



## Fuzzy Clustering for Exploratory Analysis of EEG Event-Related Potentials

Masulli, Paolo; Masulli, Francesco; Rovetta, Stefano; Lintas, Alessandra; Villa, Alessandro E.P.

*Published in:*  
IEEE Transactions on Fuzzy Systems

*Link to article, DOI:*  
[10.1109/TFUZZ.2019.2910499](https://doi.org/10.1109/TFUZZ.2019.2910499)

*Publication date:*  
2019

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Masulli, P., Masulli, F., Rovetta, S., Lintas, A., & Villa, A. E. P. (2019). Fuzzy Clustering for Exploratory Analysis of EEG Event-Related Potentials. *IEEE Transactions on Fuzzy Systems*, 28(1), 28 - 38.  
<https://doi.org/10.1109/TFUZZ.2019.2910499>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Fuzzy Clustering for Exploratory Analysis of EEG Event-Related Potentials

Paolo Masulli, *Member, IEEE*, Francesco Masulli, *Senior Member, IEEE*,  
Stefano Rovetta, *Senior Member, IEEE*, Alessandra Lintas, Alessandro E. P. Villa

**Abstract**—We introduce an analysis method for electroencephalography (EEG) data, focused on Event-Related Potentials (ERPs). Our approach is unsupervised and makes use of a fuzzy clustering algorithm based on the possibilistic framework, and includes a data-driven noise and artifact rejection phase. Our contribution provides a general analysis tool, applicable to any ERP data set, which can uncover the data set’s internal structure. The fuzzy clustering algorithm is the core of our method, since its fine-grained membership grades how much a sample belongs to a given cluster, making the method applicable even when groups have a certain overlap. Prior to the clustering step, we apply weights to the feature vectors, optimizing them in order to enhance the variance within the dataset, and we extract time-window interval based features inspired by interval arithmetic. We apply the data processing workflow to the analysis a set of ERPs recorded during an emotional Go/NoGo task. We evaluate the performance of the unsupervised analysis by computing a measure based on the clusterization rate of trials in different experimental conditions. The results on the studied data set show that the proposed method obtains a difference of clusterization rate of 69% in Go vs. NoGo trials, when weights and interval-features are applied to the data, improving previous work not including weights and interval-features which had a rate of 31%. Furthermore, when compared with the standard Fuzzy c-means, our proposed possibilistic clustering algorithm outperforms it in terms of clusterization rate. We also examine the effect of pre-processing the data with Independent Component Analysis and removing noise-related components, and observe that this does not improve significantly the obtained results. These findings demonstrate that our proposed method provides a valuable data processing workflow robust to EEG artifacts and able to produce a clustering that is coherent with the experimental conditions represented in the ERP dataset.

**Index Terms**—EEG, ERP, possibilistic clustering, unsupervised, interval features

## I. INTRODUCTION

THE BRAIN’S electrical activity is related to motor and cognitive tasks in complex ways. Shedding light on the causal links between tasks and observed regularities in brain activity can obviously benefit research in psychology

Manuscript received .... This work was partially supported by the Swiss National Science Foundation, under grant CR1311-138032. (*Corresponding author: Paolo Masulli.*)

P. Masulli is with the Department of Applied Mathematics and Computer Science DTU Compute, Section for Cognitive Systems, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark, and also with the NeuroHeuristic Research Group, University of Lausanne, Switzerland (e-mail: pamas@dtu.dk).

F. Masulli and S. Rovetta are with the Department of Computer Science, Bioengineering, Robotics and Systems Engineering DIBRIS, University of Genoa, 16146 Genoa, Italy.

A. Lintas and A.E.P. Villa are with the NeuroHeuristic Research Group, University of Lausanne, 1015 Lausanne, Switzerland.

and related disciplines. Direct brain interfaces that use brain activity as a control signal are also starting to spread in several fields, including virtual reality, advanced prosthetics, and rehabilitation. Their efficacy is crucially dependent on the ability to analyze electrical signals from the brain. This work presents a contribution to the task of discovering such links.

One of the most widely applied experimental methods for investigating human brain activity is scalp electroencephalography (EEG). This technique is not invasive, because it uses external electrodes which are placed at standard locations on the scalp. The exposition to specific stimulations (e.g., visual, auditory) causes brain activation in the form of transient electric potentials which can be measured with scalp electrodes and averaged over repeated trials [1]. The resulting EEG signals that are triggered by a specific event are called Event-Related Potentials (ERPs).

The analysis of EEG data can be used to infer behavioural and personality data. As the EEG signal is typically high-dimensional and noisy, the conventional analysis techniques for EEG data rely significantly on the experimenter’s ability to discard outliers and observe peaks and waves that are associated with different neural processes. The usual approach to stimulus-related EEG recordings is the calculation of Event-Related Potentials (ERPs), i.e. the averaging of multiple repetitions of a trial in the same experimental conditions. In fact, the assumption for ERP analysis is that the brain always reacts with the same activation as a response to a given stimulus, but the signal recorded with EEG also includes many other components, given by concurrent brain processes, noise, muscular artifacts, etc. Averaging the repeated trials allows to make the activation (the evoked potential) emerge and become observable. Neuroscientists have discovered that different stimuli cause specific waves or components in the evoked potentials, and therefore ERPs can be used to understand the brain processes that are triggered by a certain stimulus [2]. On the other hand, the averaging of many trials can cause a significant loss of information, as the noise (e.g. muscular artifacts caused by eye blinks or movements) can have an amplitude several order of magnitudes larger than the signal. To address this, approaches based on single-trial analyses [3] have been used, with the goal of including the contribution of the entire EEG recording [4].

In this study, we define an unsupervised processing workflow based on fuzzy clustering, in order to model differences and similarities between samples in a dataset, using a data-driven feature-extraction strategy. Our goal is to build a general-purpose unsupervised analysis workflow for Event-

Related Potentials. Most conventional EEG analysis pipelines have a number of processing steps requiring expert evaluation of the results. For example, the widely used Independent Component Analysis relies on an expert inspection of the components identified by the algorithm in order to decide whether they are signal or artifact. We propose an alternative approach, which aims to extract exploratory information from an EEG (ERP) dataset, minimizing human expert intervention. Our method does not need a specialist to evaluate the artifact removal, since it can extract information despite the presence of noise. The clusters that are obtained with our technique are a way for the EEG researcher to obtain a partition of the given ERP set into clusters sharing similar signal features, without having to impose any classification goal or having to perform specific preprocessing. The advantage of this approach is that it can help the research to uncover structure of the ERP set which was not known, possible groupings that were not previously studied or observed. To avoid the limitations of ERPs described above, we consider the original EEG signal (the whole set of trials) as the starting point of our analysis pipeline and implement a noise-rejection method to remove outlier epochs in our data processing workflow. Our method is not an alternative to (semi-)supervised approaches, but rather constitutes a possible first level of analysis, which can be followed by the application of other methods targeted, for instance, at classification. The results of the clustering can inform the algorithm design. To avoid the limitations of ERPs described above, we consider the original EEG signal (the whole set of trials) as the starting point of our analysis pipeline and implement a noise-rejection method to remove outlying artifacts in our data processing workflow.

In cluster analysis, the choice of a suitable number of clusters is one of the most important questions. Different methods for determining the optimal number have been proposed, included randomized algorithms based on Chebychev polynomial approximation [5]. In our approach, we choose a number of clusters that is 2 to 3 times the expected number and then merge overlapping clusters. The merging procedure relies on measuring overlap with a fuzzy version of the Jaccard index. This lets our method determine a suitable number of clusters with a data-driven approach.

Previous work on this framework was presented in [6] as a conference paper. Here we present a significantly updated treatment which includes a number of new methods. In particular:

- we apply weights to the sample vectors and we optimize those weights in order to maximize the variability of the features within the dataset,
- we define interval-valued features computed on time windows of different lengths in order to improve the comparison of signals.

The data processing workflow is outlined in Fig. 1. The remainder of this article is structured as follows: Section II presents the state of the art and Sections III to V introduce the theoretical tools that we are using: Graded Possibilistic Clustering, the Jaccard Index and Weighted Interval-valued Features. The dataset that we consider as a case study is

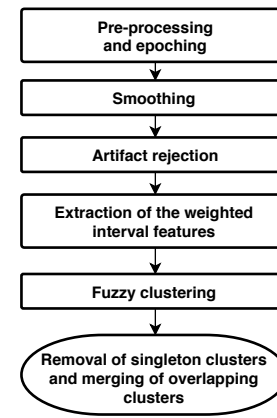


Fig. 1. Data processing workflow.

presented in Section VI. The way the data processing workflow is applied to the example dataset is the subject of Section VII. The remaining Sections VIII–IX present the results, discussions and conclusions.

It should be noted that in this article we apply the proposed method to a specific set of ERPs (from an experimental paradigm based on an emotional Go/NoGo task), but the method can be applied without significant changes to the analysis of any ERP-dataset.

## II. STATE OF THE ART

Electroencephalography is the most widely used method to record human brain activity with high temporal resolution in a non-invasive way in clinical and research contexts, and recently in consumer-grade wearable devices. Despite being inexpensive and easy to use, EEG is characterized by a high noise-to-signal ratio, caused by artifacts of different origin (e.g. muscular, electromagnetic, or caused by the poor adherence of the electrodes on the scalp giving too high impedance). Most EEG studies include in their method a manual phase of artifact detection and rejection of the samples that are too noisy: in case of ERP studies, the signals obtained in multiple trials are averaged, and the resulting waves are examined in order to identify peaks associated to specific brain processes. The cognitive processes in the brain are caused by the coordinated activation of an extremely high number of interconnected neuron, whose synchronous combined electrical fields are strong enough to be measured on the scalp. Therefore, ERPs can be seen as the brain response to a specific stimulus (e.g. visual or auditory). An ERP waveform's duration is generally not greater than one second from the stimulus onset, and along its time course one can observe positive and negative peaks or deflections, which can be considered as the combination of a number of components, positive or negative, each with a specific onset. Neuroscientists have discovered a number of components that are associated to particular classes of stimuli, and such components are characterized by a specified latency, shape, and positive or negative magnitude [2].

Most classical ERP results are obtained via supervised methods, and rely heavily on the experimenter's ability for artifact removal and for the interpretation of the wave components

associated with the studied neural processes. Machine learning techniques have been recently applied to these problems [7]. This includes methods such as Echo State Networks [8], multilayer perceptron neural networks [9], feature-extraction based on time intervals [10], or dimension reducing methods such as Independent Component Analysis [11]–[13]. It should be noted that such approaches can be used to address specific classification problems and have the disadvantage of being dependent of the choice of a set of features, designed by the experimenter in order to capture a certain aspect of the data. As a consequence, the human supervision is moved to a different step of the analysis, opening the door to biased learning and over-fitting of the model on the training data.

### III. GRADED POSSIBILISTIC CLUSTERING

To obtain a reliable clustering of EEG signals, which are noisy and highly variable, we employ a fuzzy clustering approach; in particular we focus on a robust clustering method that is rooted in the Possibilistic approach of Krishnapuram and Keller [14], [15]. The *Graded Possibilistic Clustering* method [16] is a central clustering model which inherits outlier rejection properties from the possibilistic model, while at the same time avoiding their well-known issues related to convergence and to overlapping clusters, similarly to methods which include probabilistic constraint (as in Fuzzy c-means [17], [18], and Deterministic Annealing [19], [20]). In the current work, we introduce a new version of the algorithm, called Graded Possibilistic Clustering model (GPC-II), which is an improvement on the one presented in [16].

Let  $X$  be a set of  $n$  observations (or instances)  $\mathbf{x}_l$ , for  $l \in \{1, \dots, n\}$ , and let  $C$  be a set of  $c$  fuzzy clusters denoted  $C_1, \dots, C_c$ . The clusters are represented via their centroids  $\mathbf{y}_j$ , for  $j \in \{1, \dots, c\}$ . We associate to each cluster a fuzzy cluster indicator (or membership) function  $u_{lj} \in [0, 1] \subset \mathbb{R}$  to a given observation  $\mathbf{x}_l$ .

We start by defining the *free membership* of an observation  $\mathbf{x}_l$  to cluster number  $j$  as  $v_{lj} = e^{-d_{lj}/\beta_j}$ , where  $d_{lj}$  denotes the distance between the  $j$ -th centroid and the observation  $\mathbf{x}_l$ . The coefficient  $\beta_j$  is a cluster-specific parameter of the model that plays a role in how data is represented by influencing cluster width.

The *total membership mass* for the same observation is then defined as:

$$\zeta_l = \sum_{j=1}^c v_{lj}. \quad (1)$$

The actual membership of observation  $\mathbf{x}_l$  to cluster  $j$  is written as a function of the above quantities as:

$$u_{lj} = \frac{v_{lj}}{\zeta_l} \quad (2)$$

In this equation we employ as a normalizer the *generalized partition function*:

$$Z_l = \zeta_l^\alpha = \left( \sum_{j=1}^c v_{lj} \right)^\alpha, \quad \alpha \in [0, 1] \subset \mathbb{R} \quad (3)$$

In the expression for the generalized partition function,  $\alpha$  is a parameter which controls the “possibility level” of the GPC-II

model. The value  $\alpha = 1$  corresponds to a totally probabilistic model, while  $\alpha = 0$  results in a totally possibilistic one, without normalization. The intermediate values  $0 < \alpha < 1$  allow for the grading of this model.

The following equation represents the relation between the GPC-II cluster centroids and the membership vectors:

$$\mathbf{y}_j = \frac{\sum_{l=1}^n u_{lj} \mathbf{x}_l}{\sum_{l=1}^n u_{lj}}. \quad (4)$$

The GPC-II model is trained by a Picard iteration of Eqs 2 and 4 after an initial step of random initialization of centroids.

It should be noted that, for  $\alpha = 1$ , the model has the same representation properties as Deterministic Annealing [19], [20], where the solution of a regularized minimization problem with constraints yields

$$u_{lj} = \frac{e^{-d_{lj}/\beta_j}}{\sum_{h=1}^c e^{-d_{lh}/\beta_h}}, \quad (5)$$

equivalent to (2) when  $\alpha = 1$ . In turn, this is equivalent to Fuzzy c-means, up to a non-linear transformation, as proved in [21]. When  $\alpha = 0$ , they are equivalent to those of Possibilistic c-means [15]) which is designed for robust clustering, since its clusters are not influenced by each other or by outliers. This comes at the cost of difficult convergence and collapsing clusters. The most interesting behavior is therefore obtained for  $0 < \alpha < 1$ . Whenever  $\alpha$  is positive, there is a degree of competition between clusters which improves convergence, as in probabilistic models. On the other hand, memberships eventually vanish for points sufficiently far away from the centroids, as in the possibilistic case, for noise insensitivity. This trade off is under user control, and depends in a complex way on the nature of the data (dimensionality, density, clustering structure, distribution isotropy, distribution uniformity...), so it is usually selected by grid search.

Deterministic annealing can be implemented in the Graded Possibilistic Clustering model obtaining a variation named DAGPC-II. To do so, the model parameters  $\beta_j$  are decomposed as  $\beta_j = \beta b_j$ , where the new parameter  $\beta$  is the basis for the deterministic annealing procedure (small at the start of the algorithm and increasing each time the Picard iteration converges) and  $b_j$  are the *relative* cluster scales. The values for the parameters  $b_j$  can be obtained with heuristics such as the ones proposed in [15]. This makes it possible to for the Deterministic Annealing Graded Possibilistic Clustering method to benefit from the powerful optimization technique proposed in [19], [20]: after starting from clusters that fully overlap, it performs a hierarchical clustering by progressively splitting clusters as  $\beta$  increases. “Natural” aggregations are therefore discovered.

### IV. MEASURING OVERLAP BETWEEN FUZZY CLUSTERINGS

The procedure presented requires a measure of overlap between clusters. Many approaches are available in the literature; a concise survey and an experimental comparison is presented in [22].

Several interesting indices to compare two clusterings are centered around counts of concordance patterns on pairs of data points, where any given pair can be in the same cluster

in both clusterings, in the same cluster in only one of the two clusterings, or in different clusters in both clusterings. Methods for fuzzifying such indices are usually based on replacing counts with combinations of fuzzy cluster memberships [23]–[25].

In this study we adopt the Jaccard index, defined as follows:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (6)$$

This measure is less sensitive to irrelevant matches (pairs in different clusters in both clusterings) than, for instance, the Rand index [26].

In the case of fuzzy clusters, we can use the membership function to express the fact that an element belongs to a cluster, i.e. a real quantity in the interval  $[0, 1]$ . In order to define the Jaccard index (Eq. (6)) in the fuzzy case, we adopt the following definitions of cardinality for the intersection and the union of fuzzy clusters:

$$|C_l \cap C_m| = \sum_{j=1}^N \min(u_{jl}, u_{jm}), \quad (7)$$

$$|C_l \cup C_m| = \sum_{j=1}^N \max(u_{jl}, u_{jm}), \quad (8)$$

where  $N$  is the total number of samples. The resulting fuzzy Jaccard index is still a real value in the interval  $[0, 1]$  even for possibilistic partitions, provided that the total membership of any data point does not exceed 1. With the definitions provided (Eq. (1) to (3)), this is true in realistic situations.

## V. WEIGHTED INTERVAL-VALUED FEATURES

Our strategy for comparing ERP signals is based on the following considerations: Event-related peaks might occur at slightly different moment during the signal's time course, and not all instants of the signals are equally meaningful. Therefore we use a kernel smoothing procedure based on a weight-mask, and we extract temporal, interval-valued features that summarize the data while accounting for this source of uncertainty.

### A. Interval-valued features

Comparing ERPs and estimating their similarity can be challenging because of differences in the onset time of measured potentials. We address this issue by extracting interval-valued features based on time windows.

A similar approach was used in [10]. Here we employ a lower number of features which appears to be sufficient, in order to avoid model over-fitting.

Each signal is represented by means of multiple features, each of which summarizes a specific time window as an interval.

The time window features are extracted as follows: Along a signal vector  $\mathbf{x} \in \mathbb{R}^L$ , we select onset points  $a_j$  at regular intervals such that  $a_{j+1} - a_j = R$ . Starting from each onset point, we select offset points  $b_{jk}$  such that  $b_{jk} - a_j = 2^{k+1}$ , for all the possible values of  $k$ . This procedure identifies time

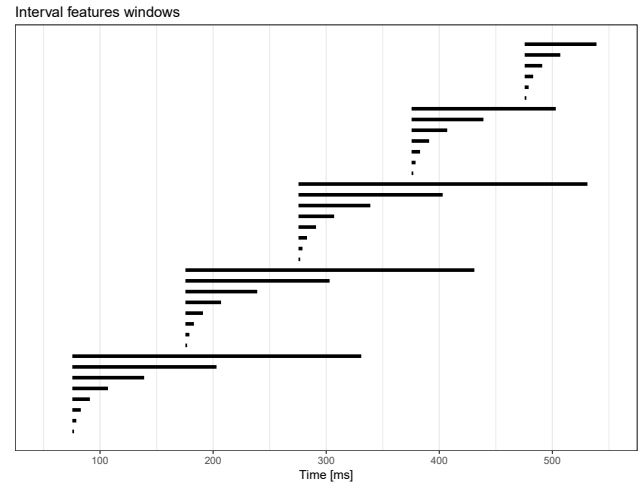


Fig. 2. Illustration of the interval feature windows. The segments represent the time intervals over which the interval-valued features [maximum, minimum] are calculated.

windows:  $I_{jk} = [a_j, b_{jk}]$  of length  $2^{k+1}$  and for each window we define the features:

$$M_{jk} = \max(x(i) \mid a_j \leq i \leq b_{jk}) \quad (9)$$

$$m_{jk} = \min(x(i) \mid a_j \leq i \leq b_{jk}), \quad (10)$$

namely the maximum and the minimum values of the signal vector in the time window. This way, the signal vector  $\mathbf{x}$  is represented by the vector formed by all the time window features  $[m_{jk}, M_{jk}]$ . An example of the obtained windows can be observed in Fig. 2.

### B. Parameter tuning

To take into account the relative importance of each feature, in comparing two signals they are multiplied by suitable weights. These weights are convex, i.e., their sum is constrained to one, and their value is selected according to the following procedure.

Since the number of parameters equals the number of features (minus 1 due to the convexity constraint), we also introduce a method to optimize their value. The resulting weights can be obtained offline.

Let  $\mathbf{x}_l \in \mathbb{R}^L$  for  $l \in \{1, \dots, N\}$  be the ERP samples and  $\mathbf{w} \in \mathbb{R}^L$  a vector of weights such that

$$\sum_{j=1}^L w_j = 1. \quad (11)$$

(convexity of weights). Then we can define a weighted similarity matrix  $K \in M_{N,N}(\mathbb{R})$  whose entry  $(i, j)$  is  $k(\mathbf{x}_i, \mathbf{x}_j, \mathbf{w})$  such that:

$$k(\mathbf{x}_i, \mathbf{x}_j, \mathbf{w}) = \exp \left( - \sum_{t=1}^L w_t (x_i(t) - x_j(t))^2 \right). \quad (12)$$

This similarity is naturally related to the Euclidean distance used in clustering.

We observe that

$$0 < k(\mathbf{x}_i, \mathbf{x}_j, \mathbf{w}) \leq 1 \quad \forall \mathbf{x}_i, \forall \mathbf{x}_j, \forall \mathbf{w} \quad (13)$$

**Given** a starting point  $\mathbf{w} \in \mathbb{R}^L$   
**repeat**  
 1.  $\Delta \mathbf{w} := -\nabla f(\mathbf{w})$ .  
 2. *Update.*  $\mathbf{w} := \mathbf{w} + \eta \Delta \mathbf{w}$   
 3. *Normalize:*  $\mathbf{w} := \mathbf{w}/|\mathbf{w}|_1$   
**until** stopping criterion is satisfied.

Fig. 3. Gradient descent algorithm

(normal similarity). Furthermore  $K$  is a symmetric matrix and all its diagonal entries are equal to 1. This implies that  $K$  has all real non-negative eigenvalues. Let  $\lambda_i$  be the  $i$ -th eigenvalue if they are sorted in descending order. Then the trace of the matrix satisfies the relation:

$$N = \text{tr}(K) = \sum_{i=1}^N \lambda_i. \quad (14)$$

This, along with normality of the similarity measure as per Eq. (13), implies that the largest eigenvalue  $\lambda_1$  or *spectral radius* satisfies  $1 \leq \lambda_1 \leq N$ .

As known in the field of spectral clustering [27], the rank of a normal similarity matrix approximates the number of clusters in the data. Specifically, the case  $\lambda_1 = 1$  corresponds to  $L$  singleton clusters, and the case  $\lambda_1 = N$  corresponds to one cluster including all the samples; the remaining values correspond to intermediate cases.

Therefore, by minimizing the spectral radius over the weight vector  $\mathbf{w}$ , we attain the solution with maximal diversity, i.e., with the highest separation of clusters. Note that convexity (11) prevents trivial solutions.

To minimize the spectral radius without computing the eigendecomposition, we minimize an upper bound. In general,  $\lambda_1$  satisfies the following:

$$\min(s_i) \leq \lambda_1 \leq \max(s_i), \quad (15)$$

where  $s_i$  is the sum of the  $i$ -th row of  $K$ . Therefore, to minimize the spectral radius we can minimize the objective function  $F(w) = \max_j(s_j)$ . To this aim we use an iterative gradient descent algorithm, illustrated in Fig. 3. In this procedure we use the following objective function gradient:

$$\begin{aligned} \frac{\partial}{\partial w_r} F(w) &= \sum_{i=1}^N \frac{\partial}{\partial w_r} k(\mathbf{x}_i, \mathbf{x}_j, \mathbf{w}) \\ &= \sum_{i=1}^N -k(\mathbf{x}_i, \mathbf{x}_j, \mathbf{w})(x_i(r) - x_j(r)), \\ &\text{for } j = \text{argmax}(s_i). \end{aligned} \quad (16)$$

As for all non-stochastic gradient descent heuristics, for a sufficiently small step size  $\eta$  this procedure will asymptotically converge to a local minimum and may benefit both from multiple restarts, for better quality minima, and from adaptive step size strategies, for faster convergence [28]. However, the problem was experimentally found to be simple, so neither of these improvements was needed.

## VI. EXPERIMENTAL DATA

### A. Participants and recording procedure

A total of nineteen volunteer participants (14 males, 5 females, mean age 28 (SD = 6.69)) participated in experimental study performed at the University of Lausanne (Switzerland). All had normal or corrected-to-normal vision, none reported a history of sustained head injury, and all were naive to the behavioral task. The participants provided written consent for their participation to the experiment approved by the local ethical committee, in line with the Declaration of Helsinki [29]. Continuous EEG was recorded using 60 scalp Ag/AgCl active electrodes (ActiveTwo MARK II Biosemi EEG System, BioSemi B.V., Amsterdam, The Netherlands) mounted on a headcap (NeuroSpec Quick Cap) and referenced to the linked earlobes. Electrophysiological signals were sampled at 2048 Hz with lower cutoff at 0.05 Hz and upper cut-off at 200 Hz, 24 bit resolution (DC amplifiers and software by Biosemi, USA). Electrode impedances were kept below 5 k $\Omega$  for all recordings. Vertical and horizontal ocular movements were also recorded using two pairs of bipolar electrodes.

### B. Behavioral task

The participants performed a behavioral task based on an emotional Go/NoGo paradigm [30]. The experimental procedure, including stimulus presentation and registration of the responses was implemented using the software the E-Prime software (Psychology Software Tools, Inc., Sharpsburg, PA 15215-2821, USA). This software was used to send triggers to the EEG acquisition system via a parallel port connection. The stimuli consisted in pictures showing a face (displayed at the center of the computer screen). The faces were either displaying an emotion (happiness, fear, anger, or sadness) or neutral. In the task studied here, the “Go”-cues were the neutral faces, for which the participants were instructed to respond as fast as possible by pressing a button, while the “NoGo”-cues were the emotional faces. In the trials with “NoGo”-cues, the participants were instructed to withhold response. The stimulus presentation lasted 500 ms per trial and was followed by a blank screen displaying only fixation cross (+) for a duration of 1000 ms, during which the possible button press was recorded. The experimental task was composed by four blocks, and each of them consisted of 30 “Go/NoGo” trials in pseudo-randomized order. In each block, 20 trials presented “Go”-cues and 10 “NoGo”-cues.

## VII. DATA PROCESSING WORKFLOW

For each participant, the data processing includes the following phases.

### A. Pre-processing

The EEG data files are imported in EEGLAB [31] and for each participant they are re-referenced with respect to the mastoidal electrodes M1 and M2. We derived two datasets for the subsequent analyses. The first dataset (*non-ICA dataset*) is formed by raw data files, as explained above. The second dataset (*ICA dataset*) is formed by raw data preprocessed with

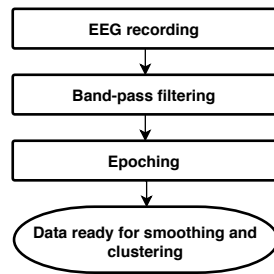


Fig. 4. EEG preprocessing workflow.

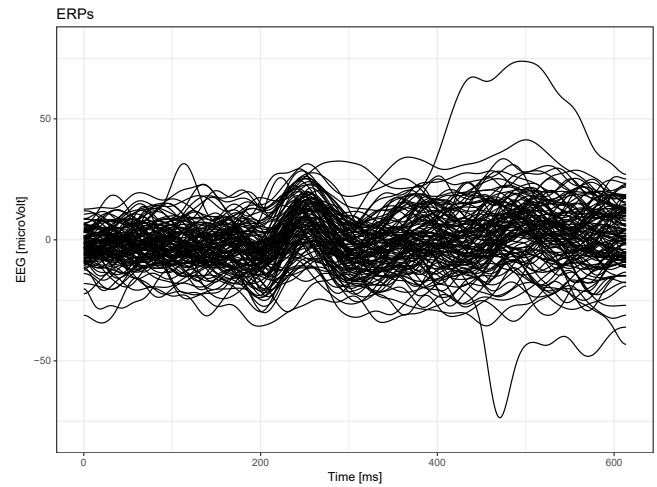
Independent Component Analysis (ICA) [2] for artifact correction: blink, saccade and eyelid artifact components are set to zero, based on their respective shape and topography [32]. The ICA dataset is also downsampled to 1024 Hz with anti-aliasing filter, prior to the ICA step. Then, for both datasets we apply an IIR Butterworth Band-Pass filter (called “Basic Filter for continuous EEG” in EEGLAB). The Half-Power threshold parameters are 0.1 Hz for High-Pass and 27.6 Hz for Low-Pass. Then the files are exported as EDF (European Data Format). The remainder of the data processing workflow is formulated for the non-ICA dataset.

### B. Segments extraction

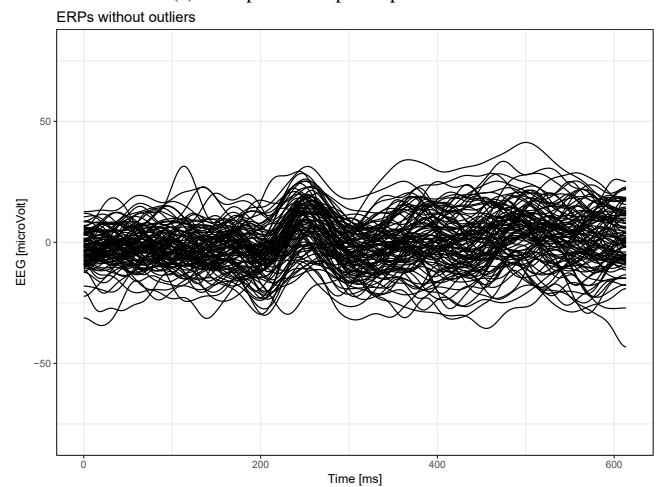
The filtered EDF data files are processed with a custom Python script. They are split into vectors corresponding to trials, using the triggers present in the EDF files. We exclude the trials in which the participant gave a wrong response (press the button for a “NoGo” cue, or not press it for a “Go” cue), which leave, on average, 108 (out of 120) trials per participant. We consider a time course of 600 ms for each vector, starting at the beginning of the trial (the instant when the visual stimulus is presented). This short duration is motivated by necessity to avoid muscular artifacts produced by the participant moving to press the button. For each signal vector, we apply a baseline correction by subtracting the average amplitude of the signal in the 200 ms preceding the trial onset. Since the sample rate is 2048 Hz, a duration of 600 ms corresponds to vectors in  $\mathbb{R}^{1228}$ , which are the input to our analysis. The phases of the EEG pre-processing and epoching are displayed in Fig. 4.

### C. Smoothing

In order to remove low-amplitude noise and artifact, we regularize the signals with the application of the 1-dimensional Anisotropic Diffusion filter by Perona & Malik [33]. The algorithm is applied to each signal vector, with 1000 iterations. The Anisotropic Diffusion parameter  $\kappa$  is set to 30: this parameter controls conduction as a function of gradient, and was chosen heuristically within the recommended range in order to allow smoothing but still preserve deflections and peaks in the signals. The parameter  $\Delta_t$ , which controls the speed of diffusion, was set to 0.33. This is recommended by the authors of the used implementation of the algorithm, in order to preserve numerical stability [34].



(a) All epochs for participant #12721



(b) Valid epochs after the removal of 11 trials

Fig. 5. Rejection of signal vectors tagged as artifact based on peak-to-peak amplitude.

### D. Artifact rejection

A simple artifact rejection method based on peak-to-peak amplitude is applied. The peak-to-peak amplitude is computed for each signal vector, and then we compute its median value and the median absolute deviation (MAD) across all vectors. A signal vector is rejected if the difference between its peak-to-peak amplitude and the median exceeds twice the MAD. This procedure rejects on average 14 signal vectors per participant. The remaining vectors for each participant are considered in the remaining part of the analysis. We denote the set of valid data as  $X \subset \mathbb{R}^{1228}$ . Fig. 5 shows an example of the application of artifact rejection.

### E. Extraction of the weighted interval-valued features

We compute a weight vector for each participant according to the procedure described in Section V-B. The iterative gradient descent is run for 700 steps with constant step size  $\eta = 10^{-6}$ , leading to convergence of the objective function. The resulting mean weight vector is displayed in Fig. 6. As the weights attain their largest values (corresponding to the largest variance among the signals) in the initial and final part

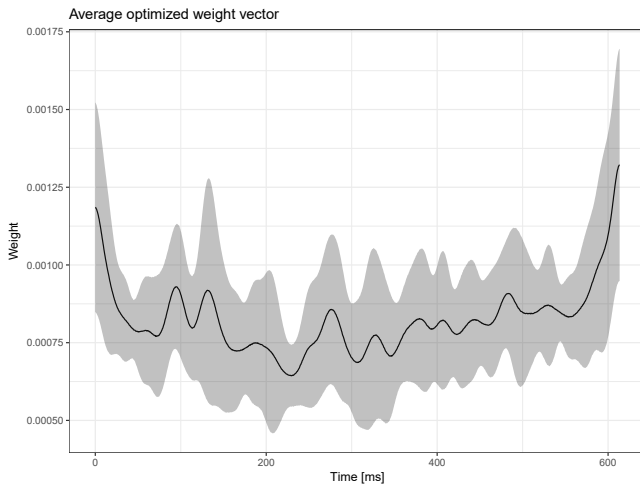


Fig. 6. The mean weight vector obtained by averaging the 19 participant weight vectors (the shaded area is the instantaneous standard deviation). The weights are computed via the algorithm described in Section V-B separately for each participant, over the entire length of the signal vectors. Here we display the average of the weight vectors for all participants in a compact form.

of the epochs' duration, which are not believed to convey meaningful information in the evoked potential, we exclude the first and last 75ms of each epoch from the remaining part of the analysis.

The participant signal vectors are then pointwise multiplied with the obtained weight vector in order to amplify the differences among them. They are furthermore multiplied by a fixed constant of 1000, in order to avoid numerical errors caused by approximations in the remainder of the data processing.

Then interval-valued features are extracted as described in Section V-A, with a step length  $r$  corresponding to 200ms. This leads to selecting 74 time windows for each signal vector (compared to the original length of the signal vectors that is 1228).

#### F. Possibilistic clustering

The Graded Possibilistic Clustering algorithm with deterministic annealing DAGPC-II (Section III) is applied to the data. We set the parameter  $\alpha$  to 0.85 and the initial number of clusters to  $c_0 = 7$ . This number is obtained by multiplying the expected number of cluster by a factor between 2 and 3. The clustering algorithm outputs, for each cluster  $C_j$ , its centroid  $y_j \in X$  and its membership vector  $u_{.j} \in \mathbb{R}^{|X|}$ . Its components express the membership of each sample to the cluster.

#### G. Singleton clusters rejection

Singleton clusters are those modeling only a single sample. We identify them by selecting the clusters whose total membership differs from the average membership across all the clusters by more than 1.5 standard deviations. If there is only one signal vector whose maximum cluster membership is in correspondence of the singleton cluster, then the vector is removed and the clustering algorithm of VII-F is applied again, reducing the number of required clusters by one.

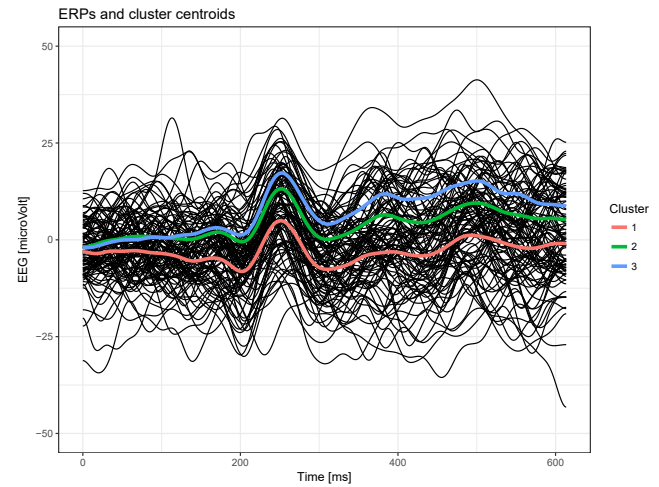


Fig. 7. ERPs for participant #12721 (thin black lines) and the three identified clusters (represented by their centroids with thicker colored lines).

The procedure is repeated until there are no more singleton clusters.

#### H. Merging of overlapping clusters

We apply the fuzzy Jaccard index, defined in Section IV, to determine the closeness of each pair of clusters. For each participant we compute (independently) the distribution of the Jaccard indices for each pair of clusters. This will be in general a multimodal distribution that can be seen as a mixture of normal distributions. The rightmost mode includes pairs of “neighbor” clusters, which are very similar to each other, since their mutual Jaccard index is close to one. The exploration of the dataset made us conclude that a cut-off point  $J_0 = 0.7$  can be heuristically chosen as a threshold above which clusters have a large overlap and can be merged. This threshold value is applied for each participant: two clusters are merged if their Jaccard index is above  $J_0$ . To merge, we compute a new centroid as the average of the two previous clusters' centroids weighted by the membership values of their elements, and then new memberships are computed with respect to the new centroid in the same fashion as done in the clustering step of Section VII-F. The procedure continues iteratively recomputing the Jaccard indices at each step, until no more neighboring clusters can be merged.

## VIII. RESULTS

The analyses and results presented in this study are limited to the signal recorded at the electrode site ‘Fz’. While the full dataset is not publicly accessible, since this was not part of the study protocol, we have included a summary of the data as Supplemental Material. For each participant, the described analysis workflow gave a number of clusters between 2 and 5. An example is displayed in Fig. 7.

The clusters could not definitely be associated to different experimental conditions (Go/NoGo condition, type of emotion). This suggests that high cognitive functions such as affective discrimination are not associated with simple wave



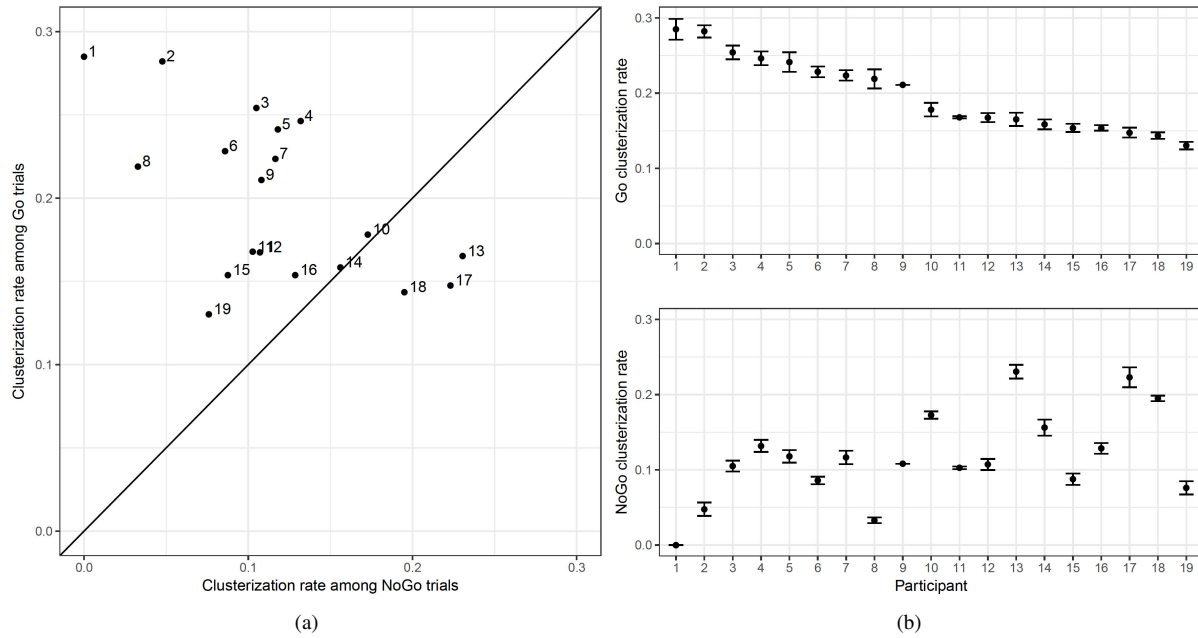


Fig. 8. Comparison of the clusterization rates of “Go” and “NoGo” trials for all participants (after the application of weight optimization procedure and the calculation of the interval-values features). (a) Scatter plot of the average values over 20 different random seed initializations (each point corresponds to one participant). For 16 participants (out of 19), the clusterization rate of the “Go” trials is higher than the one of the “NoGo” trials. This reflects the fact that the signal vectors associated to “Go” trials tend to be more similar to each other, while the “NoGo” trials are characterized by higher variability which escapes the clusters. (b) Average value and standard error of the clusterization rate of “Go” and “NoGo” trials for each participant over the same 20 different random seed initializations.

components of the brain signal and might instead require the choice of higher level features or specific indices, or the study of a larger number of EEG locations.

#### A. Clusterization rate

In order to assess the validity and interest of the obtained clusters, we measure the correspondence between the result of the analysis and the “Go” and “NoGo” experimental conditions. This is done by defining, for each participant, a measure called *clusterization rate*. The clusterization rate of the “Go” trials (respectively, the “NoGo”) trials is defined as the fraction of “Go” (respectively, “NoGo”) trials characterized by a high membership to one of the identified clusters. This is defined as the number of samples with a membership above the 95-th percentile of the cluster membership distribution for one of the clusters, divided by the total number of “Go” trials for the participant (respectively, “NoGo”). For most participants, the clusterization rate was higher in “Go” trials than in “NoGo” trials (Fig. 8).

Since the results of the data processing workflow are dependent on the random seed initialization for the clustering step, we have repeated the same analysis for 20 different values of the random seed, in order to evaluate the robustness of the result.

#### B. Weight optimizations and interval features

In order to assess the contribution to the result given by the weight optimization procedure (Section V-B) and the interval-valued features selection (Section V-A), we have performed

the same analysis switching off one or both optimizations. The results for DAGPC-II are summarized in Table I. The mean differences in each row of the table are significant (Welch Two Sample t-test, all  $p$ -values  $< 10^{-3}$ ). We can observe that, even though all four versions of the analysis gave a significant difference in clusterization rate between “Go” and “NoGo” trials, the use of both the weight optimization procedure and the interval feature selection gave the best result in terms of relative difference in clusterization rate between the two classes (69%). In this version of the analysis, we observed that the average clusterization rates for the “Go” trials were larger than the corresponding clusterization rates for the “NoGo” trials for 16 participants (Fig. 8a).

It can also be noticed that the internal variability of the clusterization rate values for the “Go” trials was smaller than for the “NoGo” trials (Fig. 8b). This is reasonable, since the “Go” trials all had neutral facial expression as a cue. This can explain more similar ERPs, which in turn would be associated with a high membership to one of the identified clusters. Conversely, the “NoGo” trials corresponded to stimuli presenting different facial expressions and therefore tended to be more different. This is also compatible with the fact that, if we inspect the data, we generally observe larger inner variability among the “NoGo” trials compared to the Go trials (Fig. 9).

#### C. Comparison with different algorithms

For comparison purposes, we have run numerical experiment replacing our proposed clustering algorithm DAGPC-II (Section III) with the Fuzzy c-means (FCM) algorithm [35].

TABLE I

**DAGPC-II ALGORITHM.** COMPARISON OF CLUSTERIZATION RATES. EACH ROW PRESENTS THE AVERAGE CLUSTERIZATION RATES FOR THE “Go” AND “NoGo” TRIALS IN THE SPECIFIED CONDITION FOR THE NON-ICA AND THE ICA DATASETS

Method		Non-ICA dataset			ICA dataset		
		Clusterization rate		# participants with	Clusterization rate		# participants with
Weights	Interval features	Go trials	NoGo trials	Go > NoGo	Go trials	NoGo trials	Go > NoGo
Yes	Yes	0.198	0.117	16 out of 19	0.229	0.187	14 out of 19
Yes	No	0.183	0.132	13 out of 19	0.213	0.179	14 out of 19
No	Yes	0.189	0.119	16 out of 19	0.169	0.156	10 out of 19
No	No	0.189	0.144	13 out of 19	0.173	0.141	10 out of 19

TABLE II

**FUZZY C-MEANS ALGORITHM.** COMPARISON OF CLUSTERIZATION RATES. EACH ROW PRESENTS THE AVERAGE CLUSTERIZATION RATES FOR THE “Go” AND “NoGo” TRIALS IN THE SPECIFIED CONDITION FOR THE NON-ICA AND THE ICA DATASETS

Method		Non-ICA dataset			ICA dataset		
		Clusterization rate		# participants with	Clusterization rate		# participants with
Weights	Interval features	Go trials	NoGo trials	Go > NoGo	Go trials	NoGo trials	Go > NoGo
Yes	Yes	0.286	0.154	15 out of 19	0.319	0.300	10 out of 19
Yes	No	0.212	0.135	13 out of 19	0.309	0.235	13 out of 19
No	Yes	0.269	0.163	15 out of 19	0.313	0.314	<i>ns</i> 9 out of 19
No	No	0.211	0.156	15 out of 19	0.294	0.241	11 out of 19

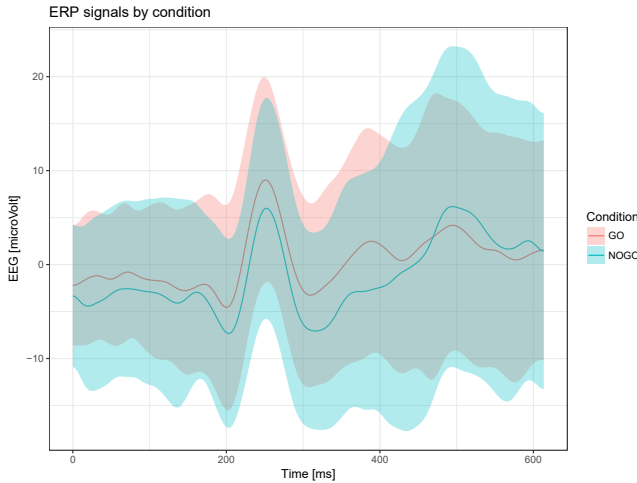


Fig. 9. Mean ERP signals for participant #12721 grouped in Go and NoGo trials. The shaded areas represent the instantaneous standard deviation. It can be observed that the variability is larger among the NoGo trial, compared to the Go trials. This is compatible with a higher clusterization rate for the Go trials.

As before, the analysis was run 20 times with different initialization of the random seed. The resulting clusterization rates are reported in Table II. The mean differences in each row of the table are significant (Welch Two Sample t-test, all  $p$ -values  $< 10^{-3}$ ), except when marked with *ns*. When running the analysis with FCM with weight optimization and interval features, the average clusterization rates for the “Go” trials were larger than the corresponding clusterization rates for the “NoGo” trials for 15 participants. This is a comparatively worse result with respect to that obtained with DAGPC-II, where the classification was correct for 16 participants.

#### D. Effect of Independent Component Analysis preprocessing

In order to assess the robustness of our proposed method to EEG artifacts, we have run the data processing workflow

both on the ICA-preprocessed dataset and the non ICA-preprocessed. The results are reported in Tables I and II for the two clustering algorithms. We notice how the ICA-preprocessing does not improve the difference of clusterization rate between the “Go” and “NoGo” trials.

## IX. DISCUSSION AND CONCLUSIONS

We presented a workflow for the unsupervised analysis of EEG signals using fuzzy clustering. It should be noted that the choices that we have made in the design are not bound to the specific dataset that was used as an example, and therefore make the data processing workflow, or parts thereof, applicable to any ERP dataset, requiring a minimal number of parameters to be decided. In comparison to Fuzzy c-means, our proposed algorithm DAGPC-II produced better results in most conditions, as a part of the data processing workflow in our experiments. However, the margin is thin, which is expected since the Graded Possibilistic algorithm is related to FCM, as pointed out in Section III. The advantage of DAGPC-II lies in fact in the grading, which makes it possible to change its level of “possibility”, by varying the parameter  $\alpha$ , and thereby changing the algorithm’s reliance on local information.

ICA is very often applied as a preprocessing step for EEG data and requires an expert to assess which components are artifact-related and should therefore be removed. Since our overall goal is to define a data processing workflow that minimizes human intervention, it is relevant to remark how the application of ICA does not improve our result in terms of difference in clusterization rate between the “Go” and “NoGo” trials. This suggests that our method is robust to EEG artifact and can indeed be applied even if the signals have not been preprocessed to manually remove noise components.

The only choice in the parameters that we believe to be data-dependent is the threshold value  $J_0$  that we used when merging similar clusters based on their fuzzy Jaccard index. We set this value  $J_0 = 0.7$  based on features specific to our

dataset derived from the distribution of its values across all samples.

Our method could be extended to the analysis of continuous EEG signals: since the analysis workflow requires signal vectors of the same length as input, the continuous EEG would have to be pre-processed with a windowing method, in order to obtain a number of signal vectors on which to apply the clustering algorithm.

Our previous work, in which the clustering algorithm compared the signals directly by their Euclidean distances [6], has now been surpassed by the use of optimized weights and interval features, which together contributed to a better result in the index we used to assess the results of the analysis. In fact, the direct comparison of signals as vectors ignores the temporal nature of the data and can lead to the loss of higher order features and inaccurate comparisons due to time shifts and delayed onset of EEG peaks. The optimized weights contribute to enhancing the differences along the signal vectors, thereby allowing the clustering algorithm to capture this variability. Using time window interval features makes the calculation less dependent on onset delays in the evoked responses.

Our work proposes an unsupervised approach to ERP analysis. This is not to mean that unsupervised methods are intrinsically better than supervised or semi-supervised ones, but their goals are different. In fact, we see our method as a preliminary and complementary to other (semi-)supervised approaches. We set out to define a data processing workflow that can be applied in generality to the exploratory analysis of ERP datasets in order to explore their internal structure captured by clustering. We imagine that an EEG researcher will be able to use our method as a first step in the study of an ERP dataset without imposing any constraints given by pre-determined labels. Subsequently they will use the results of the clustering to choose interesting classes of samples to classify with a (semi-)supervised approach.

Brain processes at all time scales are characterized by great complexity, and require analysis techniques able to capture it [36]–[39], such as methods aimed to identify specific features of the ERPs. In future work, we plan to incorporate indices targeted to the comparison of peaks and wave components, and able to measure time delays such as those used in [40]–[43].

In addition to suitable indices, data complexity and variety may also call for more powerful analysis tools. In this respect, kernel methods [44] may provide the necessary power and flexibility. Kernel methods for one-class classification have been developed by the authors [45] and may be employed in future research to deepen our understanding of the problem. Other approaches, like Infinite Kernel Learning [46], may offer better accuracy than single-kernel learning for heterogeneous data.

The possibility of capturing higher level features related to the brain processes guiding the task will make it possible to compare the EEG recordings of different participants in an unsupervised way, grouping the participants according to their mutual similarities. It will also be important to apply normalization rules to account for inter-participant amplitude

variability. As EEG hardware becomes more accessible, both in terms of cost and availability, its use becomes more common in many fields. Beyond neuroscience research, one central example is healthcare, where EEG offers an effective way to examine a patient's brain function and can be the basis for the definition of bio-markers with a high diagnostic relevance. Other future applications lie even in field of end-user products (e.g. personal monitoring of brain function for health and well-being, Brain-Computer Interfaces). EEG analyses based on computational intelligence methods are a valuable tool to address the huge quantities of data generated by EEG devices in any field and extract meaningful patterns from them. We believe that our work contributes to tackling this challenge.

## REFERENCES

- [1] T. W. Picton, S. Bentin, P. Berg, E. Donchin, S. A. Hillyard, R. Johnson, G. A. Miller, W. Ritter, D. S. Ruchkin, M. D. Rugg, and M. J. Taylor, "Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria," *Psychophysiology*, vol. 37, no. 2, pp. 127–152, Mar 2000.
- [2] S. J. Luck, *An introduction to the event-related potential technique*. MIT press, 2014.
- [3] E. Donchin, "Discriminant analysis in average evoked response studies: the study of single trial data," *Electroencephalogr. Clin. Neurophysiol.*, vol. 27, no. 3, pp. 311–4, Sep 1969.
- [4] G. A. Rousselet and C. R. Pernet, "Quantifying the Time Course of Visual Object Processing Using ERPs: It's Time to Up the Game," *Front. Psychol.*, vol. 2, p. 107, 2011.
- [5] R. Avros, O. Granichin, D. Shalymov, Z. Volkovich, and G. W. Weber, "Randomized Algorithm of Finding the True Number of Clusters Based on Chebychev Polynomial Approximation," in *Data Mining: Foundations and Intelligent Paradigms*, J. Kacprzyk, L. C. Jain, D. E. Holmes, and L. C. Jain, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, vol. 23, pp. 131–155.
- [6] P. Masulli, F. Masulli, S. Rovetta, A. Lintas, and A. E. P. Villa, "Unsupervised Analysis of Event-Related Potentials (ERPs) During an Emotional Go/NoGo Task," in *Fuzzy Logic and Soft Computing Applications. WILF 2016.*, Petrosino, Alfredo and Loia, Vincenzo and Pedrycz, Witold, Ed. Cham: Springer International Publishing, 2017, pp. 99–106.
- [7] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *Neuroimage*, vol. 56, no. 2, pp. 387–99, May 2011.
- [8] L. Bozhkov, P. Koprinkova-Hristova, and P. Georgieva, "Learning to decode human emotions with Echo State Networks," *Neural Networks*, vol. 78, pp. 112–119, Jun. 2016.
- [9] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13 475–13 481, Sep. 2011.
- [10] L. I. Kuncheva and J. J. Rodriguez, "Interval feature extraction for classification of event-related potentials (ERP) in EEG data analysis," *Lect. Notes. Artif. Int.*, vol. 2, no. 1, pp. 65–72, 2013.
- [11] A. Mueller, G. Candrian, J. D. Kropotov, V. A. Ponomarev, and G.-M. Baschera, "Classification of ADHD patients on the basis of independent ERP components using a machine learning system," *Nonlinear biomedical physics*, vol. 4, no. Suppl 1, p. S1, 2010.
- [12] L. Albero, A. Kachenoura, P. Comon, A. Karfoul, F. Wendling, L. Senhadji, and I. Merlet, "ICA-Based EEG denoising: A comparative analysis of fifteen methods," *Bulletin of the Polish Academy of Sciences: Technical Sciences*, vol. 60, no. 3, pp. 407–418, Dec. 2012.
- [13] A. X. Stewart, A. Nuthmann, and G. Sanguinetti, "Single-trial classification of EEG in a visual object task using ICA and machine learning," *J. Neurosci. Methods*, vol. 228, pp. 1–14, May 2014.
- [14] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, 1993.
- [15] —, "The possibilistic c-means algorithm: insights and recommendations," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 3, pp. 385–393, 1996.
- [16] F. Masulli and S. Rovetta, "Soft transition from probabilistic to possibilistic fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 4, pp. 516–527, Aug. 2006.
- [17] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Boston, MA: Springer US, 1981.

- [18] J. C. Dunn, "Some Recent Investigations of a New Fuzzy Partitioning Algorithm and its Application to Pattern Classification Problems," *J. Cybernetics*, vol. 4, no. 2, pp. 1–15, Jan. 1974.
- [19] K. Rose, E. Gurewitz, and G. C. Fox, "Statistical mechanics and phase transitions in clustering," *Phys. Rev. Lett.*, vol. 65, no. 8, pp. 945–948, Aug. 1990.
- [20] K. Rose, E. Gurewitz, and G. Fox, "A deterministic annealing approach to clustering," *Pattern Recognit. Lett.*, vol. 11, no. 9, pp. 589–594, 1990.
- [21] S. Miyamoto and M. Mukaidono, "Fuzzy C-Means as a regularization and maximum entropy approach," in *Proceedings of the Seventh IFSA World Congress, Prague, 1997*, pp. 86–91.
- [22] S. Rovetta and F. Masulli, "An experimental validation of some indexes of fuzzy clustering similarity," in *Fuzzy Logic and Applications, 8th International Workshop, WILF 2009, Palermo, Italy, June 9–12, 2009, Proceedings*, ser. Lecture Notes in Computer Science, V. Di Gesù, S. K. Pal, and A. Petrosino, Eds., vol. 5571. Springer, 2009, pp. 132–139.
- [23] R. K. Brouwer, "Extending the Rand, adjusted Rand and Jaccard indices to fuzzy partitions," *J. Intell. Inf. Syst.*, vol. 32, no. 3, pp. 213–235, 2009.
- [24] R. J. G. B. Campello, "Generalized external indexes for comparing data partitions with overlapping categories," *Pattern Recogn. Lett.*, vol. 31, pp. 966–975, Jul. 2010.
- [25] S. Rovetta and F. Masulli, "Visual stability analysis for model selection in graded possibilistic clustering," *Inform. Sciences*, vol. 279, pp. 37–51, 2014.
- [26] W. Rand, "Objective criteria for the evaluation of clustering methods," *J. Am. Stat. Assoc.*, vol. 66, pp. 846–850, 1971.
- [27] S. Rovetta, F. Masulli, and A. Cabri, "Measuring clustering model complexity," in *Artificial Neural Networks and Machine Learning – ICANN 2017*, A. Lintas, S. Rovetta, P. F. Verschure, and A. E. Villa, Eds. Springer International Publishing, 2017, pp. 434–441.
- [28] R. Fletcher, *Practical methods of optimization*. John Wiley & Sons, Ltd, 2000, second edition.
- [29] World Medical Association., "World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects," *JAMA*, vol. 284, no. 23, pp. 3043–3045, Dec 2000.
- [30] T. A. Hare, N. Tottenham, M. C. Davidson, G. H. Glover, and B. J. Casey, "Contributions of amygdala and striatal activity in emotion regulation," *Biol. Psychiatry*, vol. 57, no. 6, pp. 624–632, Mar. 2005.
- [31] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
- [32] M. Plöchl, J. P. Ossandón, and P. König, "Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data," *Front. Hum. Neurosci.*, vol. 6, pp. 278–301, 2012.
- [33] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, 1990.
- [34] Anisotropic Diffusion Algorithm (Perona & Malik). [Online]. Available: <https://mathworks.com/matlabcentral/fileexchange/14995-anisotropic-diffusion-perona-malik>
- [35] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, Jan. 1984.
- [36] P. Indic, "Time scale dependence of human brain dynamics," *Int. J. Neurosci.*, vol. 99, no. 1-4, pp. 195–199, Aug 1999.
- [37] V. Del Prete, L. Martignon, and A. E. Villa, "Detection of syntonies between multiple spike trains using a coarse-grain binarization of spike count distributions," *Network*, vol. 15, no. 1, pp. 13–28, Feb 2004.
- [38] R. X. Smith, L. Yan, and D. J. J. Wang, "Multiple time scale complexity analysis of resting state FMRI," *Brain Imaging Behav.*, vol. 8, no. 2, pp. 284–91, Jun 2014.
- [39] A. Wohrer and C. K. Machens, "On the number of neurons and time scale of integration underlying the formation of percepts in the brain," *PLoS Comput. Biol.*, vol. 11, no. 3, p. e1004082, Mar 2015.
- [40] M. Ihrke, H. Schrobbsdorff, and J. M. Herrmann, "Recurrence-based estimation of time-distortion functions for ERP waveform reconstruction," *Int. J. Neural Syst.*, vol. 21, no. 1, pp. 65–78, Feb 2011.
- [41] D. Lederman and J. Tabrikian, "Classification of multichannel EEG patterns using parallel hidden Markov models," *Med. Biol. Eng. Comput.*, vol. 50, no. 4, pp. 319–328, Apr 2012.
- [42] N. Karamzadeh, A. Medvedev, A. Azari, A. Gandjbakhche, and L. Najafzadeh, "Capturing dynamic patterns of task-based functional connectivity with EEG," *Neuroimage*, vol. 66, pp. 311–317, Feb 2013.
- [43] A. Zoumpoulaki, A. Alsufyani, M. Filetti, M. Brammer, and H. Bowman, "Latency as a region contrast: Measuring ERP latency differences

with Dynamic Time Warping," *Psychophysiology*, vol. 52, no. 12, pp. 1559–1576, Dec 2015.

- [44] M. Filippone, F. Camastra, F. Masulli, and S. Rovetta, "A survey of kernel and spectral methods for clustering," *Pattern Recognit.*, vol. 41, no. 1, pp. 176–190, Jan. 2008.
- [45] M. Filippone, F. Masulli, and S. Rovetta, "Applying the Possibilistic c-Means Algorithm in Kernel-Induced Spaces," *IEEE Trans. Fuzzy Syst.*, vol. 18, no. 3, pp. 572–584, Jun. 2010.
- [46] S. Özögür-Akyüz and G.-W. Weber, "On numerical optimization theory of infinite kernel learning," *J. Global Optim.*, vol. 48, no. 2, pp. 215–239, Oct. 2010.



**Paolo Masulli** (M'19) received the Ph.D. degree in mathematics from Aarhus University, Denmark, in 2014. From 2014 to 2017 he was a Junior Lecturer at the University of Lausanne, Switzerland. He is currently a Postdoctoral Researcher at the Technical University of Denmark. His research interests include machine learning and clustering applications to biosignals and interactions of topology and dynamics in networks.



**Francesco Masulli** (M'91-SM'04) is the Chair of IEEE Italy Section Computational Intelligence Society Chapter, and Full Professor of Computer Science at the University of Genoa, Italy. He received the 2008 Pattern Recognition Society Award, was chair of several international conferences and schools, and authored of more than 250 scientific papers in Clustering, Machine Learning, Neural Networks, Fuzzy Systems and Bioinformatics.



**Stefano Rovetta** (M'99-SM'12) is Associate Professor of Computer Science at the University of Genova (Italy). He authored more than 170 scientific papers in machine learning, neural networks, clustering, fuzzy systems, and bioinformatics. He received the 2008 Pattern Recognition Society Award, and was the chair of international conferences. He is a member of the Italian Neural Network Society, the European Neural Network Society, and the European Society for Fuzzy Logic And Technology.



**Alessandra Lintas** received the M.S. degree in Chemical and Pharmaceutical Technology from the University of Sassari, Italy, in 2004, and the Ph.D. degree in Pharmacology of Addiction from the University of Cagliari, Italy, in 2007. She is currently a Principal Investigator at the NeuroHeuristic Research Group, University of Lausanne, Switzerland. Her main research activity is focused on studying the brain networks involved in decision-making processes in animal models and in humans with EEG, ERP and optical brain imaging.



**Alessandro E. P. Villa** received a M.S. degree in Computer Science from the Swiss Federal School of Technology (EPFL) and the Ph.D. from the Faculty of Science of the University of Lausanne (UNIL), Switzerland. He is Full Professor of Computer Science and Neuroeconomics at UNIL and was re-elected President of the European Neural Network Society for two terms. In 2015 he was elected full member of the Swiss Academy of Sciences, in the Engineering section (SATW).