

Independent Component Analysis denoising impact on the Analysis of Visual P300 Event-related potentials with ASD subjects

Project Report for Signal Processing 2021 course

João Miguel Areias Saraiva

joaomiguelsaraiva@tecnico.ulisboa.pt

Student number 86449 of

Departments of Computer Science and of Bioengineering

Instituto Superior Técnico, Lisboa, Portugal

ABSTRACT

The P300 wave is an electroencephalogram (EEG) visual event-related potential (ERP) component that can be elicited in classic oddball paradigm experiments. Reduced P300 amplitude has been found in autism spectrum disorder (ASD) subjects, which may difficult the detection of these potentials. This work accesses the impact of pre-processing EEG signals of ASD subjects with independent component analysis (ICA). Results show that performing P300 detection only on the P300-contributing independent component (IC) yielded by ICA gives a better consensus regarding P300 latency and amplitude, which might be of great interest for deep neural network classification.

KEYWORDS

ICA, P300, ERP, ASD, artifact removal

1 INTRODUCTION

A Brain-Computer Interface (BCI) is a closed-loop system that acquires brain physiological signals, analyzes them, and translates them into operational commands for other physical or virtual systems that carry out desired actions, without requiring any physical action from the user [34, 40]. To close the loop, a BCI should present some feedback to the user. Recent works have been demonstrating BCIs to control a variety of systems, such as prosthesis, orthosis, wheelchairs, home appliances, robots and computer programs [10, 15, 40].

Scalp electroencephalography (EEG) is one of the most popular techniques to record the brain electrical activity in a BCI system [40], since it is non invasive, relatively cheap, portable, and offers excellent temporal resolution. Some known control signals that can be captured with EEG can be interpreted and transformed into commands by the BCIs to control desired systems, including event-related potentials (ERP) [18, 23, 40], sensorimotor rhythms (SMR) [31, 40, 42], slow cortical potentials (SCP) [13], or steady state visually evoked potentials (SSVEP) [3, 30, 40].

In this work we will focus on the visual ERP P300, which is characterized by a positive deflection in the EEG of central and parietal electrodes around 300 ms after a visual stimulus onset [32], caused by attention and memory processes regarding the stimulus. The P300 has been divided into two potentials: the P3a around 250–280 ms after stimulus, with maximum amplitude over fronto-central electrodes, related to the engagement of attention; and the P3b around 300–500 ms after stimulus, with maximum amplitude over parietal electrodes, related to information processing [32]. In the literature, the occurrence of P300

term when alone refers to the P3b. In 1987, Fabiani *et al.* first described the oddball paradigm as a way of eliciting the P300 potential, by means of presenting a rare deviant stimulus during a stream of frequent standard stimuli [2]. As we shall see next, the experiments conducted on the used dataset follow the same paradigm.

BCI systems are starting to prove their efficacy as assistive and rehabilitative technologies in patients with severe motor impairments [8, 26, 33, 37, 39]. The proposed BCIs are often combined with a virtual reality (VR) system, where the closed loop is intended to control and/or perceive something in the program simulation [39]. However, very few systems have been proposed to help in neurorehabilitation in patients with cognitive deficits, such as subjects with autism spectrum disorder (ASD). As a matter of fact, Friedrich *et al.* (2014) reviewed a series of neurofeedback training studies that demonstrate that quantitative EEG-based neurofeedback training could be used as a personalized therapeutic approach in ASD [19].

Autism spectrum disorder (ASD) is a set of pervasive and sustained neurodevelopmental conditions characterized by persistent deficits in social communication and social interaction, alongside restricted, repetitive patterns of behavior, interests, or activities [7], that afflicts about 1.5% of the population of developed countries [11], with recent increases primarily among those without comorbid intellectual disability [28].

The group of Amaral and Simões *et al.* performed a clinical trial [4] in 2018 to access the feasibility of their visual P300-BCI system for improvement of social joint attention (JA) in ASD patients. Patients with ASD present severe deficits in JA [17, 24], which is a social communication skill, developed early in life, defined by the non-verbal coordination of attention of two individuals towards a third object or event [9]. Patients were able to correctly identify the visual cues in the simulation environment and the corresponding P300 ERPs could be detected by inspection of the EEG signals [4].

The EEG dataset acquired by this group was object of the IFMBE Scientific Challenge on the following year [35], where groups would have compete with the best machine learning classifier that could accurately detect the P300 events. All groups opted by supervised learning approaches and all kinds of pre- and post-processing were allowed to yield the best accuracies possible. None of the groups tried independent component analysis (ICA), although one of them tried using principal component analysis (PCA) to reduce dimensionality before training their classifiers. In this work, with the same dataset, we propose to study the impact that ICA would have on the detection of the visual P300 ERP events.

2 PROBLEM AND MOTIVATION

Accurate detection of P300 events is not just a matter designing a good classifier, but a great deal of effort should also be dedicated to process the EEG signals in a way that helps the classifier to achieve the best accuracy it can achieve. For instance, one should try to remove as much noise as possible from the signal in order to maximize the signal-to-noise ratio (SNR) before trying to detect whatever ERP one wants. Intuitively we can oversee this processing will have major impacts on the classifier accuracies.

Normal EEG is not perfectly sinusoidal, nor regular, nor has a discernible pattern, making its interpretation a difficult challenge. Firstly, EEG is very sensible to other electrical activity that comes from the body, besides cortical neurons, like eye movement and blinking, heart rate, muscular activity, power-line noise and the instability of electrodes during the recording. So, all these are considered artifacts and must be rejected. Secondly, even with instrumentation amplifiers, usually recorded EEG voltages show a very low amplitude (microvolt order), being difficult to capture with a good SNR. Also, the measured voltages must cross anatomical structures, including the skull, the meninges, the skin, that act as filters and further reduce their amplitude.

Some state-of-the-art techniques have proved to be very effective in removing artifacts from noisy EEG signals, such as empirical mode decomposition (EMD) and its variations and independent component analysis (ICA). Arising on this motivation, this work accesses the impact of ICA denoising on the provided dataset of ASD subjects. We will see the implications of the number of components subtracted, the criteria to choose components to remove, and the ICA algorithm that is applied.

3 BACKGROUND

This Section describes the mathematical foundations of ICA and some algorithms to implement it. Let us denote \mathbf{x} as the vector whose elements are the recorded potentials x_i where $i \in \{1, \dots, n\}$ and n is the number of recording channels. Also, let \mathbf{s} be the vector whose elements are the true signals coming from the sources s_j where $j \in \{1, \dots, n\}$. Each recorded potential x_i is said to be a mixture of all s_j signals, each with a contribution of a_{ij} , just like a linear combination:

$$x_i = \sum_j^n a_{ij} \cdot s_j \quad (1)$$

Or in a matricial form:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} \quad (2)$$

where \mathbf{A} is the matrix of elements a_{ij} . Notice that the \mathbf{s} components cannot be observed or measured, and it is the goal of an ICA model to found the components \mathbf{s} only knowing the measured potentials \mathbf{x} . To achieve that, the parameters \mathbf{A} need to be estimated. That process relies on two key assumptions:

- (1) Components of \mathbf{s} must be statistically independent.
- (2) Components of \mathbf{s} must not follow a Gaussian distribution.

From the first assumption, we can derive important independence properties like the *factorization* of probability density functions (pdf) in Equation 3 and the consequent independence of estimations of functions, h , over the independent variables in Equation 4.

$$p(s_i, \dots, s_j) = p(s_i) \cdot \dots \cdot p(s_j) \quad (3)$$

$$E\{h(s_i) \cdot \dots \cdot h(s_j)\} = E\{h(s_i)\} \cdot \dots \cdot E\{h(s_j)\} \quad (4)$$

From the second assumption, we found the key aspect this process works on, which is to maximize non-Gaussianity between independent components. Besides, we would not be able to estimate \mathbf{A} if the components were Gaussian-distributed, but we will ignore that for now. Let us only understand why we want to maximize non-Gaussianity.

By considering the linear combination in Equation 1, we can conclude from the Central Limit Theorem that the complete sum of all components s_j will tend to be *more* Gaussian than the distribution of each one of them separated. Thus, the idea will be to estimate each component alone and find an \mathbf{A} that *maximizes* the non-Gaussianity of them all. More specifically, let us have a variable transformation $\mathbf{z} = \mathbf{A}^T \cdot \mathbf{w}$, where \mathbf{w} is a vector to be estimated. Matrix \mathbf{W} is composed of all \mathbf{w} rows, and $\mathbf{W} = \mathbf{A}^{-1}$. Then for any sample y we would have:

$$y = \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbf{A} \cdot \mathbf{s} = \mathbf{z}^T \cdot \mathbf{s} \quad (5)$$

Correspondingly, we could now find a \mathbf{w} that *maximizes* the non-Gaussianity of $\mathbf{w}^T \mathbf{x}$. In the transformed coordinate system, this is a \mathbf{z} with only one non-zero component, say z_A ; because $\mathbf{z}^T \cdot \mathbf{s}$ is *less* Gaussian when it equals one s_i component, say s_A , scenario where all other $s_i \neq s_A$ would be zero due to the multiplication by \mathbf{z} . In each z_A being non-zero, that would mean that $\mathbf{w}^T \mathbf{x}$, and consequently $\mathbf{z}^T \cdot \mathbf{s}$ would be an independent component.

This process needs to be repeated for every independent component, alternating between the non-zero element of \mathbf{z} . This corresponds to an orthogonalization process for finding all the local maxima of non-Gaussianity of $\mathbf{w}^T \mathbf{x}$, which are, in fact, $2n$ local maxima, two for each independent component (s_i and $-s_i$). But to determine such maxima, non-Gaussianity needs to be quantified. Several algorithms for computing ICA models use different measures to quantify non-Gaussianity.

3.1 Infomax algorithm

A popular algorithm to build ICA models was first proposed in 1995 called Infomax [6, 12]. Its name suggests it tries to *maximize individual information of components*, or in other words, to *minimize the amount of mutual information*. We can define mutual information for a pair of random variables, $I(x_1, x_2)$, as:

$$I(x_1, x_2) = H(x_1) - H(x_1|x_2) \quad (6)$$

where $H(x_1)$ is the entropy of x_1 , and $H(x_1|x_2)$ is the conditional entropy of x_1 given that x_2 takes a certain value, defined as:

$$H(x_1|x_2) = H(x_1, x_2) - H(x_2) \quad (7)$$

where $H(x_1, x_2)$ is the joint entropy of x_1 and x_2 . The Shannon entropy (1948) is formally defined as:

$$H(x_i) = - \int p(x_i) \cdot \log p(x_i) dx_i \quad (8)$$

$$H(x_i, x_j) = - \int p(x_i, x_j) \cdot \log p(x_i, x_j) d(x_i, x_j) \quad (9)$$

where $p(\cdot)$ are the respective pdfs. Combining Equations 6 and 7, we can state that the joint entropy of the two random variables is:

$$H(x_1, x_2) = H(x_1) + H(x_2) - I(x_1, x_2) \quad (10)$$

In information theory, entropy is the opposite of information (notice the negative signs in Equations 8 and 9). Roughly speaking, the lower the entropy we have between x_1 and x_2 , the more information x_2 can give us about x_1 , and vice-versa. In other words, to maximize $H(x_1, x_2)$ in Equation 10 we need to maximize the individual entropies of x_1 and x_2 , $H(x_1)$ and $H(x_2)$, and minimize the mutual information of both, $I(x_1, x_2)$. And maximizing $H(x_1, x_2)$ means maximizing the independence of variables x_1 and x_2 . So, generalizing to n recording channels, we can formulate the problem of finding maximum independence between all channels, \mathbf{x} , as a problem of minimizing the mutual information between all of them, $I(\mathbf{x})$. In the limit, when $I(\mathbf{x}) = 0$, all components are statistically independent [12].

Using this notion of minimizing mutual information, Amari *et al.* proposed the following algorithm [6] to learn the unmixing matrix \mathbf{W} through a stochastic gradient descent approach, with learning rate η :

- (1) Initialize $\mathbf{W}(0)$ with some arbitrary weights (e.g. random).
- (2) Repeat for each timestep t :
 - (a) Set $\mathbf{y} = \mathbf{W}(t) \cdot \mathbf{x}$
 - (b) Update $\mathbf{W}(t+1) = \mathbf{W}(t) + \eta(t) \cdot [\mathbf{I} - f(\mathbf{y}) \cdot \mathbf{y}^T] \cdot \mathbf{W}(t)$
 - (c) If \mathbf{W} converges, then halt.

where \mathbf{I} is the identity matrix, and $f(\mathbf{Y})$ is a nonlinear activation function chosen by the algorithm according to the type of distribution of \mathbf{Y} . If \mathbf{Y} is super-Gaussian, typically $f(\mathbf{Y}) = \tanh(\mathbf{Y})$; if \mathbf{Y} is sub-Gaussian, typically $f(\mathbf{Y}) = \mathbf{Y} - \tanh(\mathbf{Y})$. Convergence is reached when $|\mathbf{W}(t+1) - \mathbf{W}(t)| < \epsilon$, where ϵ is some constant chosen by the user.

Having found the unmixing matrix \mathbf{W} we can obtain an approximation, \mathbf{y} , of the original sources, \mathbf{s} , by performing operation (11). These will be called independent components (ICs) onwards.

$$\mathbf{y} = \mathbf{W} \cdot \mathbf{x} \quad (11)$$

The activation function serves an indirect approximation measure of the *amount* of mutual information. In fact the mutual information is minimized by minimizing the dependency between output components and, in turn, this dependency is measured by the Kullback-Leibler divergence between the joint and the product of the marginal distributions of the outputs:

$$D(\mathbf{W}) = \int p(\mathbf{y}) \cdot \log \frac{p(\mathbf{y})}{\prod_{a=1}^n p_a(y^a)} d\mathbf{y} \quad (12)$$

where $p_a(y^a)$ is the marginal pdf. By using Equations 8 and 9, one can see the Kullback-Leibler divergence, $D(\mathbf{W})$, relates to the average mutual information of y through Equation 13 [6].

$$D(\mathbf{W}) = -H(\mathbf{y}) + \sum_{a=1}^n H(y^a) \quad (13)$$

Since the exact pdfs are unknown, Amari *et al.* use the truncated Gram-Charlier expansion to approximate them [6], which leads to an

approximation of $D(\mathbf{W})$ in function of the observed recorded channels \mathbf{x} , using the relation (11).

$$D(\mathbf{W}) \approx -H(\mathbf{x}) - \log|det(\mathbf{W})| + \frac{n}{2} \log(2\pi e) - F(\mathbf{y}) \quad (14)$$

where $f(\mathbf{y}) = F'(\mathbf{y})$. By deriving $D(\mathbf{W})$ in reference to each $w \in \mathbf{W}$, one can reach the update rule in the algorithm step (2b). Refer to [6] for the complete proof.

3.2 Picard algorithm

Infomax maximizes the likelihood of independence using the classic stochastic gradient algorithm, which may require some hand-tuning and often fails to converge [29], or only converges slowly. Picard algorithm [1] was proposed in 2017, following the same mathematical framework as Infomax, but implementing it in a Newton optimization manner. More precisely, the limited-memory BFGS (L-BFGS) optimization algorithm [27], which is part of the family of quasi-Newton learning methods.

Picard stands for Preconditioned ICA for Real Data. The idea behind Picard is to use the Hessian approximations as preconditioners for the L-BFGS method. The algorithm refines the Hessian approximations to better take into account the true curvature [1]. The time cost per iteration of Picard is similar to the simple quasi-Newton methods, while promising far better descent directions, and that is the main reason why it was chosen to build our ICA models. See how the algorithm exploits the Hessian approximations to initialize the recursive formula:

- (1) Initialize $\mathbf{W}(0)$ with some arbitrary weights (e.g. random).
- (2) Repeat for each timestep t :
 - (a) Set $\mathbf{y} = \mathbf{W}(t) \cdot \mathbf{x}$
 - (b) Compute the relative gradient $G(t)$.
 - (c) Compute Hessian approximation $H(t)$.
 - (d) Compute the search direction $p(t) = -H^{-1}(t)$ like in L-BFGS.
 - (e) Update $\mathbf{W}(t+1) = (I + \eta(t)p(t)) \cdot \mathbf{W}(t)$
 - (f) If \mathbf{W} converges, then halt.

Refer to [1] for the complete proof.

3.3 Pre-processing considerations

It should be highlighted that before doing an ICA model, the observed data \mathbf{x} should meet some requirements in order for the algorithms to be more efficient and the model better conditioned. The algorithms used already take care of this pre-processing. The mathematical framework presented here assumes that \mathbf{x} is zero-meaned and of unit variance. So, firstly, the algorithms zero-mean \mathbf{x} and this implies the output \mathbf{s} to be centered around zero as well. So, after estimating the mixing matrix \mathbf{A} , the algorithm adds back the mean vector of \mathbf{s} to \mathbf{s} .

Secondly, the ICA algorithms also whiten the observations \mathbf{x} into a white vector \mathbf{x}' of unit variance. This can be accomplished by eigenvalue decomposition (EVD) of the covariance matrix of \mathbf{x} , ending up with an \mathbf{x}' with a covariance matrix that equals the identity matrix. In turn, this will imply having a corresponding new mixing matrix, \mathbf{A}' , that is orthogonal, hence reducing the number of parameters to be estimated, therefore reducing the algorithms time complexity. The parameters to be estimated by the mixing matrix \mathbf{A} are n^2 , whereas only $n(n-1)/2$ need to be estimated for \mathbf{A}' .

4 RELATED WORK

In 2016, Cui *et al.* reviewed 32 P300 studies with a total of 407 ASD patients and 457 non-ASD subjects and found incongruities and disparities caused by several factors. Reduced amplitude of P3b was found in ASD patients, compared with non-ASD patients, but no difference of P3b latency, P3a amplitude, or P3a latency was found between groups [16]. Moreover, they concluded the oddball paradigm elicited attenuated P3b amplitude in Pz electrode among ASD subjects. The work of Townsend *et al.* corroborates that the variability of average P300 amplitude at Pz is larger for control subjects than for ASD subjects [38].

Jung *et al.* results show the response latencies of P300 peaks were response-locked, i.e., covaried with response time [21]. They infer that ICA separated stimulus- and response-locked ERPs into different components on the basis of temporal independence and differences in spatial distribution, and allowed the time courses of response-locked components to be realigned, preventing temporal smearing of late response activity in the average P300 [21]. The group showed a larger, earlier and narrower P300 peak in corrected average signals of ASD patients, after ICA artifact-removal and latency-realignment [21], which makes them better detectable for a potential classifier.

Xu *et al.* used ICA to decompose EEG signals acquired on a speller paradigm experiment. Their trick was to spatio-temporally manipulate the ICA components to increase the P300 detection accuracy [41]. Spatially, they defend that P300 is more prominent of the Cz, C1 and C2 region, so they withdraw components expected to come from outside this region. Temporally, those components with relatively larger amplitude in the latency range of P300 ([250, 400] ms) were kept, while the others were set to zero. Afterwards, the manipulated A was back-projected to the scalp. The P300 curves before and after manipulation showed a notorious difference. As we shall explain next, a part of this approach was followed in Group B analysis.

Tahirovic *et al.* also proposed a spatial averaging technique between ICA components to better extract P300 signals [36], because sometimes, as we intuitively expect, ICA is not able to fully separate P300 into only one component (which would be the ideal).

Kahan *et al.* achieved better P300 waveforms detection when processing oddball paradigm EEG signals with constrained-ICA (cICA) than with standard ICA, using a simple linear classifier. Not only better accuracies were achieved, but also the amount of input for classification (only Cz, Pz and Fz used) and computation time were reduced [22]. The group stands by saying: " (...) predefined channels cannot be used to automatically extract P300 from a study with multiple subjects, since efficient channels for P300 extraction vary between the subjects" [22].

5 APPROACH AND UNIQUENESS

5.1 Material

5.1.1 Dataset. The EEG dataset used was collected by Amaral *et al.* team using a g.Nautilus system (g.tec medical engineering GmbH, Austria) [5], with 8 active electrodes positioned at C3, Cz, C4, CPz, P3, Pz, P4, POz locations – see Figure 1. The reference electrode was placed at the right ear and the ground electrode at the AFz location. The signals of each electrode are differentially amplified, meaning signals of all channels are the subtraction of two electrodes: one in the place we

want to measure the signal, and other the reference electrode [25]. Sampling rate was set to 250 Hz and data were acquired notch-filtered at 50 Hz and passband-filtered between 2 and 30 Hz [35]. The dataset was made publicly available [here](#) (doi: 10.34740/kaggle/dsv/1375326) in the context of 2019 IFMBE Scientific Challenge, organized by MEDICON. A copy of the dataset is also present at directory data, along with other created files for this work.

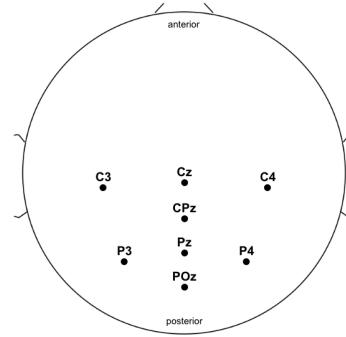


Figure 1: Channels location in the g.Nautilus system.

The dataset includes EEG signals from 15 patients (mean age = 22 years and 2 months, ranging from 16 to 38 years old) with high-functioning ASD (Full-Scale Intelligent Quotient [FSIQ]: mean = 102.53; SD = 11.64) [4]. The dataset is divided by patients, each with code SBJxx, and each patient subset is divided by seven sessions, each with code Sxx. Seven BCI sessions spread over four months were conducted, the first four sessions weekly and the remaining monthly [4]. Inside each session, there is also a division between Train and Test samples, but we will join them together, since they belong to the same session day.

The team conducted experiments on a virtual simulation environment where the patients had to demonstrate their abilities of joint attention (JA), regarding an avatar character paying attention to certain objects in a virtual bedroom space. For each patient, the experiments were divided into B blocks, where in each block the system would select one of eight objects to be the target object to pay attention (a book, a radio, a laptop, a printer, a ball, a corkboard, a wooden plane, and a picture). In each block, K experiment runs were repeated. In each run, eight epochs, E , occur, each being a flash event of each of the eight objects. In each run the flashes order is randomized. To form the dataset, each epoch was then segmented into chunks of 1400 ms, from -200 ms prior to flash onset to 1200 ms after. Notice that the dataset includes all epochs – the ones in which the target object was flashed and the ones the object flashed was not the target one.

5.1.2 ICA and EEG Analysis Software. A console version of EEGLAB (v. 2021) was used on MATLAB 2019a to do the ICA models, component removal and noise source analysis. Documentation can be found at <https://eeglab.org>. A working version is at directory eeglab2021.0, as well as a script implemented to automate analysis (`analysis_protocol.m`), as explained in Section 5.2.1. There is also another script to only remove components of a dataset already with computed ICA weights (`remove_components_only.m`). See documentation for full detail of these scripts.

5.1.3 P300 Detection Software. In directory p300 there can be found a Python script which detects the P300 latency and amplitude of a given subject (p300_plot.py), as explained in Section 5.2.2. It also plots the average ERP for each session of the subject, like the Figures in Supplementary Material (SM) 1. Another script called plot_grand_average.py provides plots like the ones in Figures 2 and 3. See documentation for full detail of these scripts.

5.2 Methods

5.2.1 ICA Analysis. We defined a protocol to analyse each subject-session dataset, composed by the following (summarized) steps:

- (1) Concatenate the training and test subsets in a unique dataset.
Train and test partitioning is only relevant for classification purposes. Not only there is no problem of combining these two since they were acquired one after the other in the same session, with no different methodology, but also the ICs removed/kept should be the same in order to remove/keep the same patterns in the train and test subsets.
- (2) Trim all epochs in time from $[-0.2, 1.2]$ s to $[0, 1]$ s.
This was considered the relevant time interval for P300 detection [14].
- (3) Inspect and remove abnormal epochs by eye.
- (4) Visually analyse average ERP image of not-P300 epochs.
- (5) Visually analyse average ERP image of P300-only epochs, and take into account the amplitude and latency of the P300 peak in all channels.
- (6) Run an ICA algorithm on the dataset to found 8 ICs.
Two different algorithms were considered: Infomax and Picard.
- (7) Run automatic artifact classification (with ICLLabel) on the discovered ICs.
- (8) Analyse IC series on time interval $[200, 550]$ s, and respective scalp maps, while taking into the account the ICLLabel classification achieved before and empirical experience on the matter.
- (9) Choose which components to remove/keep according to the strategy being followed, run the back-projection, and save the pruned dataset.
The strategy may vary. Control is given to the user. See Section 5.3.
- (10) Visually re-analyse average ERP image of P300-only epochs, and track the change in amplitude and latency of the P300 peak in all channels.

This analysis can be conducted automatically by running the script eeglab2021.0/analysis_protocol.m. To run it, open MATLAB, add the path eeglab2021.0, and type "analysis_protocol X Y" on the command line, where X is a subject number of the respective code, and Y is a session number of the respective code. Do not include the code prefixes SBJ, S or the left 0s; just the integer numbers. The script runs the analysis on the given pair subject-session and saves relevant images at directory results and new datasets (one with concatenated train-test epochs, one with ICA weights, one with the pruned ICs) at directory data. The script will pause at step (9) and ask in the command line which components to remove – ICs to remove should be given as an array with the IC numbers, e.g. [2 6 7]. Note that the steps above

are a summarized version from the user perspective – see the script documentation for a full description.

5.2.2 P300 Detection. The P300 peak was defined as the point with maximum amplitude between 200ms and 550ms. When this point is found, its amplitude (in μ V) and latency (time in ms after visual onset at $t = 0$) are found and saved for all electrodes. When more than one peak is visible we intend to capture P3b peak (which comes after P3a and it is more present in the electrodes we have). In each pair subject-session, a unique P300 amplitude was computed to be the average amplitude found in channels Cz, Pz and POz.

5.2.3 Statistical Analysis. For a given subject, its P300 amplitude and latency was considered to be the median of these values among all seven sessions. By sorting the n sampled values on ascending order the median value can be computed as in Equation 15.

$$\begin{cases} \left[\frac{n+1}{2} \right]^{th} \text{sample} & , \text{ if } n \text{ is odd} \\ \frac{\left[\frac{n}{2} \right]^{th} \text{sample} + \left[\frac{n+1}{2} \right]^{th} \text{sample}}{2} & , \text{ if } n \text{ is even} \end{cases} \quad (15)$$

The standard deviation (with Bessel's correction) of these values was computed by Equation 16, where X is the set of sampled values and \bar{X} its mean.

$$S = \sqrt{\frac{\sum_{x \in X} |x - \bar{X}|^2}{n - 1}} \quad (16)$$

In group-analysis, median values and standard deviations were averaged among all subjects by (17), where S is the number of subjects.

$$\bar{Y} = \frac{\sum_{y \in Y}(y)}{S} \quad (17)$$

The correction percentage between original amplitude and latencies, $y_{original}$, and those after ICs removal, y_{pruned} , was computed as the experimental error in Equation 18.

$$\text{Correction}(\%) = \frac{|y_{pruned} - y_{original}|}{y_{pruned}} \times 100 \quad (18)$$

5.3 Proposed Approach

The proposed solution to achieve relevant inputs for P300 detection is to use ICA to fulfill two main strategies:

- (1) **Remove common artifacts:** In some timeseries, ICA is able to isolate independent components that are most likely to be the isolated activity of temporal muscle, heart, line noise or other artifacts. When that is the case, the solution is to remove those components.
- (2) **Keep P300-contributing components:** In some timeseries, ICA is able to isolate one, two or three independent components that, together or alone, are responsible for P300 events. When that is the case, and if those components are predicted to be predominant throughout all scalp, one solution might be to keep those components only, while rejecting all others.

While goal (1) is a straight forward approach, very well documented in the literature, goal (2) might raise some questions. To test this hypothesis, and to answer those questions, the full dataset was divided in half by subjects, as enumerated below:

- **Group A:** Subjects 6, 7, 8, 9, 11, 13, 14 and 15.
- **Group B:** Subjects 1, 2, 3, 4, 5, 10 and 12.

While in Group A the components removed were chosen based on their artifact contribution, fulfilling goal (1), in Group B the components kept were chosen based on their P300 contribution, fulfilling goal (2). Note that, by definition, in the second approach the artifact-contributing components end up being removed as well. So, what is going to be seen is that the set of components removed in the first approach would be a subset of the components removed by the second approach. Hence, we can treat Group A as a control group of whether approach (2) is a viable solution for achieving better P300 detection.

6 RESULTS AND CONTRIBUTIONS

6.1 Group-analysis

Figures 2 and 3 depict the grand-average (average across all patients) of ERPs on channel Pz, by each session (session 1 not shown), for Group A and Group B, respectively. The original averaged signals are presented by the curves gray and yellow, where the gray curves show only the averaged (expected) not-P300 epochs (the ones where the object flashing was not the target one), and the yellow curves show only the averaged (expected) P300 epochs (the ones where the object flashing was the target one). We can see how pronounced it is the P3b peak in the yellow curves, between 400-480 ms and around 1.86-2.45 μ V, while the gray curves become almost flat-zero when averaged. This is the result of not-P300 epochs having most channels being quasi-sinusoidal, hence the average being near zero. In the yellow curves we can also see the next negative ERP following P300 - the N400 – further confirming the presence of visual stimulation potentials. These findings go in hand with the ones presented by the original group [4].

The green curves represent only the averaged (expected) P300 epochs only after ICA and ICs removal. For Group A, we can observe that the P300 waveform does not change much relatively to the original waveform, besides being attenuated (Figure 2). For Group B, we can observe that the P300 waveform suffers more modifications on its landscape (Figure 3). This can be explained by two major reasons:

- In most average-sessions, we can see the P3a peak was almost or completely attenuated. This was one of the goals when choosing which ICs to prune on this subjects. This can help automatic models to not misclassify P3a peak as the desired P300, in cases where P3a waves present themselves higher than P3b. See the case of average-session 4 in Figure 3 where this almost happens.
- In some average-sessions, the average P300 waveform is not visually well-defined (e.g. 2 and 5). This is because different subjects have different P300 latencies and that is not very well portraited in the yellow curves, because in the original signals the P300 duration is very wide and not a sharp peak – yielding a wider average. Thus, when selecting only the P300-contributing component(s) to remain, and eliminating the others, different latencies become clearer for different subjects, highlighting that

indeed P300 occurs at different latencies for different subjects. Consequently, the average green curve might not present a single P300 peak. See SM1 for individual P300 waveforms.

Tables 1 and 2 summarize the two most important metrics when detecting P300 peaks – amplitude and latency – for subjects of Group A and subjects of Group B, respectively. Median amplitude and latency values across sessions of expected P300 in the original signals (before ICs removal) and in the signals resultant of ICA and ICs removal are shown. The error is presented as the standard deviation of the sampled values. The *Correction* field indicates how much these values increased or decreased after ICs removal. The grand-average of all subjects is also presented. A list of the components removed for each pair subject-session may be consulted in SM2 and the IC numbers can be identified in the Figures in directory results.

The first thing one should notice is the percentage of corrections of both amplitude and latency was eager in Group B than in Group A. Although not necessary, this was expected since more ICs were removed in Group B than in Group A. So, in principle, it would cause more changes in the signal. This corroborates what we see in Figure 2.

The second thing to notice is that the average standard deviation of the P300 latency decreases significantly after ICs removal in Group B, while in Group A it increases. Having a low variation in latency across sessions is a desired characteristic if we want to perform P300 automatic detection with all sessions of a subject combined. Hence, this is a first indication that the strategy followed in Group B might be a better one than that of Group A. The similar is observed for amplitude, although less pronounced.

6.2 Case Studies of Group A

Strategy 1 (remove artifact-related components) is a very cumbersome task that should always have a human inspection case-by-case. It is also a task that requires expertise from a practical point of view. Fortunately, some auxiliary algorithms have been developed to help novices like me, such as ICLLabel. Next are presented two cases that occur typically in the process of inspecting ICs.

In Figure 4 top panel, one should notice the complete incoherence between epochs (or trials) and the bad average ERP profile that is created from that. Also, on a true brain-related component we expect to see in the power spectrum some peak at the 10 Hz corresponding to the alpha rhythm. This is not the case here – not only the power decreases at 10 Hz, but a significant peak is present from 15 Hz to 30 Hz. This might be suggestive of small myoelectric artifacts, since there is chance of these being found between 20-300 Hz. Adding to that, if we look at the scalp distribution of this component we notice that it was mostly projected onto the left temporal lobe, although no electrode was placed there (see at results/SBJ11/S04/picard). This agrees with the literature when looking for temporal myoelectric artifacts [20]. Also, we see a *bump* at 5 Hz which suggests not-brain-related activity. Furthermore, ICLLabel classified this IC as only 51% brain-related, hence this IC was removed so as any other that presented the same characteristics. This IC accounted for 14.5% of scalp variability, so for sure it altered the ERP landscape. Subject 11 had other sessions with muscle and other -related ICs, and their removal contributed to a reduction in P300 latency across sessions in approximately 24 ms – see SM3 for individual reduction of

Independent Component Analysis denoising impact on the Analysis of Visual P300 Event-related potentials with ASD subjects

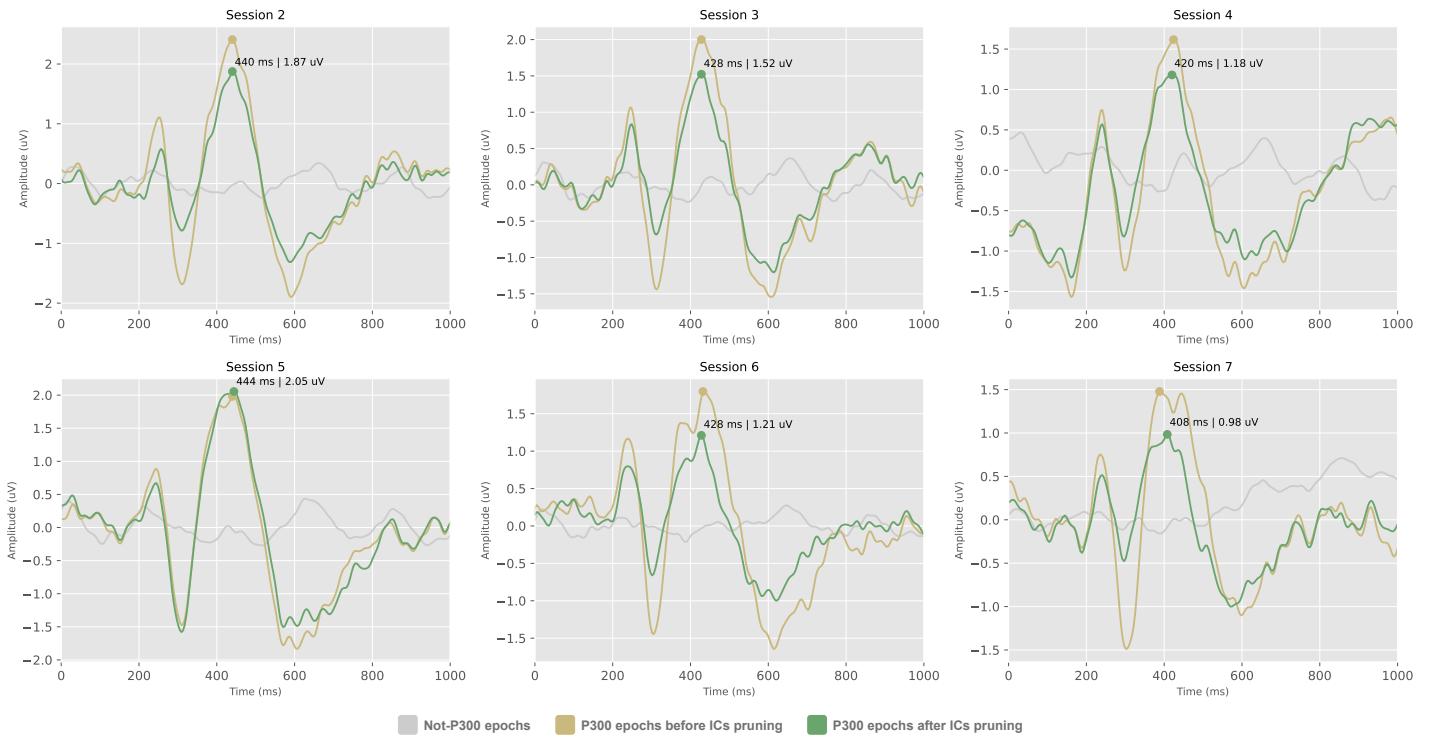


Figure 2: Group A grand-average of event-related potentials on channel Pz in each BCI session.

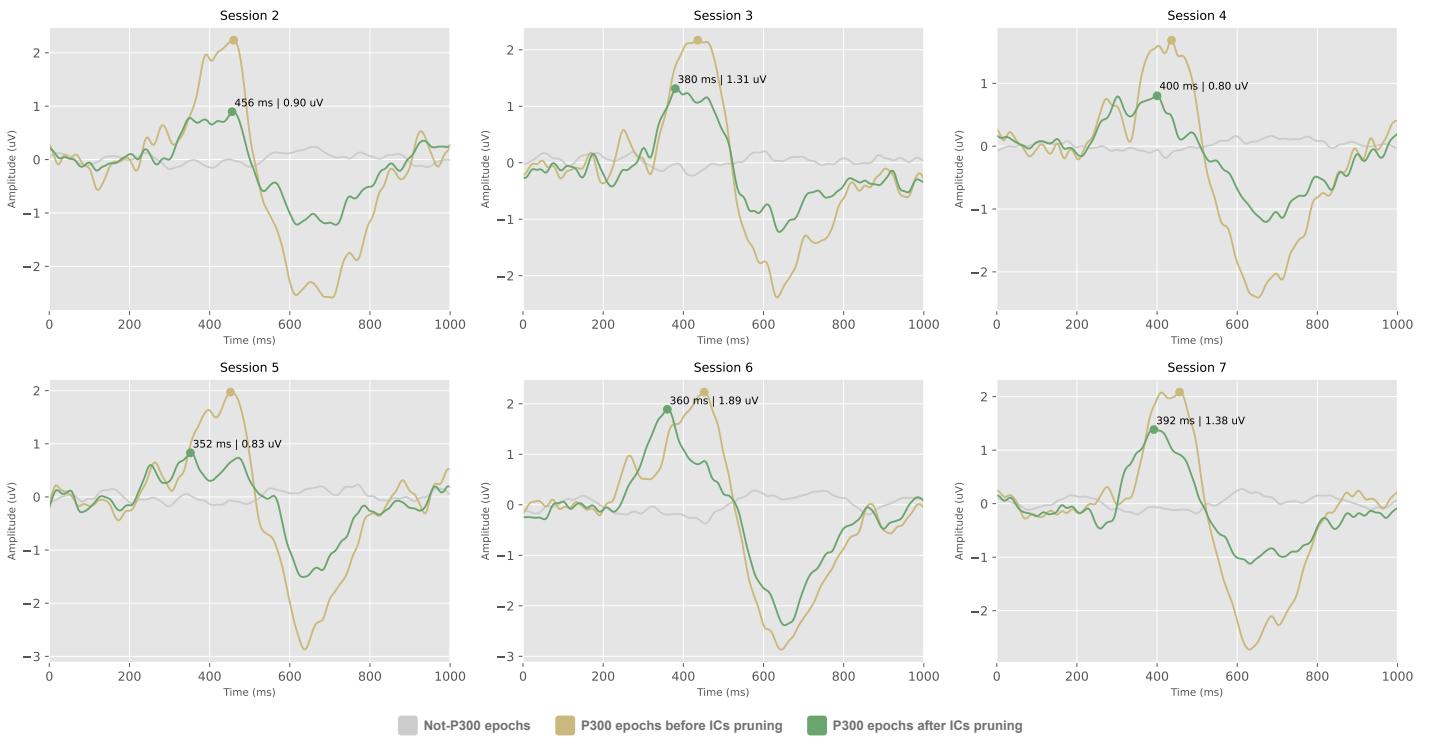


Figure 3: Group B grand-average of event-related potentials on channel Pz in each BCI session.

Table 1: P300 inter-session median amplitude and latency for each patient of Group A, before and after ICs removal.

		Subject	6	7	8	9	11	13	14	15	Grand-average
P300 amplitude (μ V)	Before	4,53 $\pm 0,88$	2,76 $\pm 1,25$	3,00 $\pm 0,66$	2,80 $\pm 0,53$	2,51 $\pm 0,75$	2,10 $\pm 1,15$	2,62 $\pm 0,98$	2,88 $\pm 0,57$	2,90 $\pm 0,63$	
	After	3,38 $\pm 0,75$	1,64 $\pm 0,68$	3,03 $\pm 0,40$	2,78 $\pm 0,67$	2,62 $\pm 0,42$	3,82 $\pm 0,57$	3,00 $\pm 1,04$	2,31 $\pm 0,51$	2,82 $\pm 0,85$	
	Correction	10,05%	25,05%	7,16%	4,47%	17,05%	24,16%	5,41%	24,52%	14,63%	
P300 latency (ms)	Before	440,0 $\pm 15,5$	430,7 $\pm 29,9$	454,7 $\pm 10,5$	400,0 $\pm 9,6$	353,3 $\pm 53,1$	373,3 $\pm 33,3$	402,7 $\pm 25,0$	238,7 $\pm 27,0$	386,7 $\pm 25,5$	
	After	433,3 $\pm 33,6$	464,0 $\pm 55,0$	453,3 $\pm 14,6$	408,0 $\pm 13,4$	293,3 $\pm 28,5$	348,0 $\pm 42,1$	401,3 $\pm 27,6$	238,7 $\pm 42,0$	380,0 $\pm 32,1$	
	Correction	7,78%	3,45%	0,31%	0,33%	22,26%	8,24%	0,33%	0,58%	5,41%	

Table 2: P300 inter-session median amplitude and latency for each patient of Group B, before and after ICs removal.

		Subject	1	2	3	4	5	10	12	Grand-average
P300 amplitude (μ V)	Before	2,87 $\pm 0,47$	3,81 $\pm 0,64$	3,02 $\pm 4,37$	3,88 $\pm 0,60$	2,40 $\pm 0,77$	3,52 $\pm 0,74$	2,19 $\pm 0,32$	3,10 $\pm 1,13$	
	After	2,28 $\pm 0,56$	2,71 $\pm 0,55$	2,30 $\pm 1,14$	2,64 $\pm 0,78$	2,74 $\pm 0,86$	2,74 $\pm 0,81$	1,62 $\pm 0,82$	2,43 $\pm 0,79$	
	Correction	25,71%	41,04%	37,48%	34,93%	45,43%	20,75%	35,86%	34,46%	
P300 latency (ms)	Before	340,0 $\pm 34,5$	445,3 $\pm 53,3$	362,7 $\pm 39,4$	374,7 $\pm 19,1$	394,7 $\pm 23,2$	414,7 $\pm 18,4$	352,0 $\pm 36,7$	383,4 $\pm 32,1$	
	After	384,0 $\pm 11,7$	444,0 $\pm 19,1$	312,7 $\pm 22,8$	354,7 $\pm 20,4$	344,0 $\pm 25,0$	432,0 $\pm 11,4$	317,3 $\pm 36,7$	369,8 $\pm 21,9$	
	Correction	9,16%	0,62%	14,71%	3,76%	16,87%	0,91%	13,54%	8,51%	

each session. The P300 result in Cz and Pz channels after removal is significantly better (see at `results/SBJ11/S04`).

In the bottom panel of Figure 4 we have an example of an heart-related artifact. ICLLabel classified this component as 24% heart-related and, indeed, we can notice small peaks in the average ERP that reassemble those of an ECG, although not with the same periodicity. We notice again an absent peak at 10 Hz, but in frequencies superior to it. Also, the epoched activity does not seem very relevant and seems to have some outlier trials (or epochs) around trial 3000. Therefore, this IC should be removed, so as any other that presented the same characteristics. However, this IC accounted for 0.6% of scalp variability, so removing it might not produce major changes on the total ERP. Indeed, for subject 15 session 2, another two more prominent components were removed and those were the main responsibles for changes in the signal – see SM2. The P300 result in Cz and Pz channels after removal is slightly worst (see at `results/SBJ15/S02`).

As a side note, notice that without frontal channels in the dataset, it is useless to try to find eye-related activity, since these artifacts are usually prominent at the frontal lobe.

6.3 Case Studies of Group B

Strategy 2 (keep only P300-contributing component(s)) is a task that can be done more automatically, although it was made manually. We simply need to look at each IC waveform for the P300 epochs and choose the one (perhaps two) that most contribute to the P300 potential. Also, this component should be projected throughout all scalp and not just a niche. Next are illustrated two cases of Group B.

In Figure 5 top left panel, we notice two major peaks around 400 ms for P300 epochs only. The highest one can vary from session to session – in this session the highest was the latest for all channels but one. Looking at the waveform of each IC (top middle panel) is where we can make the decision about which is the component responsible for the P300 peak in all channels. At first, one might think it is IC6 (solid green), but that is actually responsible for the first peak of that one channel. It is actually IC1 (dashed blue) that is responsible for the P300 in every channel, so this component was kept, and all the others were removed. Notice how the result (top middle right) yields a clean waveform for all channels, while maintaining the P300 latency around the same value (427-431 ms).

In Figure 5 bottom left panel, we observe an overall more messy waveform of P300 epochs only. It might be difficult to do automatic detection of the P300 peak. Here, EEGLAB computed it to be at 376 ms latency, as an average between all the maximum peaks. Four of the channels present a P300 wider duration at 500 ms. By looking at the decomposition of each IC (bottom middle panel) we can see that IC1 and IC2 are the responsible for the P300 peak in all channels. Again, according to our strategy, it is not IC3 because it was not manifested in all channels. Removing all components except IC1 and IC2 yields the result in the bottom right panel, where we notice the P300 peak excessive duration of those four channels was pruned. The latency got adjusted to 324 ms.

Notice how P300 latencies can be different for different subjects. Also, depending on the detection method, it can also be preferred to have different P300 waveforms among subjects, as long as they are similar

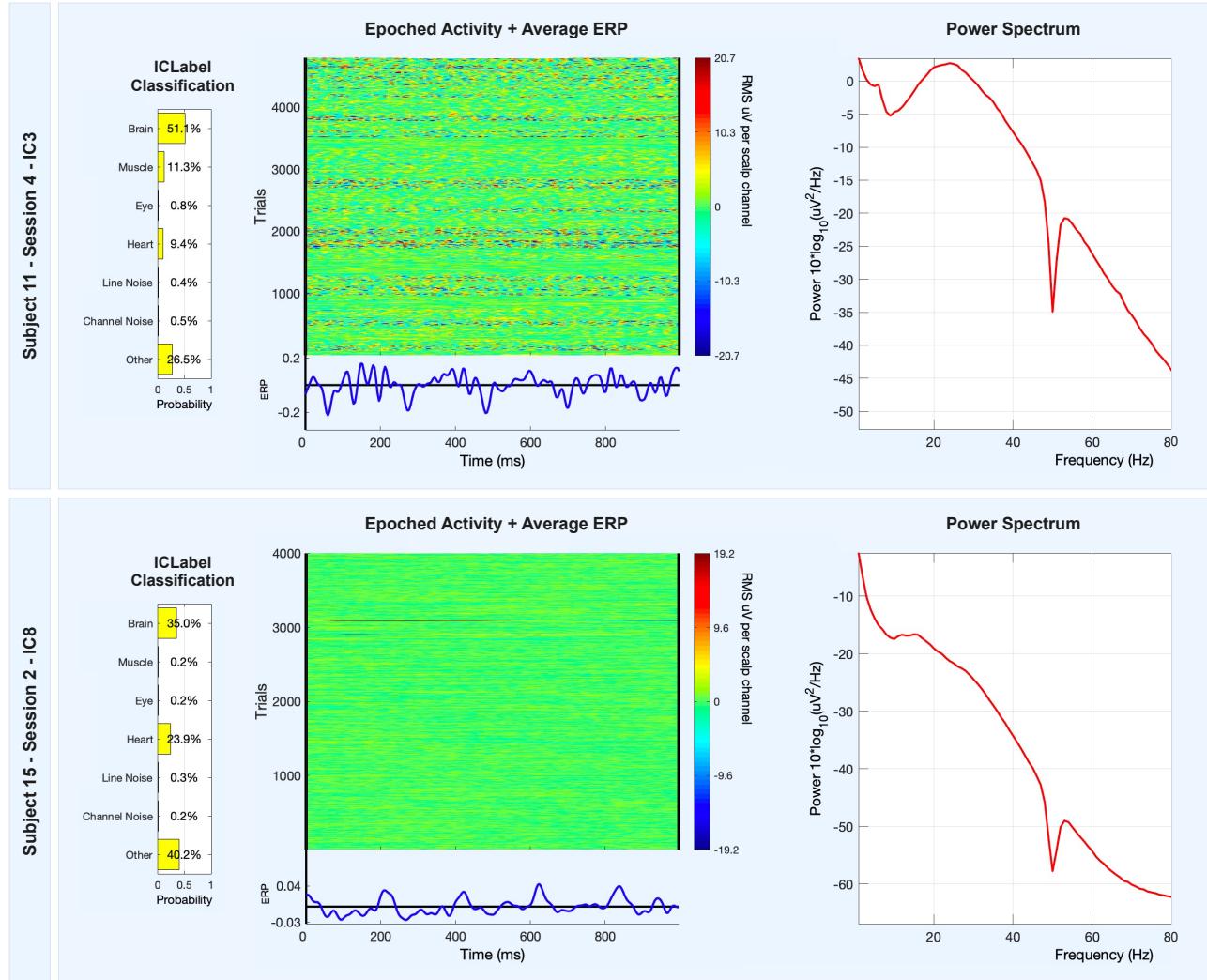


Figure 4: Two Group A case studies of two pairs subject-session for removing artifact-related ICs.

between each subject's sessions. These were two very clear examples – in some cases the P300-responsible component(s) is not very clear and might difficult the selection task. However, a good heuristic is to look at the first and second ICs, if they are ordered by scalp variability, since the most prominent components usually contain the P300 peak in P300 epochs. But this is not necessarily the case! So, even though this could be done automatically, a human double-checking should be done.

6.4 Computational remarks

There were run two different ICA algorithms: Infomax algorithm, provided in EEGLAB as `runica`, and Picard algorithm, an EEGLAB plug-in. The results presented before are all based on Picard ICA models, and there were no relevant differences in the discovered ICs between both. Although, there are significant differences in the time complexity between the two. Empirically, Picard took an average 0.923 seconds per iteration (SPI), with an average of 36 iterations per subject-session pair; while Infomax took an average of 1,238 SPI, with an average of 273

iterations¹. Finding the unmixing matrix \mathbf{W} is a challenging problem if it has to be completed quickly and accurately on large sets of real data. So, Picard proves to be a faster and cheaper algorithm with the same IC decomposition quality.

7 DISCUSSION AND CONCLUSIONS

The strategy followed for Group A yielded a larger variability in P300 latency across sessions for each patient (except for subject 11). If the goal is to detect a unanimous P300 latency for each patient this is a poor strategy. If the goal is to remove artifacts from the signals, this is a good strategy. But at what cost is one goal preferred to the other? Should we be concerned about cleaning the *whole* signal with only 8 independent components, at the cost of loosing important signal features?

We should be aware that with only 8 channels available, the signals can only be decomposed up to 8 independent components. Usually, ICA

¹Experiments conducted solely on a dual-core CPU of 3.1 GHz, and 16 GB of DDR3 memory.

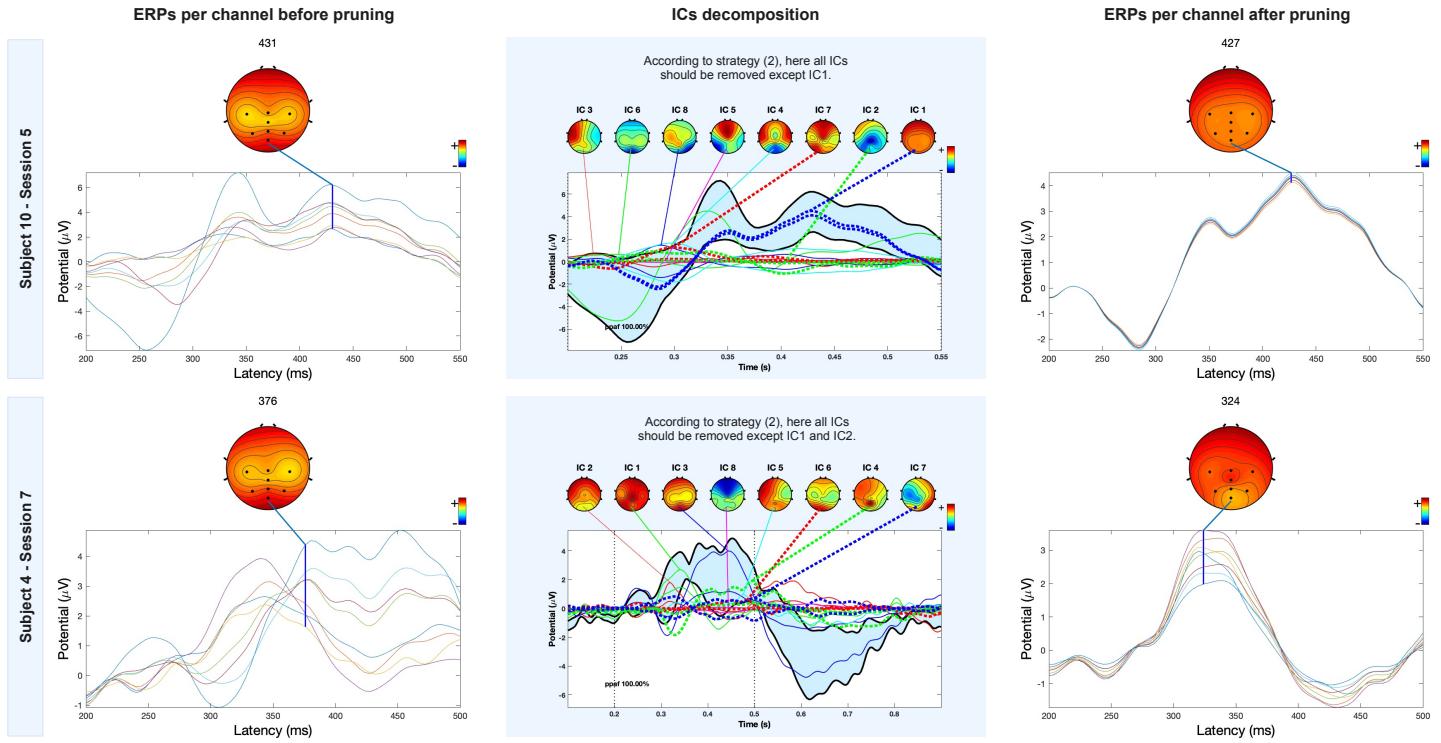


Figure 5: Two Group B case studies of two pairs subject-session for keeping P300-contributing ICs.

studies work with at least 32 channels which allow for a division in a higher number of components, perhaps favoring the chance of better isolation of individual artifactual activity. From these experiments, we can conclude that the consequence of having a division onto only 8 ICs is that each IC that could potentially contain artifact-activity will also contain significant brain-activity, such as the examples in Figure 4. And by pruning those ICs, we may as well be pruning relevant brain-activity. So, the message here is that separating the observations in eight components might be a poor-resolution separation. By having more channel observations, intuitively one can oversee that a more fine-grain separation would be obtained in terms of SNR.

Furthermore, by riskily removing ICs with both artifact-activity and brain-activity, these ICs can be different across sessions of one subject. That is, the removed ICs' y backprojection on the scalp might be different if the ICs removed are different from session to session. This explains why the P300 latency variability increased after ICs pruning in the subjects of Group A.

Regarding the strategy followed for Group B, we can notice the opposite: the variability of P300 latency decreased after ICs pruning. In general, the final ERP waveforms for each subject are much more clear across sessions and across channels. This might be of particular interest for deep learning classifiers, such as convolutional neural networks (CNNs). Indeed, if we train a CNN with all sessions of a subject, it will better learn the P300 waveform the more similar they are throughout all P300 epochs². Indeed, there is a direct correlation between the less

variability of examples of the same classification class (e.g. P300 and not-P300) and the greater accuracy of classification.

Added to that, strategy of Group B has the advantage of producing signals of lower dimensionality. Whatever it is the model we choose for P300 detection, dimensionality can be reduced after IC pruning. Usually, one or two ICs remained and all channels have a very similar waveform, meaning there would be no advantage of feeding all channels to the detection algorithm. So, from 8 channels, we can end up giving the algorithm only one, such as the Pz, where the P300 amplitude is larger.

In some sense, one might interpret the first strategy as a subtracting approach (we remove a couple of ICs), while the second as an additive approach (we make up a signal with only one IC). The strategy to follow should depend on the application – if we want to classify epochs as P300 and not-P300, or if we want to detect the P300 latency and amplitude, or both. Future efforts should be dedicated in attesting the classification accuracy in machine learning classifiers such as CNNs and support vector machines (SVMs), with the signals produced by the second strategy.

DATA AND CODE AVAILABILITY

All data and software used is publicly available at <https://github.com/jomy-kk/AutICA>. Throughout this document, all directory names and files are relatively referenced to the root of this repository.

SUPPLEMENTARY MATERIAL

SM1 – Average ERP images per subject-session pair. Access [here](#).

SM2 – List of ICs removed per subject-session pair. Access [here](#).

SM3 – Amplitudes and latencies per subject-session pair. Access [here](#).

²Although, this does not necessarily mean the model would generalize better. Empirical results would have to be drawn to affirm that.

REFERENCES

- [1] Pierre Ablin, Jean-Francois Cardoso, and Alexandre Gramfort. 2018. Faster Independent Component Analysis by Preconditioning With Hessian Approximations. *IEEE Transactions on Signal Processing* 66, 15 (Aug. 2018), 4040–4049. <https://doi.org/10.1109/TSP.2018.2844203>
- [2] Patrick K Ackles, J. Richard Jennings, and Michael G. H Coles. 1988. *Advances in psychophysiology: a research annual*. Vol.2 1987 Vol.2 1987. JAI Press Inc., Greenwich, Conn. OCLC: 1200082146.
- [3] Sangtae Ahn, Kiwoong Kim, and Sung Chan Jun. 2016. Steady-State Somatosensory Evoked Potential for Brain-Computer Interface—Present and Future. *Frontiers in Human Neuroscience* 9 (Jan. 2016). <https://doi.org/10.3389/fnhum.2015.00716>
- [4] Carlos Amaral, Susana Mouga, Marco Simões, Helena C. Pereira, Inês Bernardino, Hugo Quental, Rebecca Playle, Rachel McNamara, Guiomar Oliveira, and Miguel Castelo-Branco. 2018. A Feasibility Clinical Trial to Improve Social Attention in Autistic Spectrum Disorder (ASD) Using a Brain Computer Interface. *Frontiers in Neuroscience* 12 (July 2018), 477. <https://doi.org/10.3389/fnins.2018.00477>
- [5] Carlos P. Amaral, Marco A. Simões, Susana Mouga, João Andrade, and Miguel Castelo-Branco. 2017. A novel Brain Computer Interface for classification of social joint attention in autism and comparison of 3 experimental setups: A feasibility study. *Journal of Neuroscience Methods* 290 (Oct. 2017), 105–115. <https://doi.org/10.1016/j.jneumeth.2017.07.029>
- [6] S. Amari, A. Cichocki, and H. H. Yang. 1995. A New Learning Algorithm for Blind Signal Separation. In *Proceedings of the 8th International Conference on Neural Information Processing Systems* (Denver, Colorado) (NIPS'95). MIT Press, Cambridge, MA, USA, 757–763.
- [7] American Psychiatric Association (Ed.). 2013. *Diagnostic and statistical manual of mental disorders: DSM-5* (5. ed. ed.). American Psychiatric Assoc, Washington, DC.
- [8] Elaine Astrand, Claire Wardak, and Sulianni Ben Hamed. 2014. Selective visual attention to drive cognitive brainâ€“machine interfaces: from concepts to neurofeedback and rehabilitation applications. *Frontiers in Systems Neuroscience* 8 (Aug. 2014). <https://doi.org/10.3389/fnsys.2014.00144>
- [9] R. Bakeman and L. B. Adamson. 1984. Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development* 55, 4 (Aug. 1984), 1278–1289.
- [10] Mahdi Bamdad, Homayoon Zarshenas, and Mohammad A. Auais. 2015. Application of BCI systems in neurorehabilitation: a scoping review. *Disability and Rehabilitation: Assistive Technology* 10, 5 (Sept. 2015), 355–364. <https://doi.org/10.3109/17483107.2014.961569>
- [11] A. J. Baxter, T. S. Brugha, H. E. Erskine, R. W. Scheurer, T. Vos, and J. G. Scott. 2015. The epidemiology and global burden of autism spectrum disorders. *Psychological Medicine* 45, 3 (Feb. 2015), 601–613. <https://doi.org/10.1017/S003329171400172X>
- [12] Anthony J. Bell and Terrence J. Sejnowski. 1995. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation* 7, 6 (Nov. 1995), 1129–1159. <https://doi.org/10.1162/neco.1995.7.6.1129>
- [13] N. Birbaumer, N. Ghanyaim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor. 1999. A spelling device for the paralysed. *Nature* 398, 6725 (March 1999), 297–298. <https://doi.org/10.1038/18581>
- [14] Davide Borra, Silvia Fantozzi, and Elisa Magosso. 2020. Convolutional Neural Network for a P300 Brain-Computer Interface to Improve Social Attention in Autistic Spectrum Disorder. In *XV Mediterranean Conference on Medical and Biological Engineering and Computing – MEDICON 2019*. Jorge Henriques, Nuno Neves, and Paulo de Carvalho (Eds.). Vol. 76. Springer International Publishing, Cham, 1837–1843. https://doi.org/10.1007/978-3-030-31635-8_223 Series Title: IFMBE Proceedings.
- [15] Ujjwal Chaudhary, Niels Birbaumer, and Ander Ramos-Murgialday. 2016. Brain-computer interfaces for communication and rehabilitation. *Nature Reviews Neurology* 12, 9 (Sept. 2016), 513–525. <https://doi.org/10.1038/nrneurol.2016.113>
- [16] Tingkai Cui, Peizhong Peter Wang, Shengxin Liu, and Xin Zhang. 2017. P300 amplitude and latency in autism spectrum disorder: a meta-analysis. *European Child & Adolescent Psychiatry* 26, 2 (Feb. 2017), 177–190. <https://doi.org/10.1007/s00787-016-0880-z>
- [17] Geraldine Dawson, Karen Toth, Robert Abbott, Julie Osterling, Jeff Munson, Annette Estes, and Jane Liaw. 2004. Early Social Attention Impairments in Autism: Social Orienting, Joint Attention, and Attention to Distress. *Developmental Psychology* 40, 2 (2004), 271–283. <https://doi.org/10.1037/0012-1649.40.2.271>
- [18] L.A. Farwell and E. Donchin. 1988. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology* 70, 6 (Dec. 1988), 510–523. [https://doi.org/10.1016/0013-4694\(88\)90149-6](https://doi.org/10.1016/0013-4694(88)90149-6)
- [19] Elisabeth V. C. Friedrich, Neil Suttie, Aparajithan Sivanathan, Theodore Lim, Sandy Louchart, and Jaime A. Pineda. 2014. Brain-computer interface game applications for combined neurofeedback and biofeedback treatment for children on the autism spectrum. *Frontiers in Neuroengineering* 7 (July 2014). <https://doi.org/10.3389/fneng.2014.00021>
- [20] I.I. Goncharova, D.J. McFarland, T.M. Vaughan, and J.R. Wolpaw. 2003. EMG contamination of EEG: spectral and topographical characteristics. *Clinical Neurophysiology* 114, 9 (Sept. 2003), 1580–1593. [https://doi.org/10.1016/S1388-2457\(03\)00093-2](https://doi.org/10.1016/S1388-2457(03)00093-2)
- [21] Tzyy-Ping Jung, Scott Makeig, Marissa Westerfield, Jeanne Townsend, Eric Courchesne, and Terrence J. Sejnowski. 2001. Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping* 14, 3 (Nov. 2001), 166–185. <https://doi.org/10.1002/hbm.1050>
- [22] Ozair Idris Khan, Faisal Farooq, Faraz Akram, Mun-Taek Choi, Seung Moo Han, and Tae-Seong Kim. 2012. Robust extraction of P300 using constrained ICA for BCI applications. *Medical & Biological Engineering & Computing* 50, 3 (March 2012), 231–241. <https://doi.org/10.1007/s11517-012-0861-4>
- [23] Sonja C. Kleih, Tobias Kaufmann, Claudia Zickler, Sebastian Halder, Francesco Leotta, Fabio Cincotti, Fabio Aloise, Angela Riccio, Cornelia Herbert, Donatella Mattia, and Andrea Kübler. 2011. Out of the frying pan into the fire—the P300-based BCI faces real-world challenges. In *Progress in Brain Research*. Vol. 194. Elsevier, 27–46. <https://doi.org/10.1016/B978-0-444-53815-4.00019-4>
- [24] Ami Klin, Warren Jones, Robert Schultz, Fred Volkmar, and Donald Cohen. 2002. Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals With Autism. *Archives of General Psychiatry* 59, 9 (Sept. 2002), 809. <https://doi.org/10.10101/archpsyc.59.9.809>
- [25] Myer Kutz. 2010. *Biomedical Engineering and Design Handbook, Volumes I and II (2nd Edition)*. McGraw-Hill Professional Publishing, New York, USA. <http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=4657812> OCLC: 958553882.
- [26] Michael J. Larson, Mikle South, Erin Krauskopf, Ann Clawson, and Michael J. Crowley. 2011. Feedback and reward processing in high-functioning autism. *Psychiatry Research* 187, 1-2 (May 2011), 198–203. <https://doi.org/10.1016/j.psychres.2010.11.006>
- [27] Dong C. Liu and Jorge Nocedal. 1989. On the limited memory BFGS method for large scale optimization. *Mathematical Programming* 45, 1-3 (Aug. 1989), 503–528. <https://doi.org/10.1007/BF01589116>
- [28] Kristen Lyall, Lisa Croen, Julie Daniels, M. Daniele Fallin, Christine Ladd-Acosta, Brian K. Lee, Bo Y. Park, Nathaniel W. Snyder, Diana Schendel, Heather Volk, Gayle C. Windham, and Craig Newschaffer. 2017. The Changing Epidemiology of Autism Spectrum Disorders. *Annual Review of Public Health* 38, 1 (March 2017), 81–102. <https://doi.org/10.1146/annurev-pubhealth-031816-044318>
- [29] Jair Montoya-Martínez, Jean-François Cardoso, and Alexandre Gramfort. 2017. Caveats with Stochastic Gradient and Maximum Likelihood Based ICA for EEG. In *Latent Variable Analysis and Signal Separation*. Petr Tichavský, Massoud Babaie-Zadeh, Olivier JJ. Michel, and Nadège Thirion-Moreau (Eds.). Vol. 10169. Springer International Publishing, Cham, 279–289. https://doi.org/10.1007/978-3-319-53547-0_27 Series Title: Lecture Notes in Computer Science.
- [30] Gernot R Müller-Putz, Reinhold Scherer, Christian Brauneis, and Gert Pfurtscheller. 2005. Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components. *Journal of Neural Engineering* 2, 4 (Dec. 2005), 123–130. <https://doi.org/10.1088/1741-2560/2/4/008>
- [31] Gert Pfurtscheller and Christa Neuper. 1997. Motor imagery activates primary sensorimotor area in humans. *Neuroscience Letters* 239, 2-3 (Dec. 1997), 65–68. [https://doi.org/10.1016/S0304-3940\(97\)00889-6](https://doi.org/10.1016/S0304-3940(97)00889-6)
- [32] John Polich. 2007. Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology* 118, 10 (Oct. 2007), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>
- [33] David B. Salisbury, Marie Dahdah, Simon Driver, Thomas D. Parsons, and Kathleen M. Richter. 2016. Virtual reality and brain computer interface in neurorehabilitation. *Proceedings (Baylor University. Medical Center)* 29, 2 (April 2016), 124–127. <https://doi.org/10.1080/08998280.2016.1129386>
- [34] Jerry J. Shih, Dean J. Krusinski, and Jonathan R. Wolpaw. 2012. Brain-computer interfaces in medicine. *Mayo Clinic Proceedings* 87, 3 (March 2012), 268–279. <https://doi.org/10.1016/j.mayocp.2011.12.008>
- [35] Marco Simões, Davide Borra, Eduardo Santamaría-Vázquez, GBT-UPM, Mayra Bittencourt-Villalpando, Dominik Krzemiński, Aleksandar Miladinović, Neural_Engineering_Group, Thomas Schmid, Haifeng Zhao, Carlos Amaral, Bruno Di-reito, Jorge Henriques, Paulo Carvalho, and Miguel Castelo-Branco. 2020. BCIAUT-P300: A Multi-Session and Multi-Subject Benchmark Dataset on Autism for P300-Based Brain-Computer-Interfaces. *Frontiers in Neuroscience* 14 (Sept. 2020), 568104. <https://doi.org/10.3389/fnins.2020.568104>
- [36] A. Tahirovic, M. Matteucci, and L. Mainardi. 2015. An Averaging Technique for the P300 Spatial Distribution. *Methods of Information in Medicine* 54, 03 (2015), 215–220. <https://doi.org/10.3414/ME13-02-0037>
- [37] Ariel Tankus, Itzhak Fried, and Shy Shoham. 2014. Cognitive-motor brain-machine interfaces. *Journal of Physiology-Paris* 108, 1 (Feb. 2014), 38–44. <https://doi.org/10.1016/j.jphysparis.2013.05.005>
- [38] Jeanne Townsend, Marissa Westerfield, Echo Leaver, Scott Makeig, Tzyy-Ping Jung, Karen Pierce, and Eric Courchesne. 2001. Event-related brain response abnormalities in autism: evidence for impaired cerebello-frontal spatial attention networks. *Cognitive Brain Research* 11, 1 (March 2001), 127–145. [https://doi.org/10.1016/S0926-6410\(00\)00072-0](https://doi.org/10.1016/S0926-6410(00)00072-0)
- [39] Athanasios Vourvopoulos, Carolina Jorge, Rodolfo Abreu, Patrícia Figueiredo, Jean-Claude Fernandes, and Sergi Bermúdez i Badia. 2019. Efficacy and Brain Imaging Correlates of an Immersive Motor Imagery BCI-Driven VR System for Upper Limb Motor Rehabilitation: A Clinical Case Report. *Frontiers in Human Neuroscience* 13 (July 2019), 244. <https://doi.org/10.3389/fnhum.2019.00244>
- [40] Jonathan R. Wolpaw and Elizabeth Winter Wolpaw (Eds.). 2012. *Brain-computer interfaces: principles and practice*. Oxford University Press, Oxford ; New York.

- [41] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F. Yang. 2004. BCI Competition 2003—Data Set IIb: Enhancing P300 Wave Detection Using ICA-Based Subspace Projections for BCI Applications. *IEEE Transactions on Biomedical Engineering* 51, 6 (June 2004), 1067–1072. <https://doi.org/10.1109/TBME.2004.826699>
- [42] Han Yuan and Bin He. 2014. Brain-Computer Interfaces Using Sensorimotor Rhythms: Current State and Future Perspectives. *IEEE Transactions on Biomedical Engineering* 61, 5 (May 2014), 1425–1435. <https://doi.org/10.1109/TBME.2014.2312397>