

# A subject personalized frequency domain features selection approach using genetic algorithm to improve SVM classification in EEG N400 paradigm

Joucla C.<sup>1</sup>, Gabriel D.<sup>1,2</sup>, Grigoryeva L.<sup>3</sup>, Ortega J-P.<sup>4,5</sup>, Haffen E.<sup>1,2,6</sup>

<sup>1</sup> Neurosciences intégratives et cliniques EA 481, Université de Franche-Comté, F-25000 Besançon, France  
<sup>2</sup> INSERM CIC 1431, CHRU Besançon F-25000 Besançon, France  
<sup>3</sup> Department of Mathematics and Statistics, Universität Konstanz, Germany

<sup>4</sup> Centre National de la Recherche Scientifique (CNRS)  
<sup>5</sup> Faculty of Mathematics and Statistics, University of St. Gallen, Switzerland  
<sup>6</sup> Service de Psychiatrie de l'adulte, CHRU Besançon, F-25000 Besançon, France

Context

In recent years, progress in medicine allowed to keep alive an increasing number of patients with disorders of consciousness. The neuroimaging technique of electroencephalography (EEG) is widely exploited to investigate brain mechanisms involved in cognitive processes and *in extenso* to monitor consciousness activity.

These data are used for the supervised training and testing of a Support Vector Machine (SVM), but recent studies have evidenced serious difficulties in achieving good performances both in unresponsive patients and healthy control subjects.

In this study, we test a features extraction optimization method, assuming that only a method that shows a good accuracy in fully aware subjects is likely to perform well in the detection of covert awareness.

Purpose

The object of the present study is to address the following questions:

- 1- How to extract relevant features after EEG recording of brain mechanisms during N400 stimuli ?
- 2- How to manage inter-subjects variability and optimize features extraction for each one of them ?
- 3- What improvements of classification accuracy and pValue can be expected using this method ?

## Part 1 : N400 paradigm and features extraction by Fourier transform

N400 is an event-related potential that allows to appraise the semantic processing ability of the subject. It appears when a word of different meaning occurs in a sequence of words. We used 120 word quartets:

- 3 first words set a semantic context
- 4<sup>th</sup> word is related or not to this context.

Bandpass filter: 1-30 Hz  
Baseline correction: 100 ms pre-stimulus  
Interpolation of electrodes: Voltage >100 µV

29 healthy subjects  
64 EEG channels  
1000 Hz sampling

EEG classification is highly sensitive to noise. Instead of processing raw data, leading to low classification accuracy, we extract relevant signal features on a period of interest. We select the time period ranging from 250 to 550 ms post-stimulus.

The most interesting information stands in the frequency content. We compute a Fast Fourier Transform of the signal multiplied by a sliding Hamming window function to retrieve both spectral components and their temporal location.

Considering the wide inter-subjects variations, we considered necessary to customize the period of interest, the sliding window width and the overlap length.

## Part 2 : Optimization of features extraction subject by subject

Creation of Initial Population (●●●●)x50

Genetic algorithms: natural genetics and evolution. Solving problems of global optimization by developing a population of candidate solutions.

Fitness evaluation

Signal transformation of each trial based on the 4 parameters given in each chromosome vector. The transformed trials are then classified using SVM.

Random permutation of trials Division into 4 equal groups: 1 group: testing set Others: training set Repeated 1000 times

Classification accuracy: (trials correctly classified / total number of trials)x100

Are terminus criteria met ?

YES: END The best set of parameters is kept

NO: Selection and reproduction

Offspring generation

Selection and reproduction

Constitution of the next vectors generation Selection probability based on performance.

3 types of children: Elite: highest accuracy, saved. Crossover: recombined. Mutation: random switch of a value in a chromosome

Terminus criteria: Average relative change in the best value over 50 generations is ≤ 1e-6 100% accuracy is reached

The entire process is repeated individually for each subject.

## Part 3 : Optimization results regarding accuracy and pValue

We plot as a histogram the percentage of subjects whose period of interest is included for a given time.

Red rectangle represents typical N400 window.

N400 window is relevant, since all subjects presented relevant features at that time.

We clearly see that inter-subject deviation makes the relevant features exceed N400 window, and should therefore be taken in account.

Plotting accuracy for each subject before (in red) and after (in green) optimization allows us to estimate global improvement about 20%.

pValues show that obtained accuracies are significant for all but one participant.

These results confirm that features extraction optimization permits classification improvement which allows for better brain activity discrimination in semantic cognitive processes.