PAPER

Identifying evoked potential response patterns using independent component analysis and unsupervised learning

Biomedical Physics & Engineering Express

PAPER

Identifying evoked potential response patterns using independent component analysis and unsupervised learning

RECEIVED
17 July 2018

REVISED
21 October 2018

ACCEPTED FOR PUBLICATION
7 November 2018

PUBLISHED
30 November 2018

Ana R Teixeira[1,2], **Isabel M Santos**[3] **and A M Tomé**[1]

[1] DETI-IEETA Universidade Aveiro 3810-193 Aveiro, Portugal
[2] ESEC Instituto Politécnico Coimbra 3030-329 Coimbra, Portugal
[3] CINTESIS, Departamento de Educação e Psicologia, Universidade Aveiro 3810-193 Aveiro, Portugal

E-mail: ateixeira@ua.pt

Keywords: ICA, HCT, K-means, ERP, FRN

Supplementary material for this article is available online

Abstract

Independent Component Analysis (ICA) is a pre-processing step widely used in brain studies. One of the most common problems in artifact elimination or brain activity related studies is the ordering and identification of the independent components (ICs). In this work, a novel procedure is proposed which combines ICA decomposition at trial level with an unsupervised learning algorithm (K-means) at participant level in order to enhance the related signal patterns which might represent interesting brain waves. The feasibility of this methodology is evaluated with EEG data acquired with participants performing on the Halstead Category Test. The analysis shows that it is possible to find the Feedback Error Negativity (FRN) Potential at single-trial level and relate its characteristics with the performance of the participant based on their knowledge of the abstract principle underlying the task.

1. Introduction

During the last decade, electroencephalographic (EEG) signals have been used to develop applications such as personal authentication [1], control of drowsy driving [2], emotion recognition [3] and so on. These applications are still developed and tested in research laboratory environments, however, the EEG recording is achieved under less constrained conditions [4] than those which are traditionally applied in brain cognitive studies. Therefore, only the combination of more powerful signal processing and machine learning algorithms are able to extract useful information about brain processing. One of the signal processing strategies is to decompose the multichannel recordings into a set of independent components. ICA is an exploratory matrix factorization method which decomposes the data matrix into a linear combination of independent components. When applied to multichannel recordings like EEG, ICA decomposes the raw data into maximally statistically independent components and a set of weights which explain the spread of the components over the scalp (scalp maps). The method has been introduced to EEG analysis by Makeig *et al* [5, 6] and it has been disseminated by the EEGLAB, a widely used software package running under MATLAB [7]. Since then, the development of graphical tools to assist the EEG reviewer in the selection of components has been a wide-spread trend among the EEGLAB users [8]. Moreover, most of the early studies were focused on evaluating different ICA or blind source separation (BSS) algorithms using either artificially mixed data or real data [9, 10]. The success of ICA in EEG analysis is largely due to the plausibility of the solution returned by the method [11]. Usually, the scalp maps are regular and smooth and the independent components present a unique waveform pattern or shape. Recently in [12] it is pointed out that ICA is by far the most widely used method for artifact elimination and correction. One of the key issues of any signal decomposition via ICA is the selection or identification of components related with artifacts or with brain signals. In [13] several methods to detect these artifacts are evaluated, and the authors conclude that the methods were not consistent in the selection of components. There, a new tool is offered, whose goal is to guide the experimenter's decisions. An alternative approach discussed in the literature is inspired by the observation of experts during visual classification of components. For this process, the 2D

scalp maps are observed (sometimes in conjunction with power spectrum of corresponding components) and a decision about artifact or non-artifact is made. This procedure can be modeled as a supervised machine learning problem and in [14, 15] a complete set-up is proposed to identify artifact related components. With that proposal, the scalp maps (columns of the mixing matrix) are interpolated and texture features are computed from the resulting images. Furthermore, to form the training data, the scalp maps are also labeled by experts.

In most of the applications of brain computer interfaces, the label coincides with the task performed, thus avoiding the manual data labeling [16, 17]. This feature is shared with emotion recognition systems, where the type of stimulus to elicit the emotions serves as label to the supervised learning algorithm [3, 18]. When ICA is introduced as a pre-processing step, the goal is to find out which of the components is of interest [19]. However, BCI systems use algebraic methods to compute the column of the separation matrix which enhances the component related with one of two possible tasks [4]. This procedure is called common spatial filter calculation [20, 21]. The aim of the operation is to transform the multichannel recording into one component from which several measures result which form the inputs to a classifier (supervised machine learning algorithm) [22]. A variant of the common spatial filter incorporates the average amplitudes of a chosen wave (for instance the *P*300) on the calculation of the filter coefficients [23]. The prior knowledge about stimulus type (target versus non-target) and and its latency are also essential information to calculate filter coefficients [24, 25].

Several attempts also have been made to incorporate available signal characteristics and to develop the so-called semi-blind source separation techniques. These strategies incorporate constraints in the contrast function, related with source independence, to be optimized [26]. The constraints might be applied to the columns of the mixing matrix [27, 28] or to the extracted components [29–32]. These approaches imply the existence of either a reference signal or a model to be used as constraint and the choice is dependent on the problem under study. Unsupervised machine learning techniques have also been introduced in brain studies with the advantage of working without labels [33, 34].

In this work, K-means clustering is applied to deal with the permutation indeterminacy of the ICA decomposition. The methodology is evaluated with EEG signals acquired while participants performed on the Halstead Category Test [35], which is used to assess cognitive executive functions. The complete test involves a short number of trials and it is meant to measure the paticipant's ability to formulate abstract principles based on the feedback received about his/her performance. Therefore, the goal is to track the ICA component related with the participant's performance at single trial level. The analysis shows that it is possible to find the Feedback Related Negativity (FRN) potential at the single-trial level and relate its characteristics with the performance. This approach was also applied to identify one of most studied ERP which is related to visual processing mechanisms (the *P*100). In this study, the trial averages were estimated in synchronization with the stimulus onset and with the feedback onset related with participant's response. The former is important to study the event-related potential *P*100 which peaks around 100 ms after the stimulus onset. The feedback onset is relevant to study the FRN component, which peaks around 250 ms after providing the information about the performance. These waves were chosen as case studies to demonstrate the adequacy of the proposed analysis technique because they are relevant markers of perceptive and cognitive performance in the present paradigm. The *P*100 is one of the most studied neural correlates of visual processing, being elicited in posterior areas in response to visual stimuli [36]. This component has been localized to activity in lateral extrastriate cortex and inferior occipital-temporal cortex [37]. The FRN component is a neural response related to error-processing which has been demonstrated during performance on the HCT (please see [35], for a more detailed description of the paradigm). It occurs at fronto-central electrodes after feedback regarding the accuracy of the performance is provided. This negative component is typically more prominent for incorrect compared to correct feedback and seems to reflect performance monitoring considering specifically the processing of external cues about one's performance [38]. This is particularly relevant in the present paradigm, where participants must adjust their response strategy according to the feedback they receive, in order to avoid making an error in the following trial. Although it has been suggested that multiple cortical sources might be involved [39, 40], the generators of the FRN have been frequently located around the anterior cingulate cortex [41]. This region plays an important role in cognitive control, planning and executing behaviour, and in adapting behaviour as a function of task demands and situations [42].

The paper is organized as follows: section 2 describes task and the signal acquisition protocol and the main steps of the signal processing methodology and the appendix complements the formulation; section 3 presents the results which is followed by the discussion (section 4) and finally the conclusion (section 5).

2. Methods

Early cognitive brain studies with ICA were focused on event-related response averages but some authors considered that event related single-trial EEG epochs would be more promising [7, 43, 44]. However, one of

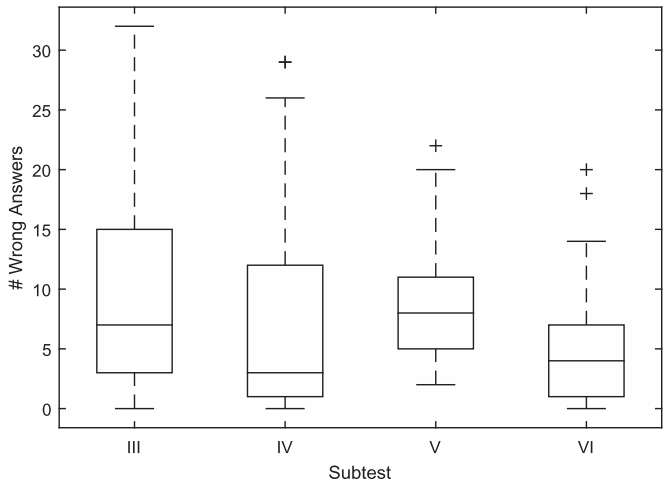


Figure 1. Number of *Wrong* answers in subtests III, IV, V and VI for all participants.

the main drawbacks of standard ICA is its permutation indeterminacy, i.e. that the extracted components are not ordered. Therefore, to limit the complexity of the selection of components, the trials are concatenated into a single data set, or ICA is applied to the continuous version of the signal. The latter is often considered after applying low-pass and high-pass filters and down-sampling, in order to assure the convergence of contrast-based ICA algorithms in a reasonable time. Furthermore, to achieve group inferences, the individual components of each subject must be matched. One strategy, borrowed from functional Magentic Ressonace is to apply ICA decomposition to a concatenated data set formed by the principal components of all subjects. The technique is called Group-ICA and the indivual componets are found by backprojecting the data [40, 45]. Another alternative is to construct feature vectors comprising characteristics of the individual components and the corresponding scalp maps and apply clustering techniques [33].

2.1. Participants and task

Fifty eight (58) participants (39 females and 19 males) were enrolled in the study. They performed a computerized version of the Halstead Category Test (HCT) which was used to assess cognitive executive frontal lobe functions. This test is used to measure a person's ability to formulate abstract principles based on receiving feedback after their response to each specific test item [35]. The EEG was continuously registered, using Ag–AgCl sintered electrodes which were placed according to the 10–20 system, with a Neuroscan SynAmps2 amplifier through an Easy-Cap with 26 channels and recorded, sampling rate of 1 kHz and with an analogue notch filter at 50 Hz, with the software Scan 4.3 (Neuroscan Systems) using the tip of the nose as reference.

The HCT involves the presentation of 208 trials divided into 7 subtests. The first two subtests, with 28 trials, are training subtests. The subtests III and IV, each with 40 trials, measure spatial/positional reasoning. The subtests V and VI, each with 40 trials, assess proportional reasoning. Finally, the last subtest is a memory test and comprises 20 trials. Each test item shows one or more figures that, altogether, suggest a number ranging from one to four. In subtests I to VI, participants are instructed to determine or guess the correct number based on their conceptualization of the abstract principle represented by the stimulus [35]. Visual feedback is provided after each response, to indicate if the participant's response was *Right* or *Wrong*. The information is presented 1500 ms after the response and remains on the screen for 750 ms. Based on this feedback, the participants must maintain or change their response strategy accordingly. The average time for the participant to answer was 1790 ms with a standard deviation of 400 ms.

Observing the sequence of responses, in general, the participants provide more often answers *Wrong* in subtest III and subtest V than in subtests IV and VI (see figure 1). Anyway, note that the number of *Wrong* answers in a total of 40 trials is usually not more than 15.

2.2. Signal processing

The main signal processing steps are the ICA decomposition, clustering of normalized scalp maps to find the prototypes which are used to look for components at trial-level. ICA is applied at single-trial level with a segment of data taken from continuous data and having as reference the participant's response. Then, for each participant, the spatial scalp maps (columns of mixing matrices) of all trials are clustered and a prototype for each cluster is estimated. This template resulting from each cluster is the average of all scalp maps of the cluster. In order to assure that scalp prototypes represent similar scalp localizations across participants, a re-ordering of these templates is performed. Afterwards, and for each trial individually, the component related with the scalp map closest to

the template was chosen as the representative of that particular cluster. These relevant cluster related components were then back-projected onto the scalp electrodes and average signals were estimated.

2.2.1. Single trial ICA

After the application of an ICA algorithm, it is possible, to explain the original trial t data with N samples of P channels as

$$\mathbf{X}^{(t)} = \mathbf{A}^{(t)}\mathbf{S}^{(t)} \tag{1}$$

where $\mathbf{X}^{(t)}$ is a $P \times N$ matrix related with the t-th trial, the matrix $\mathbf{A}^{(t)}$ is the $P \times L$ mixing matrix and $\mathbf{S}^{(t)}$ is the $L \times N$ matrix of the components. In this work, an algebraic variant of the ICA decomposition is applied to decompose the segment related with the t-th trial. The second order blind identification (SOBI) algorithm [46], without any dimension reduction ($L = P$) at the singular value decomposition pre-processing step, is applied (see appendix and [26]). The outcome of the algorithm is the separation matrix, from which it is possible to estimate the components and the mixing matrix. Note that several works report that SOBI has good performance in artifact reduction applications [10, 12].

2.2.2. Clustering scalp maps

As mentioned before, the mixing matrix is in fact estimated as the inverse (or pseudo-inverse) of the demixing (or separation) matrix which is used to estimate the components. And in most of the algorithms, the energy of the estimated components has the same value (even equal to one). Therefore, the columns of the mixing matrix give the relative projection strengths of the respective components onto each of the scalp sensors [5]. These may be interpolated [14] and the values might be normalized in order to facilitate comparisons. The latter are usually achieved by visual inspection of the color-coded column values displayed on the corresponding scalp positions. In this work, in order to apply a clustering algorithm, it is necessary to perform a normalization that allows having an identical scale in all trials. Therefore each element of the mixing matrix is written as

$$\rho_{ij} = \frac{|a_{ij}|}{\sqrt{a_{i1}^2 + a_{i2}^2 + \ldots + a_{iP}^2}} \tag{2}$$

The value ρ_{ij} measures the importance of the j-th source contribution to the i-th sensor signal. This is because the denominator of the second term of the equation involves all coefficients that contribute to the i-th sensor signal, while the numerator reflects the weight for the j-th source. These values are naturally in the range (0, 1) and the $P \times P$ normalized mixing matrix related with the t-th trial can be written as

Table 1. Scalp regions and pooling of EEG channels [47]. The EEG channels corresponding to the 10–20 montage used in the experimental set-up.

Pool	Channels name	Partitions of indexes
pre-frontal	*Fp*1, *Fp*2, *Fpz*	$J_1 = \{1, 2, 3\}$
frontal	*F*7, *F*3, *Fz*, *F*4, *F*8	$\{4, 5, 6, 7, 8\}$
fronto-central	*FC*3, *FCz*, *FC*4, *C*3, *Cz*, *C*4	$J_2 = \{9, 10, 11, 13, 14, 15\}$
parietal	*P*7, *P*3, *Pz*, *P*4 , *P*8	$J_3 = \{17, 18, 19, 20, 21\}$
parieto-occipital	*PO*7, *PO*8, *O*1, *Oz*, *O*2	$J_4 = \{22, 23, 24, 25, 26\}$
temporal	*T*7, *T*8	$J_5 = \{12, 16\}$

$$\mathbf{\Psi}^{(t)} = \begin{pmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1P} \\ \rho_{21} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{P1} & \cdots & \cdots & \rho_{PP} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\rho}_1^{(t)} & \boldsymbol{\rho}_2^{(t)} & \ldots & \boldsymbol{\rho}_P^{(t)} \end{pmatrix} \tag{3}$$

Note that j-th column $\boldsymbol{\rho}_j^{(t)}$ of the matrix $\mathbf{\Psi}^{(t)}$ represents the relative spread of the j-th source on the scalp.

The scalp map data set is obtained by concatenating the matrices $\mathbf{\Psi}^{(t)}$, $t = 1 \ldots T$. For the sake of simplicity, let's represent the participant data matrix $P \times I$, where $I = P \times T$ and where the column index i is $i = P \times (t - 1) + c$, with $t = 1, \ldots, T$ and $c = 1 \ldots P$ representing the indexes of the trials and the columns of the normalized scalp maps of each trial, respectively.

$$\mathbf{\Phi} = (\boldsymbol{\rho}_1 \quad \boldsymbol{\rho}_2 \quad \boldsymbol{\rho}_3 \quad \cdots \quad \boldsymbol{\rho}_I) \tag{4}$$

2.2.2.1. Pooling scalp maps

The P sensors are related to different positions on the scalp and might be pooled together into different scalp regions. Table 1 shows the pooling of electrodes into six scalp regions. For each normalized scalp map, a feature vector is estimated where the entries correspond to the median value of the channels pooled in each of the regions. This new data set is clustered using the *K-means* algorithm and using the cosine distance as proximity criterion.

2.2.2.2. Clustering

After the application of K-means algorithm, the set of indices is distributed among all K clusters. Therefore, the data is assumed to be partitioned in K groups (or clusters). Let us denote by $\mathbf{p}$ the vector, of length I, representing the partition of the data. The entry p_i is equal to the integer k, where $1 \leqslant k \leqslant K$, of the cluster that the normalized scalp map $\boldsymbol{\rho}_i$ belongs to.

Let us denote by $\hat{I}_k$ the set of indices i of the observations $\boldsymbol{\rho}_i$ belonging to the cluster

$$\hat{I}_k = \{i | p_i = k\} \tag{5}$$

The centroid of the k-th cluster is defined as the mean of the elements of the group

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{N_k} \sum_{i \in \hat{I}_k} \boldsymbol{\rho}_i \tag{6}$$

where N_k is the number of elements of the set $\hat{I}_k$. Different works discuss the use of scalp maps to identify artifact or task related components [4]. In this work, the centroids constitute the templates which are used to select signal components. Hence, the column of the mixing matrix close to the template is considered an implicit spatial signature for the component. Preliminary tests, using our data set, confirm the results of the referred works. In particular using $K = 5$ clusters the corresponding templates show that the largest values in each template are concentrated into distinct scalp localizations.

2.2.2.3. Ordering of templates

Different runs of *K-Means* conduct to similar templates, e.g., its visual analysis show similar patterns. However these patterns do not show up in the same order. This problem is relevant because the clustering step is applied individually to each participant. In order to combine the results across participants, it is necessary to identify the clusters according to their scalp localizations. Therefore a re-ordering step is introduced to assure that each cluster C_k represents identical scalp maps for all participants. The re-ordering is based on the computation of scores which represent the relative contribution of the pools of electrodes to the global template. For each template, $\hat{\boldsymbol{\mu}}_k$, the following scores are estimated

$$r_{kj} = \frac{\sum_{j \in J_k} \hat{\mu}_{kj}}{\sum_{j=1}^{C} \hat{\mu}_{kj}}, \quad k = 1, 2, \ldots, 5 \tag{7}$$

where the channel index partition J_k is defined as indicated in the last column of table 1. An empirical sequential procedure as described by the algorithm 1 is applied. Therefore the template that has the largest contribution among all the templates not ordered is selected as representative of each pool of electrodes.

Algorithm 1. Ordering clusters

INPUT: The estimated scores r_{kj} for all clusters.
for k = 1 **to** 5 **do**
  FIND $j^+ = \{j | r_{kj^+} > r_{kj} \forall j\}$
  C_k cluster's template is $\boldsymbol{\mu}_k = \hat{\boldsymbol{\mu}}_{j^+}$ and corresponding elements are $\mathbf{I}_k = \hat{\mathbf{I}}_{j^+}$
  $r_{kj} = 0 \forall j$
end for

The goal is to assure that the clusters of the different participants are organized according to the relevance of the scalp localizations. Hence, the cluster C_k whose members are identified by the $\mathbf{I}_k$ and the corresponding template $\boldsymbol{\mu}_k$ correspond to a similar scalp localization for all participants.

2.2.2.4. Selection of the ICA components

In this work we propose, in each trial, only to select one component according to its spread on the scalp. Note that ICA decomposes the single trial segment into P scalp maps and related components. To select components related with cluster C_k, the set of indexes for this group, $j \in \mathbf{I}_k$ are converted into a pair of indexes: the first indicating the trial and the second the column index for the mixing matrix. Those indexes are related with integer and rest of the division of the index j by P. Then, for all the elements of the partition $\mathbf{I}_k$ the steps are the following

- Compute the trial index for all $j \in \mathbf{I}_k$,

$$t = \left\lfloor \frac{j}{P} \right\rfloor + 1$$

- Compute the column number of the mixing matrix for the trial *t-th*.

$$c = j \mod P$$

More than one column might belong to *t-th* trial because the number of scalp maps is larger than the number of clusters. To select the relevant component, first it is chosen the scalp map $\boldsymbol{\rho}_{c^+}^{(t)}$ which is closer to the template $\boldsymbol{\mu}_k$. The relevant component is the one related with the selected scalp map. The *t-th* trial relevant component is back projected to the scalp by

$$\hat{\mathbf{X}}^{(t)} = \mathbf{A}_{*c^+}^{(t)} \mathbf{S}_{c^+*}^{(t)} \tag{8}$$

Which corresponds to the outer product of the selected row $c^+ - th$ of the matrix of components $\mathbf{S}^{(t)} = \mathbf{B}^{(t)} \mathbf{X}^{(t)}$ (see appendix A.1 for more details). Note that the matrix $\mathbf{X}^{(t)}$ represents the data of one trial.

2.3. Data analysis

The length of the multichannel segment to apply an ICA algorithm is an open issue which naturally has impact on the computational load and it influences the quality of the decomposition. In [12] the authors report as common choices a length of 10 *s* or a number of samples corresponding to a few times the square of number of channels. In this case, being interested in signals after the stimulus and also after the feedback, which occurs 1500 ms after the participant's response onset (to analyse the *P*100 and the FRN potentials, respectively), the multichannel EEG was segmented from 6000 ms prior to response onset to 3000 ms after, corresponding to 9000 samples. Note that in this way, at each trial level, the segment comprises the following sequence of time stamps: presentation of the stimulus, the participant's response and the feedback information. As referred before, the time interval between stimulus presentation and the participant's response is variable, while the time interval between the participant's response and the presentation of feedback is 1500 ms (for more details see [35]).

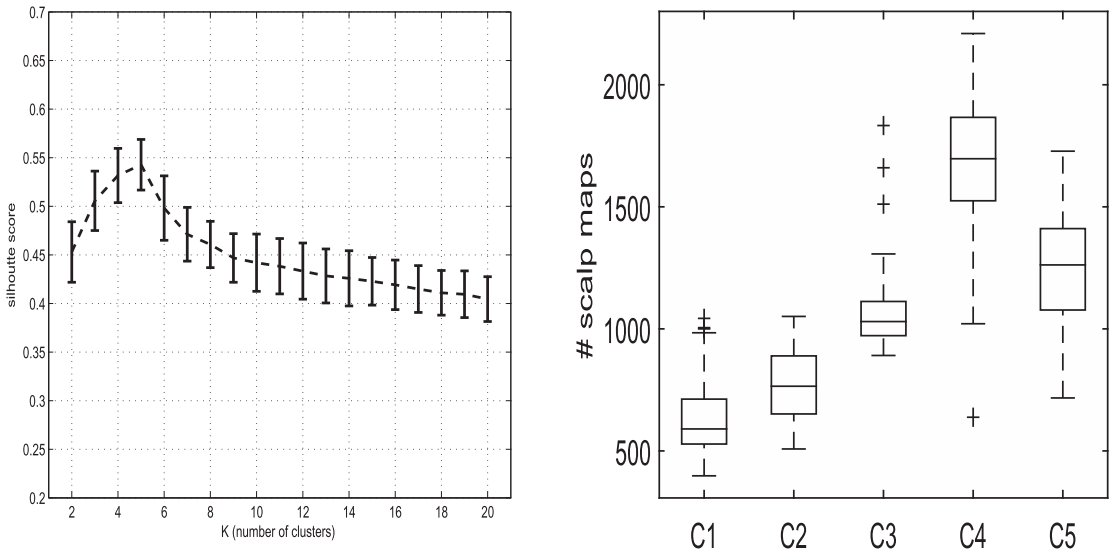


Figure 2. Clustering the scalp maps of all participants: *Left*: Silhouette score (mean and standard deviation) varying the number of clusters; *Right*: $K = 5$ clusters—size of the clusters.

The data segments were bandpass filtered in the range (1–40) Hz using the discrete Fourier transform. After the application of the FFT algorithm, the coefficients outside of the defined range were eliminated and the inverse discrete Fourier Transform (IFFT) was then applied.

The 26×26 ICA separation matrix was estimated using Second Order Blind Source Separation (SOBI) algorithm [46]. The mixing matrix for each trial and the corresponding signal components were then estimated as described in appendix A.1.

2.3.1. Scalp maps and selection of components

The clustering step was performed as described before. The number of clusters K is a pre-assigned value which is often determined by trial and error and validated experimentally. The quantitative evaluation of clustering algorithms is often achieved using a global measure that reflects the intra-cluster versus the inter-cluster similarities [48, 49]. The silhouette value is a common choice to evaluate the quality of clustering algorithms. Figure 2 shows the mean silhouette score assigning the number of clusters $K = 2$ to $K = 20$ to partition the scalp maps of all participants of the study. The average value of silhouette close to one means that the data is appropriately clustered (see appendix A.3). In the different runs of the algorithm, the best values are achieved with $K = 5$ to $K = 7$ (see figure 2). The study was finally conducted with $K = 5$ because the reordering of templates gave visually comparable patterns across participants.

Figure 3 shows the cluster templates (the centroids of the clusters) of the normalized scalp maps of 5 participants. Although there is variability among participants, the different clusters still show different scalp localizations. In all cases the differences among the participants concern the symmetry and the focus of the scalp patterns. Some participants present symmetric patterns while in others the activity is more concentrated on one of the sides of the scalp. Some participants have scalp values that change abruptly, corresponding to a more precise localization, while in others the values spread smoothly over the scalp with peaks occurring in distinct scalp regions. For instance, the template corresponding to cluster C_1 (first column of figure 3) is localized in the frontal region, but in participant 1 it is focused while in participant 2 it is diffuse. In participant 1 a symmetric pattern is observed while in participant 4 there is a non-symmetric pattern.

The cluster C_2 (second column of figure 3) represents the fronto-central region and the participants have symmetric and non-symmetric patterns. However all of them involve the most relevant electrodes of this region which are *FCz* and *Cz*. The cluster C_3 (3rd column of the figure) represents mostly the parietal region but the spread is also different among the participants resulting in more or less focused patterns. The cluster C_4 (4th column of the figure) is related with the occipital region and cluster C_5 (5th column of the figure) is related with temporal regions. The latter also shows symmetric and non-symmetric patterns among hemispheres.

Figure 2 (on the right) shows the number of scalp maps belonging to the different regions in the total of $208 \times 26 = 5408$ scalp maps for each participant. These results show that it is possible to find members of all clusters in the 26 scalp maps of every trial. The cluster C_1 has the smallest number of elements. Note that, in spite of the filtering operation before ICA,

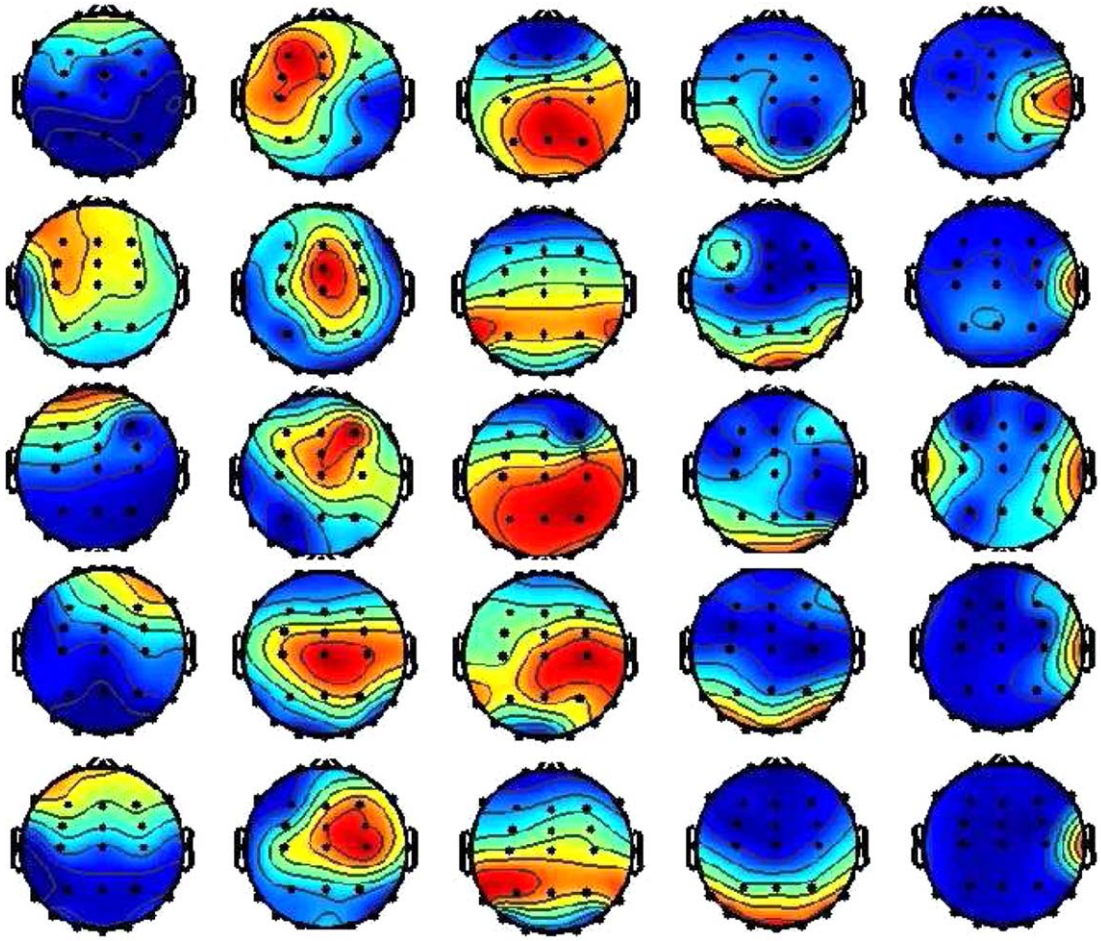


Figure 3. Centroids of the cluster for 5 participants: Rows (participants) columns centroids of clusters: $1st - C_1$, $2nd - C_2$, $3rd - C_3$, $4th - C_4$ and $5th - C_5$. For visualization purposes each scalp map was normalized individually. Minimum amplitude (very small positive number) corresponds to blue and maximum to red.

there are 2–3 components which are artifact-related components in the 26 components of each trial.

3. Results

Regarding the previously described clusters that emerged from the component selection process, the cluster C_4 (occipital region) is the one that best provides the scalp signature of the *P*100 component, while the cluster C_2 provides the signature of the FRN (fronto-central region, encompassing electrodes *FCz* and *Cz*, where the FRN is most conspicuous).

In the following sections, we will use these two case studies to demonstrate the effectiveness of the previously described signal processing methodology. This methodology allows applying ICA at the single trial level using the unsupervised learning algorithm K-means to automatically select the relevant components and enhance the signal patterns related to specific neural activity. We will thus demonstrate that the waves *P*100 and FRN are identifiable in the reconstructed signal by back-projection of the relevant components to the scalp, both when average signals are computed and at the single-trial level.

3.1. P100

After the selection of the component of each trial related with clusters C_4 and C_3 (C_3 was also included in the analysis in order to clearly demonstrate that the P100 component is indeed more prominent in cluster C_4), the component segment centered on the onset of the visual stimuli (from 1000 ms prior to visual stimuli onset to 2000 ms after) is back-projected to the scalp.

Figure 4 represents the ERP image [7] and the ERP average of one participant using the back-projected signal into *O*1. The positive deflection around 100 ms is present either using C_3 or C_4 components, although being more evident if the scalp map of the cluster C_4 is considered, as expected.

The P100 pattern clearly visible in the average ERP with a peak with latency equal to 100 ms, is also visible in the ERP image where a larger intensity value is visible with latency of 100 ms across all trials.

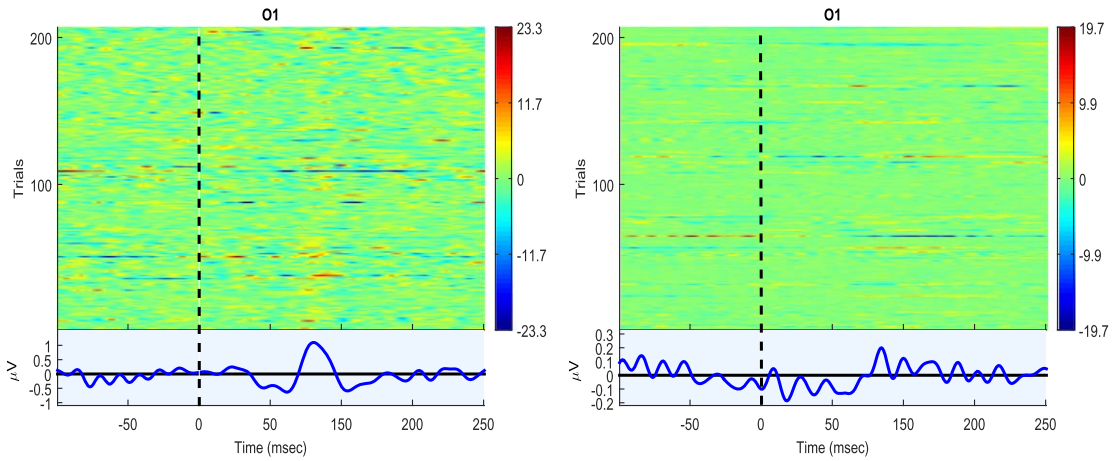


Figure 4. ERP image of the back-projected components in O1 channel for one participant. The vertical traces indicate stimulus presentation. *Left*: Component related with cluster C_4 ; *Right*: Component related with cluster C_3.

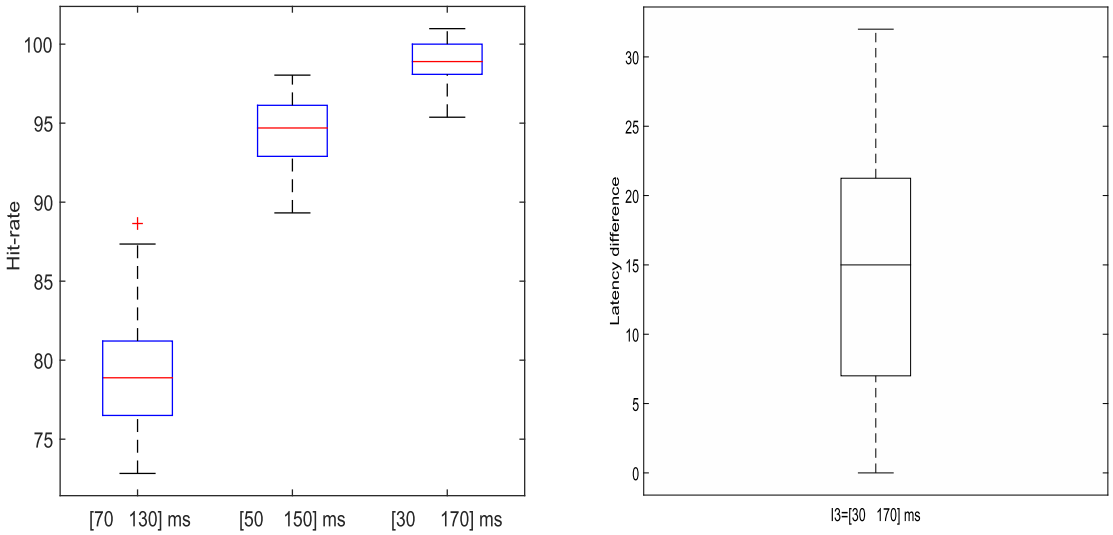


Figure 5. *Left*: Hit-rate for finding the *P*100 within the respective interval, considering all participants and three latencies intervals. *Right*: Absolute differences between P100 latency of ERP average signal and the single-trial P100 using I_3 interval.

3.1.1. Single trial P100

The *P*100 event-related potential, defined as a positive deflection around 100 ms, is detected automatically for all trials of all subjects defining three different intervals to look for the local peak maximum. In the literature, the variability of the latency of the P100 in single trial detection is also described [50]. The three intervals are I_1: [70 130] ms, I_2: [50 150] ms and I_3 [30 170] ms. Our goal was to find out if the pattern is present in all trials and to understand the latency variability. The hit rate was defined as the number of trials where the condition is accomplished, divided by the total number of trials (208). Figure 5 represents the results for all participants.

The larger interval allows to identify the positive deflection at least in 95% of the trials. These single-trial *P*100 peaks were compared with the peak detected on the participants' ERP average signal. The absolute differences, for all participants and trials, are also shown in figure 5. It is possible to see that the median absolute difference is around 15 ms, which is a value smaller than what is observed between ERP averages [50].

3.2. FRN single-trial level

In [35] it was shown that the FRN has larger amplitude after the feedback of *wrong* answers, i.e., it was demonstrated that wrong answers elicit an FRN of larger amplitude than correct answers, given that this ERP component is related to our error-monitoring system, which allows to adapt behavioral responses based on the success of previous performance.

In this case, we were interested in studying the FRN at the subtest level and trial level. In particular, we were interested in subtest transitions, when participants are still unsure about the abstract principle underlying the following subtest. At the beginning of the test, participants are informed that the underlying

Table 2. Performance in the first three trials: entries represent the number of participants answering 0, 1, 2 or 3 times *wrong*.

Number of *wrong*	0	1	2	3
sub-test III	7	7	30	14
sub-test IV	31	19	5	3

abstract principle may change from one subtest to the next. Sometimes this principle remains the same, and other times it changes. Participants are never sure of this on the first trials, and they will only be able to determine the correct abstract principle on the basis of the feedback they receive about their performance.

Table 2 shows how the participants perform in the first three trials of each subtest. The number participants giving a higher number of*Wrong* answers is larger on sub-test III. It has to be noticed that, in subtest IV, more than half of the participants do not give any *Wrong* answer. On subtest IV the principle did not change from the previous subtest.

The ERP single-trial signal, resulting by back-projecting the components related to cluster C_2, was studied and the mean of the three first trials of subtest III and subtest IV of the HCT was estimated. Then, a negative deflection was detected in the interval [150 350] ms. Figure 6 (left side) shows the amplitude of the FRN for the group of participants which have similar values in both subtests. And in both subtests there are outliers. The same figure (right side) also illustrates the range of latencies of the wave. It has to be noticed that, although the median value is similar in both subtests, in subtest IV the range is larger. Figure 7 shows an ERP like image with the signals of all participants. Note that for visualization purposes the images only show truncated values of the amplitudes to avoid the influence of the outliers.

3.3. FRN—Case study of one participant

As referred before, the participants must adjust their response strategy according to the feedback they receive, and therefore their performance changes along each subtest.

Figure 8 shows the sequence of *Wrong* and *Right* answers in two subtests of one participant. The goal of the following study was to compare the wave patterns in three performance scenarios. The consecutive trials were chosen as follows:

- Case1: The first three trials (t_1, t_2 and t_3). Note that in both sub-tests the answers are type *Wrong* and *Right*.
- Case2: Three trials in the middle (t_{25}, t_{26} and t_{27}). In this case the answer is always *Right* in both subsets.
- Case3: Three consecutive trials (t_{19}, t_{20} and t_{21} subtest III) corresponding to answers type *Wrong* (t_{19}, t_{20}) and *Right* (t_{21}) after a long sequence of answers *Right*.

The following figures represent the analysis of the dataset considering the criteria described above. Regardless of the type of response, the FRN wave, a negative deflection with latency around 250 ms, is visible in the first trials of each subtest (see figures 9 and 10 on the left—top pannel in each figure, Case 1). The FRN is also visible if, after a sequence of answers type *Right*, the participant gives answers type Wrong, as it is visible in figure 9 (bottom pannel in the figure, Case 3). The FRN is not so clearly present if the answers are of type *Right* after a long sequence of correct answers (Case 2). The topoplots of the average amplitudes in the defined interval also reflect the presence of a clear negativity associated with the FRN. When the FRN wave pattern is present, the signals in fronto-central channels are clearly negative (see Case 1 and Case 3 of figure 9). If the FRN pattern is not so clear, either the negativity has smaller amplitude (subtest III, Case 2) or does not appear (subset IV, Case 2). This strategy of analysis was repeated with the signals of different participants and similar outcomes were obtained. In the supplementary material available online at stacks.iop.org/BPEX/5/015019/mmedia results concerning four more participants are discussed. This preliminary results show that the FRN wave reflects the performance of the participant. Its amplitude is higher in the case of *Wrong* answers or in the transition of the test when the participant is not certain about the abstract principle of the stimulus.

4. Discussion

EEG registration while participants answer to the HCT test provides an acquisition scenario where it is useless to instruct participants to avoid movements or blinks. Therefore the ERP analysis should be conducted in pre-processed data because the number of trials is too small. The ERP study [35] shows that the FRN is visible and its amplitude is different in the average of *Right* and *Wrong* answers. In that work, the average signals were pre-processed with a singular spectrum analysis filter.

The HCT is a neuropsychological test that traditionally provides only an overall error score indicative of global impairment [51]. However, recent works have attempted to develop and validate several scoring subscales that allow measuring cognitive skills in more detail and provide more information about the individuals performance, which increases the clinical utility of the test [52, 53]. Some of these subscales (ex., the CAT-2 and CAT-2A subscales) highlight the importance of examining how performance changes across a sequence of trials, instead of an overall score or a single trial, in order to evaluate the process of learning and applying the underlying abstract principle [52]. The

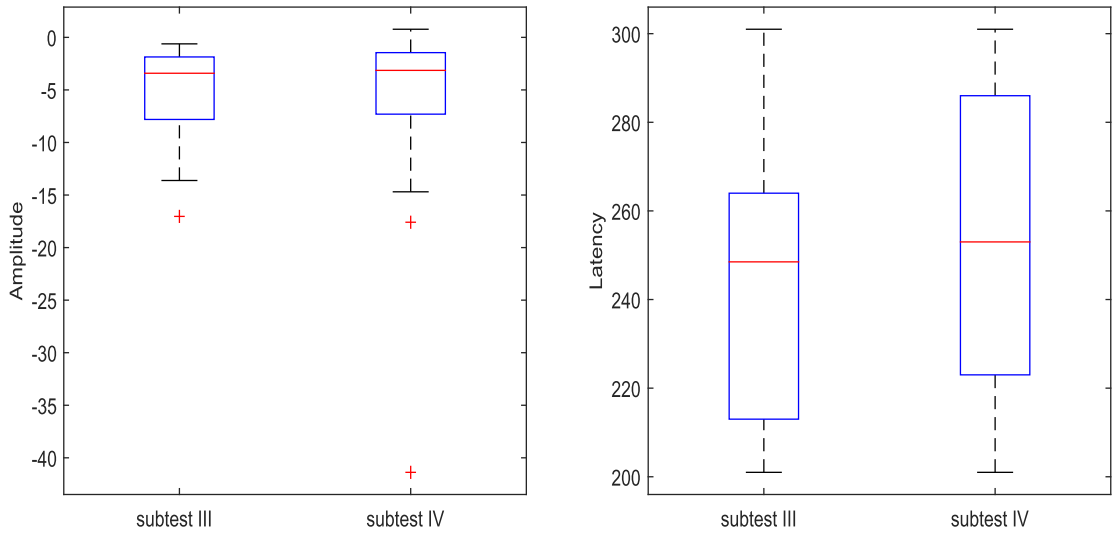


Figure 6. Amplitude and Latency of the FRN wave in all participants considering the mean of first three trials in each subtest III and subtest IV of the HCT.

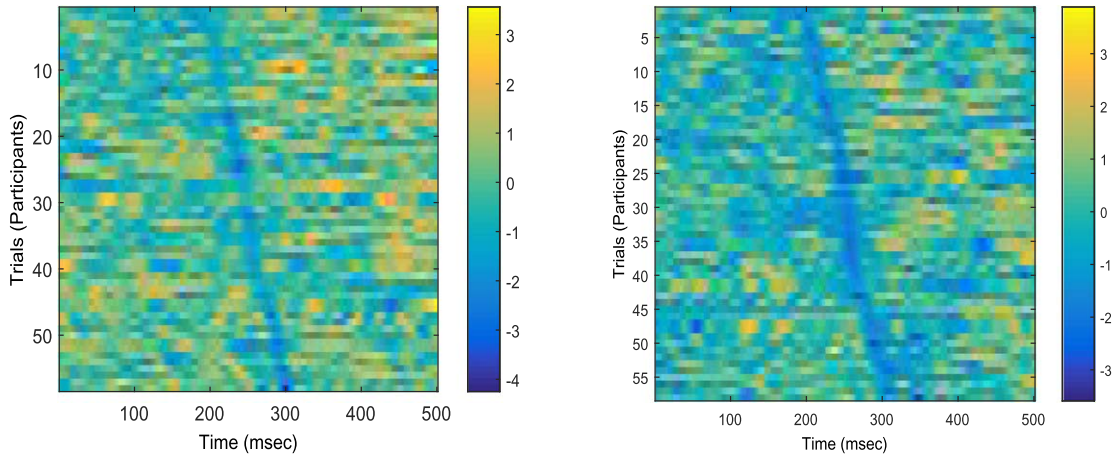


Figure 7. ERP image of the back-projected components related to cluster C_2 in channel *Cz* for all participants. *Left*: subtest III of the HCT; *Right*: subtest IV of the HCT. The participant's signals are ordered according to the latency of the detected peak.

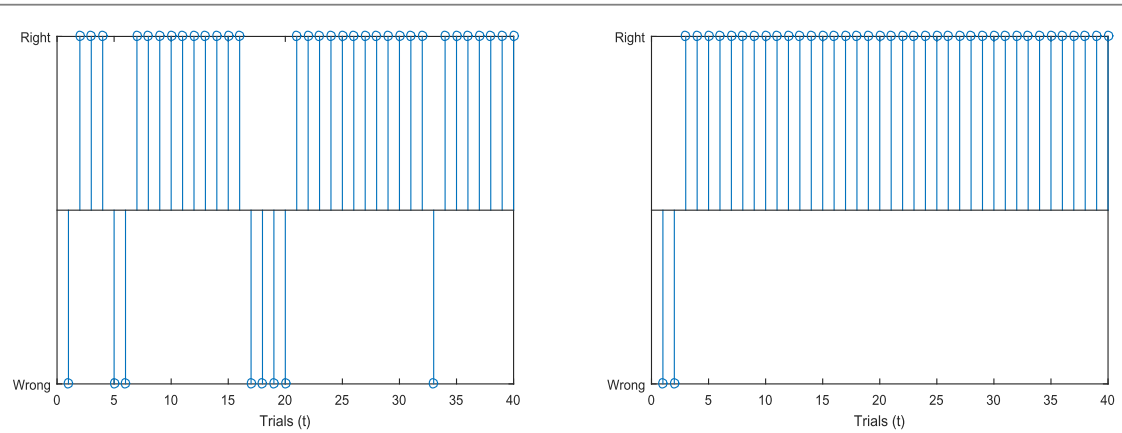


Figure 8. Sequence of Wrong and Right answers: *Left*: subtest III *Right* and subtest IV

coupling of information about brain activity recorded during test performance can be useful in increasing the diagnostic potential of the HCT. Therefore is it important to develop signal analysis techniques that allow examining brain activity on a trial-by-trial basis. In the present work, in order to relate the performance

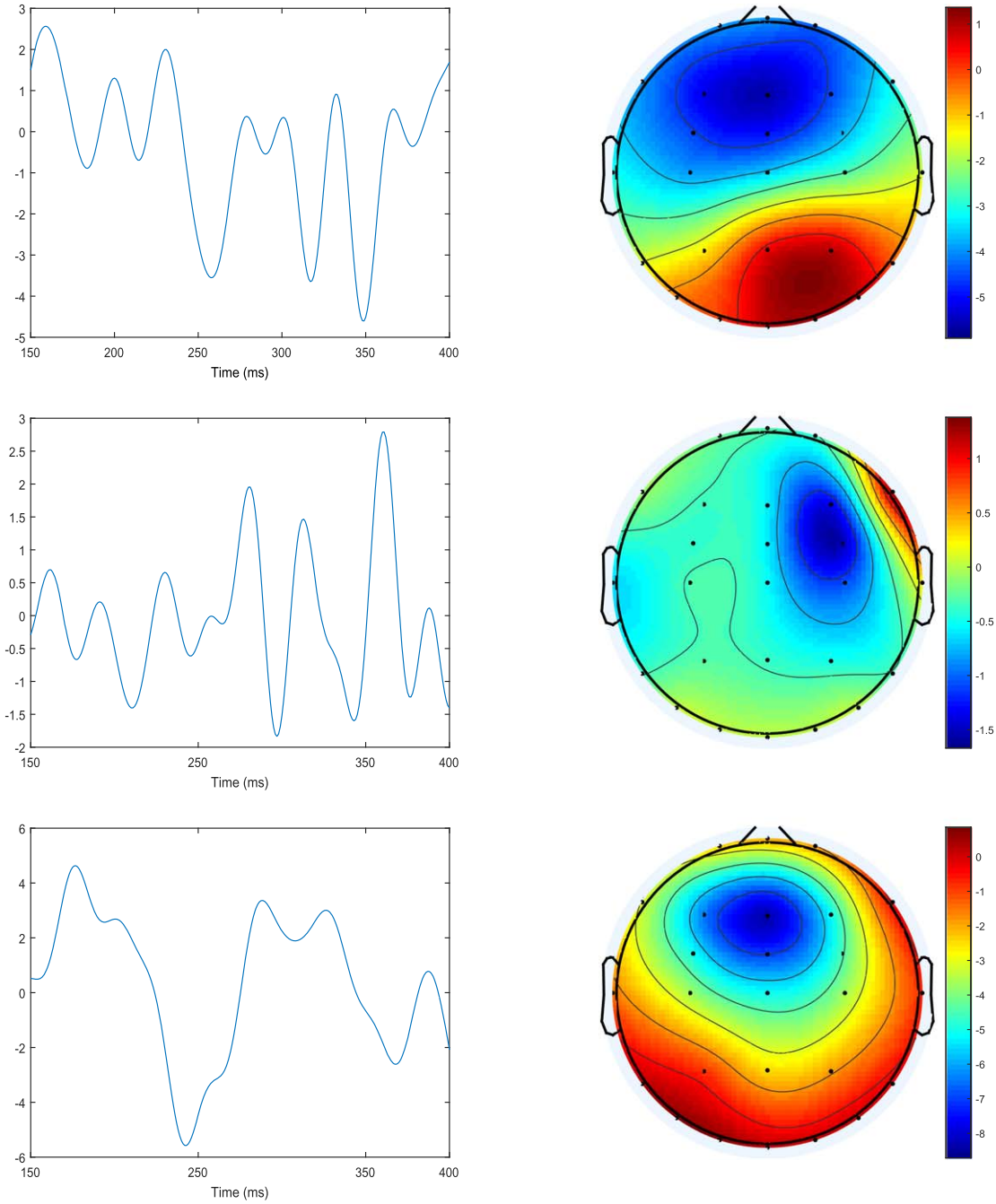
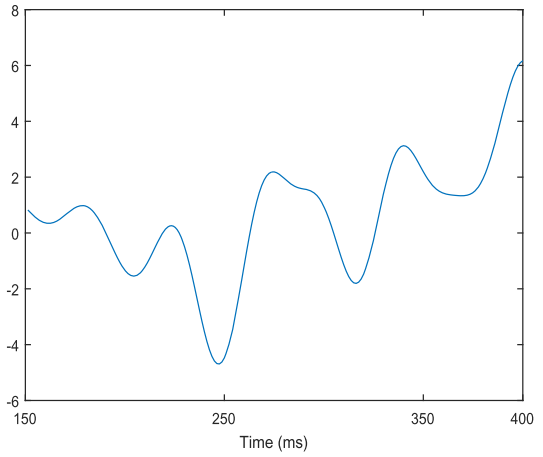
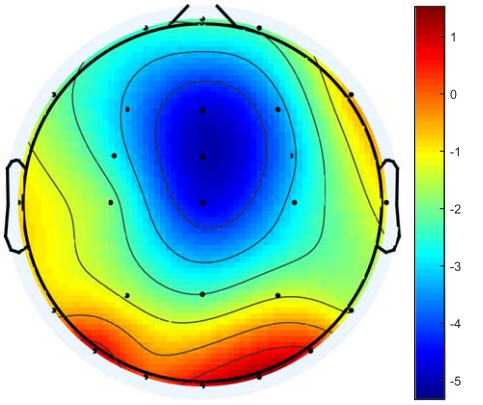
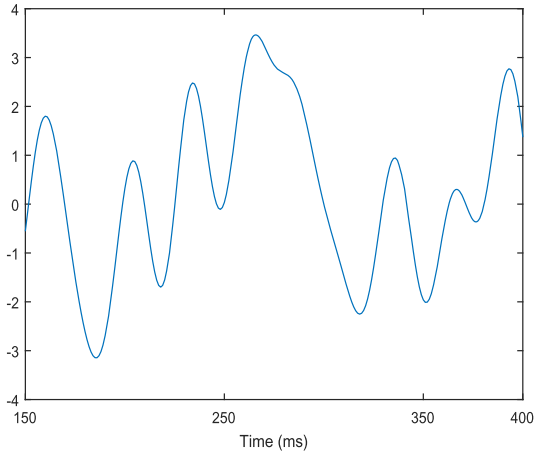


Figure 9. Subtest III: *Left*: ERP signal in FCZ; *Right*: topoplots associated to each ERP signal in all electrodes, between [230 270] ms. *Top* to *Bottom*: Case 1, Case 2 and Case 3.

of the user in short sequences of answers with the corresponding signals, ICA at single-trial level was applied. The study shows that it is possible to find back-projected ICA components where ERP events are clearly visible. The *P*100 study shows that ERP image of the proper back-projected component is clearly visible on the occipital channel. The single-trial *P*100 peak latencies and the peak latency detected on the participant's average signal have differences whose median is around 15-ms which is a value smaller than what is observed between ERP averages [50]. Related with the HCT test, the FRN wave is visible in all participants in transitions between sub-tests either when the principle changes (sub-test III) or when does not change (sub-test IV) even when the partipant is answering *Right* , but is still unsure about the abstract principle in that particular subtest. The exploratory study conducted with different participants indicates that the FRN wave is related with the performance of the user. This is a very important result because it demonstrates that the proposed analysis method allows the detailed inspection of this potential on a trial-by-trial basis. With this detailed analysis, it is possible to make inferences about the underlying cognitive processing and how the participant approaches each individual stimulus in terms of their certainty







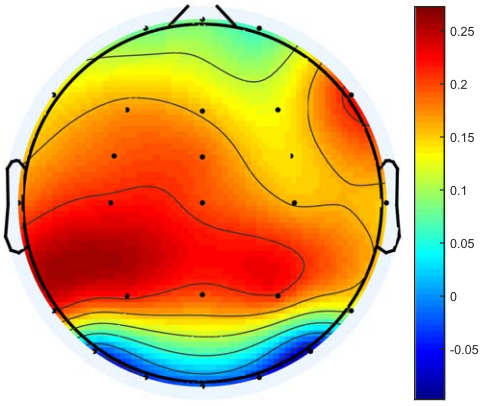


Figure 10. Subtest IV: *Left*: ERP signal in FCZ; *Right*: topoplots associated to each ERP signal, in all electrodes between [230 270] ms. *Top* to *Bottom*: Case 1 and Case 2.

about the underlying abstract principle. This knowledge is important to understand the neural mechanisms associated with executive function and abstract reasoning.

5. Conclusion

In this work a new approach is proposed to deal with the main drawbacks of independent components analysis: ordering and scaling indeterminacies. The ordering drawback was achieved by introducing a well known algorithm of clustering. This step proves to be essential in order to accomplish the application of the algorithm at the single-trial level. Furthermore, with a simple heuristic the centroids (templates) of the clusters of the different participants are pairwise to perform inter-participant analysis. The scaling indeterminacy was partially solved by choosing only one component by trial. The main idea was to select the component that can be the best representative of each cluster in a particular trial. The experimental study conducted with EEG data acquired with participants performing the HCT test shows the usefulness/viability of the methodology. The P100 was detected and characterized at the single-trial level and the study shows that the characteristics of this visual potential are according to what is expected. However, the more promising application was the detection of the FRN potential and the possibility of using its detection to study the dynamics of the brain while assessing performance on a specific task, monitoring performance for errors and adjusting behavior on the basis of feedback. The analysis of a sequence of three trials in different context shows that the amplitude of the FRN is related with the performance of the user in the HCT test. Further developments of this work are to study the dynamics of the signals detected and to relate them with the participants' performance.

Acknowledgments

This work was partially supported by national funds through FCT—Foundation for Science and Technology, within the R&D Units IEETA (UID/CEC/00127/2013) and CINTESIS (UID/IC/4255/2013), and by European funds through FEDER, under the COMPETE 2020 and Portugal 2020 programs, in the context of the projects PTDC/EEI-SII/6608/2014 and POCI-01-0145-FEDER-007746. This research

was also supported by Bial Foundation with the Grant ref. 136/08 to Isabel M Santos.

Appendix

A.1. Blind source separation model

Most of BSS/ICA algorithms follow a two step procedure to estimate the separation (or de-mixing) matrix [26]. The first step is based on Principal Component Analysis or Singular Value Decomposition (SVD) of the data matrix **X** [26]. For the second step, different approaches have been proposed [26, 54]. The separation matrix **B** is estimated as the product of matrices computed in both steps. For convenience, these steps are reviewed

- Performing the SVD of the original data **X**, two matrices are computed: the $C \times L$ eigenvector matrix **V** and the $L \times L$ diagonal singular value matrix **D**. The user can decide to perform dimension reduction after the first step by discarding the smallest singular values and corresponding eigenvectors. In that case $L < P$ and the number of sources will be smaller than the number of measured signals. Without dimension reduction, the number of singular values and corresponding eigenvectors is $L = P$.
- After whitening the original data **X**, i.e $\mathbf{Z} = \mathbf{D}^{-1}\mathbf{V}^{T}\mathbf{X}$, time-delayed correlation matrices are estimated [26, 46]. The approximate joint diagonalization of this set of matrices leads to an orthogonal matrix **W**.

The separation matrix is defined as $\mathbf{B} = \mathbf{W}^{T}\mathbf{D}^{-1}\mathbf{V}^{T}$ and its pseudo-inverse as $\mathbf{A} = \mathbf{VDW}$. The mandatory parameter of this algorithm is Q, the number of matrices of the second step. Eventually, the user can decide to perform dimension reduction after the first step by discarding the smallest singular values and corresponding eigenvectors. In this work, the dimension was maintained equal to the number of sensors (channels), i.e. C, and Q is assigned to 100. And note that the energy of the components, i.e., the rows of the matrix $\mathbf{S} = \mathbf{BX}$, is equal to one. For the sake of simplicity the identification of each trial by a proper index is omitted here as it is explained in the text. But remember, the ICA algorithm is applied to each trial separately.

A.2. K-means and cosine distance

K-means is a recursive procedure which groups the data into K clusters by computing a similarity measure between data instances and the cluster's centroid. The inputs of the algorithm are the number of clusters and the choice of the similarity measure. In this work the cosine distance was applied. Considering two data instances ($\mathbf{a}$, $\mathbf{b}$), the distance is defined as

$$d(\mathbf{a}, \mathbf{b}) = 1 - \frac{\mathbf{a}^{T}\mathbf{b}}{\|\mathbf{a}\|_2 \|\mathbf{b}\|_2}$$

where $\|.\|$ means the length of the argument, then the second term on the right side of the equality is the cosine of the angle between the two vectors It has to be noticed that $d(\mathbf{a}, \mathbf{b}) = d(\mathbf{a}, \alpha\mathbf{b})$, $\alpha > 0$. And if the vectors $\mathbf{a}$, $\mathbf{b}$ have the same direction, the distance is zero. Note that all the entries of the normalized scalp maps are positive, so the distance ranges between (0–1).

A.3. Silhouette score

The score [48, 49] is estimated individually for each *i-th* example of the data set as

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

where $a(i)$ represents the intra-cluster distance, eg, is the average distance of the *i-th* example to all the elements of its cluster while $b(i)$ represents the inter-cluster distance, then is the average distance of the *i-th* example to the examples of the closest cluster. Therefore ideally $a(i) \ll b(i)$ and the $s(i)$ is close to 1, meaning that the example $s(i)$ is properly assigned to its cluster. A global silhouette score is the average of the individual scores allowing to have a global score to evaluate the clustering results.

ORCID iDs

Ana R Teixeira https://orcid.org/0000-0002-8120-0148

References

[1] Palaniappan R and Mandic D P 2007 Biometrics from brain electrical activity: a machine learning approach *IEEE Trans. Pattern Anal. Mach. Intell.* **29** 738–42
[2] Khushaba R N, Kodagoda S, Lal S and Dissanayake G 2011 Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm *IEEE Trans. Biomed. Eng.* **58** 121–31
[3] Frantzidis C A, Bratsas C, Papadelis C L, Konstantinidis E, Pappas C and Bamidis P D 2010 Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli *IEEE Transactions on Information Technology in Biomedicine* **14** 589–97
[4] Chuang C H, Ko L W, Lin Y P, Jung T P and Lin C T 2014 Independent component ensemble of eeg for brain-computer interface *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **22** 230–8
[5] Jung T P, Makeig S, Westerfield M, Townsend J, Courchesne E and Sejnowski T J 2000 Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects *Clinical Neurophysiology* **111** 1745–58
[6] Makeig S, Westerfield M, Jung T P, Covington J, Townsend J, Sejnowski T J and Courchesne E 1999 Functionally independent components of the late positive event-related potential during visual spatial attention *The Journal of Neuroscience* **19** 2665–80
[7] Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis *J. Neurosci. Methods* **134** 9–2

[8] Delorme A, Sejnowski T and Makeig S 2007 Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis *NeuroImage* **3** 1443–9

[9] Daly I, Nicolaou N, Nasuto S J and Warwick K 2013 Automated artifact removal from the electroencephalogram: a comparative study *Clinical EEG and Neuroscience* **44** 291–306

[10] Iriarte J, Urrestarazu E, Valencia M, Alegre M, Malanda A, Viteri C and Artieda J 2003 Independent component analysis as a tool to eliminate artifacts in EEG: a quantitative study *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society* (https://doi.org/10.1097/00004691-200307000-00004)

[11] Jung T P, Makeig S, Westerfield M, Townsend J, Courchesne E and Sejnowski T J 2001 Analysis and visualization of single-trial event-related potentials *Human Brain Mapping* **14** 166–85

[12] Urigüen J A and Garcia-Zapirain B 2015 EEG artifact removal-state-of-the-art and guidelines *J. Neural Eng.* **12** 031001

[13] Chaumon M, Bishop D V and Busch N A 2015 A practical guide to the selection of independent components of the electroencephalogram for artifact correction *J. Neurosci. Methods* **250** 47–63

[14] Radüntz T, Scouten J, Hochmuth O and Meffert B 2015 EEG artifact elimination by extraction of ICA-component features using image processing algorithms *J. Neurosci. Methods* **243** 84–93

[15] Radüntz T, Scouten J, Hochmuth O and Meffert B 2017 Automated EEG artifact elimination by applying machine learning algorithms to ICA-based features *J. Neural Eng.* **14** 046004

[16] Esfahani E T and Sundararajan V 2012 Classification of primitive shapes using brain-computer interfaces *CAD Computer Aided Design* **44** 1011–9

[17] van Gerven M *et al* 2009 The brain-computer interface cycle *J. Neural Eng.* **6** 041001

[18] Hidalgo-Munoz A, López M, Santos I, Pereira A, Vázquez-Marrufo M, Galvao-Carmona A and Tomé A 2013 Application of SVM-RFE on EEG signals for detecting the most relevant scalp regions linked to affective valence processing *Expert Syst. Appl.* **40** 2102–8

[19] Serby H, Yom-Tov E and Inbar G F 2005 An improved P300-based brain-computer interface *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **13** 89–98

[20] Blankertz B, Tomioka R, Lemm S, Kawanabe M and Muller K R 2008 Optimizing spatial filters for robust EEG single-trial analysis *IEEE Signal Process. Mag.* **25** 41–56

[21] Cohen M X 2017 Comparison of linear spatial filters for identifying oscillatory activity in multichannel data *J. Neurosci. Methods* **278** 1–12

[22] Wang Y and Jung T P 2013 *Improving Brain- Computer Interfaces Using Independent Component Analysis* (Berlin Heidelberg: Springer) pp 67–83

[23] Treder M, Porbadnigk A, Shahbazi Avarvand F, Müller K and Blankertz B 2016 The lda beamformer: optimal estimation of erp source time series using linear discriminant analysis *Neuroimage* **129** 279–91

[24] Congedo M, Korczowski L, Delorme A and Lopes da Silva F 2016 Spatio-temporal common pattern: a companion method for erp analysis in the time domain *J. Neurosci. Methods* **267** 74–88

[25] Rivet B, Souloumiac A and Attina V 2009 Xdawn algorithm to enhance evoked potentials: application to brain-computer interface *Applied Neuropsychology: Adult* **56** 2035–43

[26] Common P and Jutten C 2010 *Handbook of Blind Source Separation: Independent Component Analysis and its Applications* (New York: Academic)

[27] Hesse C W and James C J 2006 On semi-blind source separation using spatial constraints with applications in EEG analysis *IEEE Trans. Biomed. Eng.* **53** 2525–34

[28] Vos M D, Lathauwer L D and Huffel S V 2011 Spatially constrained ICA algorithm with an application in EEG processing *Signal Process.* **91** 1963–72

[29] Götz T, Stadler L, Fraunhofer G, Tomé A M, Hausner H and Lang E W 2017 A combined cICA-EEMD analysis of EEG recordings from depressed or schizophrenic patients during olfactory stimulation *J. Neural Eng.* **14** 016011

[30] James C J and Gibson O J 2003 Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis *IEEE Trans. Biomed. Eng.* **50** 1108–16

[31] Lu W and Rajapakse J C 2005 Approach and applications of constrained ICA *IEEE Trans. Neural Netw.* **16** 203–12

[32] Zhang Z-L 2008 Morphologically constrained ICA for extracting weak temporally correlated signals *Neurocomputing* **71** 1669–79

[33] Onton J and Makeig S 2006 Information-based modeling of event-related brain dynamics *Progress in Brain Research* **159** 99–120

[34] Vijean V, Hariharan M, Yaacob S and Sulaiman M N B 2014 Application of clustering techniques for visually evoked potentials based detection of vision impairments *Biocybernetics and Biomedical Engineering* **34** 169–77

[35] Santos I *et al* 2016 ERP correlates of error processing during performance on the Halstead Category Test *International Journal of Psychophysiology* **106** 97–105

[36] Luck S J 2005 *An Introduction to the Event-Related Potential Technique* (Cambridge, MA: MIT Press)

[37] Pratt H 2011 *Sensory ERP components* (New York: Oxford University Press)

[38] Bernat E M, Nelson L D, Steele V R, Gehring W J and Patrick C J 2011 Externalizing psychopathology and gain/loss feedback in a simulated gambling task: dissociable components of brain response revealed by time-frequency analysis *Journal of Abnormal Psychology* **120** 352–64

[39] Gehring W J and Willoughby A R 2002 The medial frontal cortex and the rapid processing of monetary gains and losses *Science* **295** 2279–82

[40] Kokmotou K, Cook S, Xie Y, Wright H, Soto V, Fallon N, Giesbrecht T, Pantelous A and Stancak A 2017 Effects of loss aversion on neural responses to loss outcomes: An event-related potential study *Biological Psychology* **126** 30–40

[41] Walsh M M and Anderson J R 2012 Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice *Neuroscience and Biobehavioral Reviews* **36** 1870–84

[42] Hauser T U, Iannaccone R, Stämpfli P, Drechsler R, Brandeis D, Walitza S and Brem S 2014 The feedback-related negativity (FRN) revisited: new insights into the localization, meaning and network organization *NeuroImage* **84** 159–68

[43] Stewart A X, Nuthmann A and Sanguinetti G 2014 Single-trial classification of EEG in a visual object task using ICA and machine learning *J. Neurosci. Methods* **228** 1–14

[44] Williams N, Nasuto S and Saddy J 2015 Method for exploratory cluster analysis and visualisation of single-trial ERP ensembles *J. Neurosci. Methods* **250** 22–33

[45] Eichele T, Rachakonda S, Brakedal B, Eikeland R and Calhoun V 2011 EEGIFT: group independent component analysis for event-related EEG data *Computational Intelligence and Neuroscience* **44** 129365

[46] Belouchrani A, Abed-Meraim K, Cardoso J F and Moulines E 1997 A blind source separation technique using second-order statistics *IEEE Trans. Signal Process.* **45** 434–44

[47] Kong W, Zhou Z, Hu S, Zhang J, Babiloni F and Dai G 2013 Automatic and direct identification of blink components from scalp eeg *Sensors* **13** 10783–801

[48] Fujita A, Takahashi D Y and Patriota A G 2014 A non-parametric method to estimate the number of clusters *Comput. Stat. & Data Analysis* **73** 27–39

[49] Rousseeuw P J 1987 Silhouettes: a graphical aid to the interpretation and validation of cluster analysis *J. Comput. Appl. Math.* **20** 53–65

[50] Ahmadi M and Quiroga R Q 2013 Automatic denoising of single-trial evoked potentials *NeuroImage* **66** 672–80
[51] Choca J P, Laatsch L, Wetzel L and Agresti A 1997 The halstead category test: a fifty year perspective *Neuropsychology Review* **7** 61–75
[52] McNally S, Dsurney J, McGovern J, DeFilippis N and Chan L 2016 Concurrent validity of new subscale scores for the booklet category test *Assessment* **23** 333–41
[53] Roye S, Calamia M, Greve K, Bianchini L, Aguerrevere K and Curtis K 2016 Further validation of booklet category test subscales for learning, set loss, and memory in a mixed clinical sample *Applied Neuropsychology: Adult* **25** 11–8
[54] Tomé A M and Lang E 2003 Approximate diagonalization approach to blind source separation with a subset of matrices *Seventh International Symposium on Signal Processing and Its Applications, 2003. Proceedings* **vol 2** 105–8 IEEE