columns[J]. Clinical Neurophysiology, 2010, 121(7): 1109-1120. DOI: 10.1016/j.clinph.2010.01.030.

[30] CECOTTI H, RIVET B. One step beyond rows and columns flashes in the P300 speller: a theoretical description[J]. International Journal of bioelectromagnetism, 2010, 13(1): 39-41.

[31] SCHÖNER G, KELSO J A S. Dynamic Pattern Generation in Behavioral and Neural Systems[J]. Science, 1988, 239(4847): 1513-1520. DOI: 10.1126/science.3281253.

[32] GRABEN B P, SADDY J D, SCHLESEWSKY M, et al. Symbolic dynamics of event-related brain potentials[J]. Physical Review E, 2000, 62(4): 5518-5541. DOI: 10.1103/PhysRevE.62.5518.

[33] GRABEN B P. Estimating and improving the signal-to-noise ratio of time series by symbolic dynamics[J]. Physical Review E, 2001, 64(5): 051104. DOI: 10.1103/PhysRevE.64.051104.

[34] JOHNSON R A, WICHERN D W. Applied Multivariate Statistical Analysis[M]. Pearson Prentice Hall, 2007.

[35] BELL C. On the Motions of the Eye, in Illustration of the Uses of the Muscles and Nerves of the Orbit[J]. Philosophical Transactions of The Royal Society of London, 1823, 113: 166-186. DOI: 10.1098/rstl.1823.0017.

[36] IWASAKI M, KELLINGHAUS C, ALEXOPOULOS A V, et al. Effects of eyelid closure, blinks, and eye movements on the electroencephalogram[J]. Clinical Neurophysiology, 2005, 116(4): 878-885. DOI: 10.1016/j.clinph.2004.11.001.

[37] COLLEWIJN H, van der STEEN J, STEINMAN R M. Human eye movements associated with blinks and prolonged eyelid closure[J]. Journal of Neurophysiology, 1985, 54(1): 11-27. DOI: 10.1152/jn.1985.54.1.11.

[38] BARRY W, MELVILL JONES G. Influence of Eye Lid Movement Upon Electro-Oculographic Recording of Vertical Eye Movements[J]. Aerospace medicine, 1965, 36: 855-8.

[39] MAKEIG S, BELL A, JUNG T P, et al. Independent Component Analysis of Electroencephalographic Data[C/OL]//: vol. 8. MIT Press, 1995. https://proceedings.neurips.cc/paper_files/paper/1995/hash/754dda4b1ba34c6fa89716b85d68532b-Abst

ract.html.

[40] MULLEN T, KOTHE C, CHI Y M, et al. Real-Time Modeling and 3D Visualization of Source Dynamics and Connectivity Using Wearable EEG[J]. Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2013, 2013: 2184-2187. DOI: 10.1109/EMBC.2013.6609968.

[41] ROMERO S, MAÑANAS M A, BARBANOJ M J. A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: A simulation case[J]. Computers in Biology and Medicine, 2008, 38(3): 348-360. DOI: 10.1016/j.compbiomed.2007.12.001.

[42] BENTIVOGLIO A R, BRESSMAN S B, CASSETTA E, et al. Analysis of blink rate patterns in normal subjects[J]. Movement Disorders, 1997, 12(6): 1028-1034. DOI: 10.1002/mds.870120629.

[43] NARA K. Removal of Ocular Artifacts from EEG Signal using Joint Approximate Diagonalization of Eigen Matrices (JADE) and Wavelet Transform[C]//4: vol. 1. 2010.

[44] EVANS A, COLLINS D, MILLS S, et al. 3D statistical neuroanatomical models from 305 MRI volumes[C]//. 1993: 1813-1817 vol.3. DOI: 10.1109/NSSMIC.1993.373 602.

[45] DOANE M G. Interaction of Eyelids and Tears in Corneal Wetting and the Dynamics of the Normal Human Eyeblink[J]. American Journal of Ophthalmology, 1980, 89(4): 507-516. DOI: 10.1016/0002-9394(80)90058-6.

[46] CRUZ A A V, GARCIA D M, PINTO C T, et al. Spontaneous Eyeblink Activity[J]. The Ocular Surface, 2011, 9(1): 29-41. DOI: 10.1016/S1542-0124(11)70007-6.

[47] PATIL I. Visualizations with statistical details: The 'ggstatsplot' approach[J]. Journal of Open Source Software, 2021, 6(61): 3167. DOI: 10.21105/joss.03167.

[48] DELORME A, MAKEIG S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis[J]. Journal of Neuroscience Methods, 2004, 134(1): 9-21. DOI: 10.1016/j.jneumeth.2003.10.009.

[49] LIU C C, GHOSH HAJRA S, PAWLOWSKI G, et al. Differential neural processing of spontaneous blinking under visual and auditory sensory environments: An EEG

investigation of blink-related oscillations[J]. NeuroImage, 2020, 218: 116879. DOI: 10.1016/j.neuroimage.2020.116879.

[50] PION-TONACHINI L, KREUTZ-DELGADO K, MAKEIG S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website[J]. NeuroImage, 2019, 198: 181-197. DOI: 10.1016/j.neuroimage.2019.05.026.

[51] GRAMFORT A, LUESSI M, LARSON E, et al. MEG and EEG data analysis with MNE-Python[J/OL]. Frontiers in Neuroscience, 2013, 7. https://www.frontiersin.org/articles/10.3389/fnins.2013.00267.

[52] SACCHET M D, KNUTSON B. Spatial smoothing systematically biases the localization of reward-related brain activity[J]. NeuroImage, 2013, 0: 270-277. DOI: 10.1016/j.neuroimage.2012.10.056.

[53] LIBERATI A, ALTMAN D G, TETZLAFF J, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration[J]. PLOS Medicine, 2009, 6(7): e1000100. DOI: 10.1371/journal.pmed.1000100.

[54] ECCLESTON C, WILLIAMS A C D C, MORLEY S. Psychological therapies for the management of chronic pain (excluding headache) in adults[J/OL]. Cochrane Database of Systematic Reviews, 2009(2). https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD007407.pub2/full?cookiesEnabled. DOI: 10.1002/14651858.CD007407.pub2.

[55] SNYDER S M, HALL J R. A Meta-analysis of Quantitative EEG Power Associated With Attention-Deficit Hyperactivity Disorder[J]. Journal of Clinical Neurophysiology, 2006, 23(5): 441-456. DOI: 10.1097/01.wnp.0000221363.12503.78.

[56] MULLIN J P, SHRIVER M, ALOMAR S, et al. Is SEEG safe? A systematic review and meta-analysis of stereo-electroencephalography‐related complications[J]. Epilepsia, 2016, 57(3): 386-401. DOI: 10.1111/epi.13298.

[57] ARNS M, CONNERS C K, KRAEMER H C. A Decade of EEG Theta/Beta Ratio Research in ADHD: A Meta-Analysis[J]. Journal of Attention Disorders, 2013, 17(5): 374-383. DOI: 10.1177/1087054712460087.

[58] HUNTER J E, SCHMIDT F L. Fixed Effects vs. Random Effects Meta-Analysis Models: Implications for Cumulative Research Knowledge[J]. International Journal

of Selection and Assessment, 2000, 8(4): 275-292. DOI: https://doi.org/10.1111/14 68-2389.00156.

[59] HEDGES L V, PIGOTT T D. The power of statistical tests in meta-analysis[J]. Psychological Methods, 2001, 6(3): 203-217. DOI: 10.1037/1082-989X.6.3.203.

[60] VALENTINE J C, PIGOTT T D, ROTHSTEIN H R. How Many Studies Do You Need?: A Primer on Statistical Power for Meta-Analysis[J]. Journal of Educational and Behavioral Statistics, 2010, 35(2): 215-247. DOI: 10.3102/1076998609346961.

[61] JURCAK V, TSUZUKI D, DAN I. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems[J]. NeuroImage, 2007, 34(4): 1600-1611. DOI: 10.1016/j.neuroimage.2006.09.024.

[62] HALDER S, AGORASTOS D, VEIT R, et al. Neural mechanisms of brain–computer interface control[J]. NeuroImage, 2011, 55(4): 1779-1790. DOI: 10.1016/j.neuroimage.2011.01.021.

[63] KOESSLER L, MAILLARD L, BENHADID A, et al. Automated cortical projection of EEG sensors: Anatomical correlation via the international 10–10 system[J]. NeuroImage, 2009, 46(1): 64-72. DOI: 10.1016/j.neuroimage.2009.02.006.

[64] SCHIPPERS M B, KEYSERS C. Mapping the flow of information within the putative mirror neuron system during gesture observation[J]. NeuroImage, 2011, 57(1): 37-44. DOI: 10.1016/j.neuroimage.2011.02.018.

[65] VARNEY N R, DAMASIO H. Locus of Lesion in Impaired Pantomime Recognition [J]. Cortex, 1987, 23(4): 699-703. DOI: 10.1016/S0010-9452(87)80061-8.

[66] DANCKERT J, FERBER S, DOHERTY T, et al. Selective, Non-lateralized Impairment of Motor Imagery Following Right Parietal Damage[J]. Neurocase, 2002, 8(3): 194-204. DOI: 10.1093/neucas/8.3.194.

[67] DRESEL C, CASTROP F, HASLINGER B, et al. The functional neuroanatomy of coordinated orofacial movements: Sparse sampling fMRI of whistling[J]. NeuroImage, 2005, 28(3): 588-597. DOI: 10.1016/j.neuroimage.2005.06.021.

[68] KRÓLICZAK G, PIPER B J, FREY S H. Specialization of the left supramarginal gyrus for hand-independent praxis representation is not related to hand dominance [J]. Neuropsychologia. The Neural Bases of Hemispheric Specialisation 2016, 93: 501-512. DOI: 10.1016/j.neuropsychologia.2016.03.023.

[69] GERARDIN E, SIRIGU A, LEHÉRICY S, et al. Partially Overlapping Neural Networks for Real and Imagined Hand Movements[J]. Cerebral Cortex, 2000, 10(11): 1093-1104. DOI: 10.1093/cercor/10.11.1093.

[70] STIPPICH C, OCHMANN H, SARTOR K. Somatotopic mapping of the human primary sensorimotor cortex during motor imagery and motor execution by functional magnetic resonance imaging[J]. Neuroscience Letters, 2002, 331(1): 50-54. DOI: 10.1016/S0304-3940(02)00826-1.

[71] KRAEUTNER S, GIONFRIDDO A, BARDOUILLE T, et al. Motor imagery-based brain activity parallels that of motor execution: Evidence from magnetic source imaging of cortical oscillations[J]. Brain Research, 2014, 1588: 81-91. DOI: 10.1016/j.brainres.2014.09.001.

[72] BURIANOVÁ H, MARSTALLER L, SOWMAN P, et al. Multimodal functional imaging of motor imagery using a novel paradigm[J]. NeuroImage, 2013, 71: 50-58. DOI: 10.1016/j.neuroimage.2013.01.001.

[73] BEGLIOMINI C, SARTORI L, DI BONO M G, et al. The Neural Correlates of Grasping in Left-Handers: When Handedness Does Not Matter[J/OL]. Frontiers in Neuroscience, 2018, 12. https://www.frontiersin.org/articles/10.3389/fnins.2018.00192/full. DOI: 10.3389/fnins.2018.00192.

[74] BEGLIOMINI C, NELINI C, CARIA A, et al. Cortical Activations in Humans Grasp-Related Areas Depend on Hand Used and Handedness[J/OL]. PLoS ONE, 2008, 3(10). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2561002/. DOI: 10.1371/journal.pone.0003388.

[75] HARDWICK R M, CASPERS S, EICKHOFF S B, et al. Neural Correlates of Motor Imagery, Action Observation, and Movement Execution: A Comparison Across Quantitative Meta-Analyses[J]. bioRxiv, 2017: 198432. DOI: 10.1101/198432.

[76] VARGAS C, OLIVIER E, CRAIGHERO L, et al. The Influence of Hand Posture on Corticospinal Excitability during Motor Imagery: A Transcranial Magnetic Stimulation Study[J]. Cerebral Cortex, 2004, 14(11): 1200-1206. DOI: 10.1093/cercor/bhh080.

[77] WARD N S, FRACKOWIAK R S J. Age-related changes in the neural correlates of motor performance[J]. Brain, 2003, 126(4): 873-888. DOI: 10.1093/brain/awg071.

[78]   CRAMMOND D J. Motor imagery: never in your wildest dream[J]. Trends in Neurosciences, 1997, 20(2): 54-57. DOI: 10.1016/S0166-2236(96)30019-2.

[79]   RIZZOLATTI G, CRAIGHERO L. The mirror-neuron system[J]. Annual Review of Neuroscience, 2004, 27(1): 169-192. DOI: 10.1146/annurev.neuro.27.070203.1442 30.

[80]   GRAFTON S T, ARBIB M A, FADIGA L, et al. Localization of grasp representations in humans by positron emission tomography[J]. Experimental Brain Research, 1996, 112(1): 103-111. DOI: 10.1007/BF00227183.

[81]   WILLEMS R M, TONI I, HAGOORT P, et al. Body-specific motor imagery of hand actions: neural evidence from right- and left-handers[J/OL]. Frontiers in Human Neuroscience, 2009, 3. https://www.frontiersin.org/articles/10.3389/neuro.09.039.2 009/full. DOI: 10.3389/neuro.09.039.2009.

[82]   MOKIENKO O A, CHERVYAKOV A V, KULIKOVA S N, et al. Increased motor cortex excitability during motor imagery in brain-computer interface trained subjects [J/OL]. Frontiers in Computational Neuroscience, 2013, 7(2013-11-22). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3837244/. DOI: 10.3389/fncom.2013.00168.

[83]   PORRO C A, CETTOLO V, FRANCESCATO M P, et al. Ipsilateral involvement of primary motor cortex during motor imagery[J]. European Journal of Neuroscience, 2000, 12(8): 3059-3063. DOI: 10.1046/j.1460-9568.2000.00182.x.

[84]   KROLAK-SALMON P, HÉNAFF M A, TALLON-BAUDRY C, et al. Human lateral geniculate nucleus and visual cortex respond to screen flicker[J]. Annals of Neurology, 2003, 53(1): 73-80. DOI: 10.1002/ana.10403.

[85]   RUSSO F D, PITZALIS S, APRILE T, et al. Spatiotemporal analysis of the cortical sources of the steady-state visual evoked potential[J]. Human Brain Mapping, 2007, 28(4): 323-334. DOI: 10.1002/hbm.20276.

[86]   SRINIVASAN R, FORNARI E, KNYAZEVA M G, et al. fMRI responses in medial frontal cortex that depend on the temporal frequency of visual input[J]. Experimental Brain Research, 2007, 180(4): 677-691. DOI: 10.1007/s00221-007-0886-3.

[87]   SRINIVASAN R, BIBI F A, NUNEZ P L. Steady-state visual evoked potentials: distributed local sources and wave-like dynamics are sensitive to flicker frequency [J]. Brain topography, 2006, 18(3): 167-187. DOI: 10.1007/s10548-006-0267-4.

[88] LUU P, FERREE T. Determination of the HydroCel Geodesic Sensor Nets' Average Electrode Positions and Their 10–10 International Equivalents[J]. Technical Note, Electrical Geodesics, Inc. 11.

[89] NUNEZ P L, NUNEZ E P O B E P L, SRINIVASAN R, et al. Electric Fields of the Brain: The Neurophysics of EEG[M]. Oxford University Press, 2006.

[90] ZHU D, BIEGER J, GARCIA MOLINA G, et al. A Survey of Stimulation Methods Used in SSVEP-Based BCIs[EB/OL]. (2010-03-07). https://www.hindawi.com/journals/cin/2010/702357/.

[91] REGAN D. Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine[M]. New York: Elsevier, 1989.

[92] SCHMOLESKY M T, WANG Y, HANES D P, et al. Signal Timing Across the Macaque Visual System[J]. Journal of Neurophysiology, 1998, 79(6): 3272-3278. DOI: 10.1152/jn.1998.79.6.3272.

[93] NICOLAS-ALONSO L F, GOMEZ-GIL J. Brain computer interfaces, a review[J]. Sensors (Basel, Switzerland), 2012, 12(2): 1211-1279. DOI: 10.3390/s120201211.

[94] PERLSTEIN W M, COLE M A, LARSON M, et al. Steady-state visual evoked potentials reveal frontally-mediated working memory activity in humans[J]. Neuroscience Letters, 2003, 342(3): 191-195. DOI: 10.1016/S0304-3940(03)00226-X.

[95] MARTINEZ P, BAKARDJIAN H, CICHOCKI A. Fully Online Multicommand Brain-Computer Interface with Visual Neurofeedback Using SSVEP Paradigm [EB/OL]. (2007-07-26). https://www.hindawi.com/journals/cin/2007/094561/.

[96] SPENCER K M, DIEN J, DONCHIN E. Spatiotemporal analysis of the late ERP responses to deviant stimuli[J]. Psychophysiology, 2001, 38(2): 343-358. DOI: 10.1111/1469-8986.3820343.

[97] GAETA H, FRIEDMAN D, HUNT G. Stimulus characteristics and task category dissociate the anterior and posterior aspects of the novelty P3[J]. Psychophysiology, 2003, 40(2): 198-208. DOI: 10.1111/1469-8986.00022.

[98] FJELL A M, WALHOVD K B. Life-span changes in P3a[J]. Psychophysiology, 2004, 41(4): 575-583. DOI: 10.1111/j.1469-8986.2004.00177.x.

[99] WEST R, SCHWARB H, JOHNSON B N. The influence of age and individual differences in executive function on stimulus processing in the oddball task[J]. Cortex.

The Cognitive Neuroscience of Aging 2010, 46(4): 550-563. DOI: 10.1016/j.cortex
.2009.08.001.

[100]   MENON V, FORD J M, LIM K O, et al. Combined event-related fMRI and EEG ev-
idence for temporal—parietal cortex activation during target detection[J]. NeuroRe-
port, 1997, 8(14): 3029-3037.

[101]   FRIEDMAN D. Cognition and Aging: A Highly Selective Overview of Event-
Related Potential (ERP) Data[J]. Journal of Clinical and Experimental Neuropsy-
chology, 2003, 25(5): 702-720. DOI: 10.1076/jcen.25.5.702.14578.

[102]   ROSSINI P M, ROSSI S, BABILONI C, et al. Clinical neurophysiology of aging
brain: From normal aging to neurodegeneration[J]. Progress in Neurobiology, 2007,
83(6): 375-400. DOI: 10.1016/j.pneurobio.2007.07.010.

[103]   RAMIREZ-QUINTANA J A, MADRID-HERRERA L, CHACON-MURGUIA M I,
et al. Brain-Computer Interface System Based on P300 Processing with Convolu-
tional Neural Network, Novel Speller, and Low Number of Electrodes[J/OL]. Cog-
nitive Computation, 2020(2020-06-18). https://doi.org/10.1007/s12559-020-09744
-2. DOI: 10.1007/s12559-020-09744-2.

[104]   ARDEKANI B A, CHOI S J, HOSSEIN-ZADEH G A, et al. Functional magnetic
resonance imaging of brain activity in the visual oddball task[J]. Cognitive Brain
Research, 2002, 14(3): 347-356. DOI: 10.1016/S0926-6410(02)00137-4.

[105]   MULERT C, POGARELL O, JUCKEL G, et al. The neural basis of the P300 poten-
tial[J]. European Archives of Psychiatry and Clinical Neuroscience, 2004, 254(3):
190-198. DOI: 10.1007/s00406-004-0469-2.

[106]   CLARK V P, FANNON S, LAI S, et al. Responses to Rare Visual Target and Distrac-
tor Stimuli Using Event-Related fMRI[J]. Journal of Neurophysiology, 2000, 83(5):
3133-3139. DOI: 10.1152/jn.2000.83.5.3133.

[107]   BLEDOWSKI C, PRVULOVIC D, HOECHSTETTER K, et al. Localizing P300
Generators in Visual Target and Distractor Processing: A Combined Event-Related
Potential and Functional Magnetic Resonance Imaging Study[J]. Journal of Neuro-
science, 2004, 24(42): 9353-9360. DOI: 10.1523/JNEUROSCI.1897-04.2004.

[108]   IKEGAMI S M, TAKANO K, WADA M M, et al. Effect of the Green/Blue Flicker
Matrix for P300-Based Brain–Computer Interface: An EEG–fMRI Study[J/OL].

Frontiers in Neurology, 2012, 3. https://www.frontiersin.org/articles/10.3389/fneur.2012.00113/full. DOI: 10.3389/fneur.2012.00113.

[109]　BÉNAR C G, GROVA C, KOBAYASHI E, et al. EEG–fMRI of epileptic spikes: Concordance with EEG source localization and intracranial EEG[J/OL]. NeuroImage, 2006, 30. https://www.sciencedirect.com/science/article/abs/pii/S1053811905024523.

[110]　GROVA C, DAUNIZEAU J, KOBAYASHI E, et al. Concordance between distributed EEG source localization and simultaneous EEG-fMRI studies of epileptic spikes [J/OL]. NeuroImage, 2007, 39. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3792086/.

[111]　PODDER P, HASAN M M, ISLAM M R, et al. Design and Implementation of Butterworth, Chebyshev-I and Elliptic Filter for Speech Signal Analysis[J]. 2014. DOI: 10.5120/17195-7390.

[112]　YAEL D, VECHT J J, BAR-GAD I. Filter-Based Phase Shifts Distort Neuronal Timing Information[J]. eNeuro, 2018, 5(2). DOI: 10.1523/ENEURO.0261-17.2018.

[113]　TSUCHIMOTO S, SHIBUSAWA S, IWAMA S, et al. Use of common average reference and large-Laplacian spatial-filters enhances EEG signal-to-noise ratios in intrinsic sensorimotor activity[J]. Journal of Neuroscience Methods, 2021, 353: 109089. DOI: 10.1016/j.jneumeth.2021.109089.

[114]　HAGHIGHI S, JASEMI M, HESSABI S, et al. PyCM: Multiclass confusion matrix library in Python[J]. Journal of Open Source Software, 2018, 3(25): 729. DOI: 10.21105/joss.00729.

Appendix A   Meta-analysis

First author	Task	Recording method	Hand-edness	Sample size (N)	Year of publi-cation	Paradigm	Cleaning	Normalization	Spatial filtering	Channel re-refer-encing
1. Gerardin [104]	MI & ME	MRI	Right	8	2000	Simple (simultaneous flexion/extension of the fingers) or complex (selective flexion/extension of the index and the little finger) continuous movement	Motion correction	Normalization to stereotaxic Talairach coordinates	Gaussian spatial filter to a final smoothness of 5 mm	No
2. Grafton [114]	MI	PET	Right	7	1996	Precision grasp of common objects ranged in size from small to large	No	Spatial normalization using mean PET image from each individual + Global flow normalization	Smoothing to a final isotropic resolution of 18 mm	No
3. Stippich [105]	MI & ME	fMRI	Right	14	2002	Right and left sided complex finger opposition involving digits 1 – 5	Correction of linear trends, high frequency fluctuations, and motion; Exclusion when strong artifacts	Linear cross correlations between each image set and the supplied reference function	Threshold cluster size of 36 voxels	No
4. Mokienko [124]	MI	fMRI* & nTMS & EEG	Right	11	2013	Slow grasping movement with the right or left hand	Fat suppression and correction for motion	fMRI translation to standard Montreal Neurological Institute (MNI) coordinates	Smoothing	No
5. Willems [122]	MI	fMRI	Right & Left	32	2009	Imaging performance of manual and nonmanual action verbs	Realignment through rigid body registration	Normalization to Montreal Neurological Institute (MNI) space	Spatial smoothing (8 mm FWHM kernel)	No

Figure A–1 Main features of included studies; analyzed data is from the imaging technique marked by an asterix * (only for multiple recordings)

First author	Task	Recording method	Hand-edness	Sample size (N)	Year of publi-cation	Paradigm	Cleaning	Normalization	Spatial filtering	Channel re-refer-encing
6. Kraeutner [106]	MI & ME	MEG* & EMG	Right	16	2014	Sequential button press with the non-dominant (left) hand	Exclusion when head movements	Normalization to Talairach-Tournoux coordinate space	Dual-state beamforming	No
7. Burianová [107]	MI & ME	fMRI & MEG	Right	14	2013	Specific finger movements	Realignment onto the mean image for head-motion correction	Spatial normalization into a standard stereotaxic space with voxel size of 2 mm3 using the Montreal Neurological Institute (MNI) EPI template	Spatial smoothing filter for each volume by convolving it with an isotropic Gaussian kernel	No
8. Begliomini [107]	ME	fMRI	Right & Left	34	2008	Grasping with the index and the thumb of spherical plastic objects (3 cm diameter)	Motion correction	Normalization using the MNI152 template	Single-subject spatial normalization using the high-resolution T1 template	No
9. Begliomini [115]	ME	fMRI	Left	16	2018	Grasping with the index an+d the thumb of a spherical object (3 cm diameter)	Correction for head motion (translations/ rotations)	Normalization using the MNI152 template	Smoothing using a 3D Gaussian kernel	No
10. Di Russo [121]	SSVEP	fMRI & EEG	Right	15	2007	Circular Gabor grating sinusoidally modulated in black and white	(1) fMRI = Removal of constant and linear terms + motion correction through a rigid body transformation with a least squares approach (2) EEG = Rejection of epochs with artifacts (eye movements, blinks or amplifier blocking + Bandpass filter from 5 to 25 Hz on time-averaged VEPs	(1) fMRI = Normalization according to the overall sensitivity of each subject (2) EEG = Normalization to the standardized finite element model of BESA 2000	No	No

First author	Task	Recording method	Hand-edness	Sample size (N)	Year of publi-cation	Paradigm	Cleaning	Normalization	Spatial filtering	Channel re-refer-encing
11. Srinivasan [117]	SSVEP	fMRI	NA	6	2007	High-contrast sinusoid-dally reversing checker-board with stimulus frequencies ranging from 3 to 14 Hz	No	Normalization to the MNI template	Smoothing using a Gaussian filter of size 7.65 mmFWHM	No
12. Srinivasan [118]	SSVEP	EEG	Right	11	2006	Patterns composed of 600 dots with stimulus frequencies ranging from 3 to 30 Hz	Removal of 10 EEG channels due to artifacts in some subjects	Signal-to-Noise Ratio units	Surface Laplacian calculated with the New Orleans 3D Spline Laplacian algorithm	Average reference of the 110 recorded EEG channels
13. Perlstein [116]	SSVEP	EEG	Right	18	2003	Gray-scale pictures representing pleasant, neutral and unpleasant images flickering at 10 Hz	Offline correction for electro-oculographic contamination	Linear regression-base de-meaning and de-trending of the averaged data	LORETA source localization with a 7mm spatial resolution	Average reference of the 128 recorded EEG channels
14. Martinez [119]	SSVEP	EEG	NA	5	2007	Navigation of a small car on the computer screen in real time using an array of four small checkerboard images flickering with different frequencies	Blind source separation (BSS) algorithm for artifact rejection	Energy normalization of each flickering frequency directly by using a dedicated tuned filter	No	No
15. Menon [108]	P300	fMRI* & EEG	Right	11	1997	Auditory oddball paradigm with eighty 1000 Hz (frequent) and twenty 2000 Hz (infrequent) tones	Movement correction using least square minimization	Normalization to stereotaxic Talairach coordinates	Spatial smoothing with a uniform 3D Gaussian filter	No
16. Ikegami [120]	P300	EEG & fMRI*	Right	12	2012	Visual-flicker stimuli in a 6 × 6 alphabet flicker matrix modified from the P300 speller" with two types conditions (green/blue luminance and chromatic and white/gray luminance)	Removal of low-frequency noise with a high-pass filter (cut-off period of 128 s) + Correction of short-range serial correlations with an autoregressive (order one) model	Normalization to a Montreal Neurological Institute (MNI) EPI template	Smoothing with an 8-mm full-width at half-maximum Gaussian kernel	No

First author	Task	Recording method	Hand-edness	Sample size (N)	Year of publication	Paradigm	Cleaning	Normalization	Spatial filtering	Channel re-referencing
17. Ramirez-Quintana [109]	P300	EEG	NA	8	2020	Donchin Speller configured in columns (letter speller)	Third-order Butterworth bandpass filter from 1 to 15 Hz	Normalization to obtain a zero mean and unitary variance in all the channels	No	No
18. Ardekani [110]	P300	fMRI	NA	7	2002	Classic visual oddball paradigm with a standard visual stimulus of characters 'OOOOO' and a target image with characters 'XXXXX'	Motion correction + Centering of the voxels' time-series to obtain a zero mean + Trend removal by PCA	Normalization to the Talairach–Tourneaux coordinates	Resampling to a voxel size of 3x3x3 mm3 and a matrix size of 54x64x50	No
19. Clark [112]	P300	fMRI	Right	6	2000	Classic visual oddball paradigm with a standard visual stimulus using the letter 'T', a distractor letter 'C' and a target letter 'X'	Movement correction	Normalization to the Talairach–Tourneaux coordinates	Threshold clusters of 10 or more contiguous voxels	No
20. Mulert [111]	P300	EEG	NA	50	2004	Auditory oddball paradigm with 80 % non-target stimuli (500 Hz) and 20 % target stimuli (1000 Hz)	Rejection of trials using an amplitude criterion (±70 µV) for every EEG and EOG channel	Two approaches: Averaging of the LORETA-images and grand averaging of the ERPs	LORETA source localization with a 7mm spatial resolution	Average reference of the 63 recorded EEG channels
21. Bledowski [113]	P300	fMRI* & EEG	Right	10	2004	Three-stimulus oddball task with simple blue shapes	3-D motion correction to estimate the three translation and three rotation parameters of rigid body transformation	Normalization to the standard Montreal Neurological Institute (MNI) template head surface	Spatial smoothing of EPI images with a Gaussian kernel (8 mm FWHM kernel)	No

Acknowledgements

I am deeply grateful to all those who have supported and contributed to the completion of this dissertation. Their guidance, encouragement, and assistance have been invaluable throughout this challenging yet rewarding journey.

First of all, I would like to express my gratitude to my supervisor, Prof. Xiangyang Zhu, for letting me discover the amazing field of Brain-Computer Interfaces. I am truly fortunate to have had the opportunity to work freely on this topic that I find so fascinating. I extend my heartfelt appreciation to Prof. Xinjun Sheng and Prof. Jianjun Meng for their valuable feedback and editorial advice. Their expertise and perspectives have enriched the quality of this thesis. I am also grateful to Shanghai Jiaotong University for providing the necessary resources and infrastructure to conduct this research. Access to state-of-the-art equipment has been crucial in conducting the main experiment of this thesis and has broadened my understanding of the subject matter.

I especially want to thank my uncle, Claude Guttmann, for his continuous help and meticulous guidance throughout my Ph.D. studies, for kindling my love of computing and statistics, for introducing me to R and reviewing every aspect of my code, for his constructive criticism on algorithms, ideas, and most of the words and syntax in all of my scientific papers. It is an understatement to say that this thesis would look completely different and might not even exist without his never-ending efforts and commitment.

I also want to thank M. Ebrahim M. Mashat for his methodological advice and support in designing and conducting the experiment, as well as for all the insightful discussions over coffee breaks. I also wholeheartedly thank Nicolas Moutaux for his technical assistance with the 3D printing of the headrest and his recommendations on the design of the EMG electrodes adapted to the eyelid. Last but not least, I warmly thank Thomas Dumur for lending me and helping me set up his server, on which most of the computations were done.

My deepest gratitude goes to all the participants who generously volunteered their time and participated in this study. Their willingness to be involved in this research has been instrumental in advancing our understanding of the field. Their cooperation and patience during data collection have been greatly appreciated.

Furthermore, I want to express my gratitude to my family for their unwavering support, encouragement, and belief in my abilities. Their love and understanding have been my constant source of strength and motivation throughout this challenging journey. I am especially thankful for their patience over my sempiternal question about which colors and shapes look better in this thesis figures.

Finally, I want to thank the late teacher Wanren Wang, who taught me Chinese with poems and songs when I was a high school student in Beijing. His pedagogy and altruism serve as the best example of how one person can make the world a better place.

In conclusion, the completion of this PhD thesis would not have been possible without the contributions and support of all those mentioned above. I am sincerely grateful for their involvement and look forward to their continued guidance and encouragement in my future endeavors.

Research Achievements

Publications

[1] **E. Guttmann-Flury**, X. Sheng, D. Zhang, and X. Zhu, "A new algorithm for blink correction adaptive to inter- and intra-subject variability," Comput. Biol. Med., vol. 114, p. 103442, Nov. 2019, doi: 10.1016/j.compbiomed.2019.103442.

[2] **E. Guttmann-Flury**, X. Sheng, and X. Zhu, "Channel selection from source localization: A review of four EEG-based brain–computer interfaces paradigms," Behav. Res. Methods, Jul. 2022, doi: https://doi.org/10.3758/s13428-022-01897-2.

[3] **E. Guttmann-Flury**, X. Sheng, D. Zhang, and X. Zhu, "A Priori Sample Size Determination for the Number of Subjects in an EEG Experiment," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul. 2019, pp. 5180–5183. doi: 10.1109/EMBC.2019.8857482.

[4] **E. Guttmann-Flury**, X. Sheng, D. Zhang, and X. Zhu, "Preliminary Results on a New Algorithm for Blink Correction Adaptive to Inter- and Intra-Subject Variability," presented at the 9th International IEEE/EMBS Conference on Neural Engineering (NER), San Francisco, CA, USA, Mar. 2019, pp. 585–588. doi: 10.1109/NER.2019.8717029.

[5] **E. Guttmann-Flury**, X. Sheng, D. Zhang, and X. Zhu, "Characterize Differences between Frustration and Contentment to Analyze Errors in Brain-Computer Interfaces," presented at the 7th International IEEE EMBS Neural Engineering Conference, Montpellier, France, Apr. 2015.