Noname manuscript No.
(will be inserted by the editor)

On the EEG footprint of image saliency

Dario Izzo · Marek Ruciński · Christos Ampatzis · Eduardo Martin Moraud · Graham Healy · Peter Wilkins · Alan F. Smeaton · Ashkan Yazdani · Touradj Ebrahimi

Received: date / Accepted: date

Abstract This article investigates the possibility to classify images according to their subjective interest by looking at recorded EEG signals. We classify subjects' EEG recordings while they are shown images in a rapid serial visualization (RSVP) task. Our protocol is essentially derived from the well-known 'oddball paradigm'. Our exerimental methodology focusses on maximum usability of the device. We show promising results obtained with a simple 'wearable' 4-node setup; we evaluate the accuracy that can be obtained under such experimental conditions, and quantify the effect of image-display speed in this context.

Keywords Autonomous curiosity · EEG classification

1 Introduction

The concept of "science autonomy", defined as the capacity for a robotic agent to act compelled by some scientific motivation [24], has been discussed in the context of future technologies. Such an advance implies that features of some scientific value may be autonomously recognised and selected, just as scientists would choose them according to their expertise and interest. A possible approach is to first define characteristics that are scientifically relevant and then upload such definition into the agent so that it can autonomously detect features of interest. This approach, however, might prove unsuccessful to detect anything that falls out of the defined boundaries and that is still interesting.

The definition of what is to be considered as "scientifically relevant" (in images for example), is indeed a difficult one: it not only should encompasses scientific knowledge of *a-priori* defined features, but also "curiosity" toward what is unknown, novel or out of context. Curiosity itself is an abstract concept for which no easy definition is possible. It has been suggested for instance to be linked with surprise, creativity or beauty [20]. Examples of computational approaches to replicate curiosity have addressed the problem from the perspective of Bayesian reasoning – based on the discrepancy between prior and posterior beliefs [13, 25] – or reward-based motivation and learning [4, 21, 16]. Classical machine learning approaches need to be careful however not to confuse "curious" features with mere data outliers.

In this paper, we propose to use the classification of EEG readings recorded during a rapid serial visualization experiments to decide upon the "interest" of images. It is known from neuroscience experiments that the ability to detect and respond to novel events is inherent to human nature, and electrophysiological evidence shows that novelty and surprise triggers specific neural responses that may be recorded and further exploited [17]. In particular, the P300 brain-wave is an event-related potential (ERP) – a brain response recorded when participants exposed to different kinds of discrete stimuli – shown to be linked with switching of attention to deviant events [9]. It is usually recorded

Dario Izzo, Marek Ruciński, Christos Ampatzis and Eduardo Martin Moraud
European Space Research and Technology Center (ESTEC)
E-mail: dario.izzo@googlemail.com
Present address: Marek Ruciński
Centre for Robotics and Neuroscience, University of Plymouth

Graham Healy, Peter Wilkins and Alan F. Smeaton
CLARITY: Centre for Sensor Web Technologies, Dublin City University
E-mail: alan.smeaton@dcu.ie

Ashkan Yazdani and Touradj Ebrahimi
Multimedia Signal Processing Group, Institute of Electrical Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL)
E-mail: touradj.ebrahimi@epfl.ch

in the context of the so-called 'oddball paradigm', an experimental setup where a target stimulus is presented among more frequent background stimuli, and has been employed to assess aspects of human cognitive information processing [12], or in the framework of Brain-Computer Interfaces to recognise and automatically classify images which belong to different classes, e.g. letters of the alphabet [7] or faces, cars, animals, etc. [23].

Our ultimate goal would be to train a system able to replicate experts' scientific choices, i.e. an image classifier with broad – and fuzzy – classification borders that would return not only the strictly defined, but also images with unexpected, yet potentially relevant properties. Note that this implies that the system embodies the participants' *subjective* reactions to what they see, and may represent the first step to endowing agents with real human-like curiosity. Related work by [11] shares this goal and uses knowledge provided by scientists to create clusters of images according to their degree of scientific interest. Their protocol, however, implies asking experts to *manually* indicate their opinion on the images displayed. It is thus slow and not necessarily repetitive, making the procedure tedious and the results noisy and un-realiable. In comparison, this paper aims at proving that the same clusters can be built classifying brain recordings, thus allowing to speed up the process, while also exploiting implicit neural processing, thereby avoiding over-analysis. Central to our methodology is, thus, the wearability of the recording device and the need to reach high-speed image displays, even at the cost of accuracy loss. Note indeed that the final goal targeted here also differs from other applications of EEG reading classification, since it is tailored to automatically discard non-interesting images among huge amounts of data (i.e., to compress) rather than to communicate a message in which each single unit carries essential meaning: a unique error under such circumstances would be disastrous and might disturb the whole communication; false positives in our case however have less negative implications than in aforementioned systems, the essential point being that relevant images are classified correctly. Therefore a highly accurate but cumbersome system would not be appropriate for our purposes. The performance of our system must thus be understood as responding to the need of a global *compromise* between accuracy, speed and usability.

Our contribution is thus twofold. First, we measure the P300 ERP in the framework of a Rapid Serial Visual Presentation experiment (RSVP) using a simple 4-node electrode, easy to wear and simple to manipulate. We evaluate under these conditions the main parameters that affect our classification scheme, attempting to find a compromise between speed and accuracy in the framework of a two-class oddball paradigm. Secondly, we go beyond this paradigm and employ classes that are not clearly labelled, but which are instead 'fuzzy' in that they require scientific expertise to be understood (i.e. images might appear interesting to someone in the field, and banal to non-experts).

The remainder of the paper is organised as follows. Section 2 presents the background of using P300-based brain-machine interfaces to measure event-related potentials (ERPs). We then describe the experimental procedures in section 3 and the data analysis in section 5.

2 Background

The idea of using EEG recordings of event-related potentials for Brain-Computer Interfaces (BCI) applications has been in the literature for quite some time. A BCI enables a user to communicate with the world through her or his brain signals alone. The first P300-based system was originally described in [8], where the authors implemented a matrix of 6×6 cells containing the 26 letters of the alphabet. These were displayed on a screen and employed as a virtual keyboard in which the P300 was used as to capture the participants' choice of letter in, for example, a spelling task. Results showed that it served reliably as a communication channel up to 2.3 characters/min. Further improvements of such a system were undertaken by [7], reaching up to 7.8 characters a minute and 80% classification accuracy. Similar approaches have also been reported for both able-bodied adults and 'locked-in' patients (e.g., amyotrophic lateral sclerosis patients [22]) with similar degrees of success. The effects of parameters such as matrix size have also been analyzed [2].

Motivated by neurophysiological studies on the processing capabilities of the human visual system [26], recent techniques have attempted to apply this concept for classifying images using rapid serial visual presentation (RSVP) tasks [10,19]. Visual recognition in humans is indeed extremely fast (less than 150ms after the stimulus) and even for a brief presentation, individuals are able to extract a tremendous amount of information that enables a general characterization of the image content. It has been shown that the extraction of meaning under such conditions follows two stages: one early perceptual process, task independent at around 70ms after the stimulus, followed by a higher-level decision process that helps evaluate the relevance of the visual information in terms of goals and expectations [27]. BCI based on RSVP tasks thus not only help speed up the process, but they may also enable to exploit implicit recognition that avoids over-analysis of images which may result in a behavioral decision threshold higher than the one wanted in a classification task. Results for binary recognition tasks have been recently reported to reach 93% correct classification with 64 electrodes at speeds of up to 10 images/s [19]. Yet issues such as the impact of image saliency, image presentation rate, user expertise or target-to-target interval have been pointed out as central for a correct recognition, and as requiring further analysis.

Note that all the approaches mentioned employ complex EEG systems (32 or more electrodes), which make those experiments cumbersome. The recognition itself is restricted to images belonging to clearly defined classes, i.e. which are recognizable according to common features that may be a-priori listed (e.g., faces, cars). Here, we intend to widen the scope by displaying images which ellicit subjective repsonses according to the participants' own interest and knowledge. In addition, we show that promising results can be achieved with a very simple device (only 4-nodes), hence opening capabilities in a wide range of applications where the wearability of the device is to be taken into account.

3 Experiments

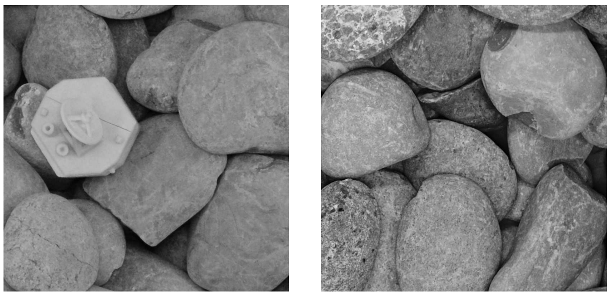
Each experiment described here was carried out independently by two groups located in different premises. The experimental environment was carefully replicated, and the ITU-R BT. 500-11 recommendation [3] was used as a baseline. To ensure equal experimental setups and precise timings during image projections, an image visualization software named Curiosity Cloning Viewer (*CCViewer)* was developed [18] and used throughout the project.[1]

The two setups differed solely in the EEG recording apparatus used. The first group operated at Dublin City University (DCU) and used two two-channel devices with a sampling rate of 254 samples per second and a 12-bit sampling resolution. Electrodes were placed at Pz, Cz, P3 and P4 according to the international 10-20 system. A joint mastoid reference was used between both of the EEG devices which when joined provided 4 sampling channels. A ground electrode was placed on the chin of the participant. In parallel, a second group operated at the Swiss Federal Institute of Technology (EPFL) in Lausanne, Switzerland, and recorded with a 36 channel device as a way to validate the results. The EEG signals were acquired at 2048 Hz and 24-bit sampling rate from 32 electrodes that were placed on the scalp of the subjects according to the 10-20 international electrode positioning system. A 'Biosemi Active Two' amplifier was used for amplification and analog to digital conversion of the recorded EEG signal.

In this paper the results reported on the classification performances originate from the low spatial, 'wearable' resolution device (4 nodes). The 32 node setups experiments have only been accounted to backup the general conclusions drawn.

3.1 Experiment 1 - Reliability vs. Speed

The aim of this first experiment is to analyze the reliability that can be expected with our simple experimental setup, and to further evaluate how this reliability is affected by the rate at which the images are presented, i.e. how fast the images can be shown to participants while still being able to classify their EEG readings. This proof of concept is important as a first step to get a clear idea of the compromise (in terms of reliability vs. speed) that can be reached under the present conditions when tested with a classical 'oddball paradigm' [12] task. Visual stimuli consisted of a subset of 3,204 images of grey stones illuminated with a uniform ambient light, among which 25 contained (in addition to the stones) a small model of a spacecraft, and thus constituted oddball images. The spacecraft position was different in each one of these images but the object itself was clearly visible in all cases. Examples of background and oddball images for this first phase experiments are given in figure 1.



(a) (b)

Fig. 1 Examples of (a) an oddball image (notice the small white model of a spacecraft) and (b) a non-oddball image used for Experiment 1

No. of subjects	*No. of seqs*	*Images in seq.*	*Oddballs in seq.*	*Reps*	*IDP/IIP (ms)*	*T (s)*
4	5	40	4	2	500/500	40
4	5	67	7	2	300/300	40
4	5	133	13	2	150/150	40
4	5	200	20	2	100/100	40
4	5	400	40	2	50/50	40

Table 1 Parameters of the 'Reliability vs. Speed' experiment

[1] The software has been released under BSD license and can be downloaded from `sourceforge.net/projects/ccviewer/`

The experiment involved the presentation of different sequences of images at varying speed rates. The subjects were initially familiarized with several examples of oddball and non-oddball images. After that, for each trial, they were instructed to count the number of images containing the spacecraft model while the sequences of images were presented to them and their EEG signals recorded. Each sequence was always preceded by a countdown screen of 5 seconds duration that allowed the participants to prepare for the experiment, hence reducing the surprise effect at the start. The experiment involved 4 subjects, 5 different image rates and 5 different sequences of images per image rate. The procedure was repeated twice for each subject, with the same 5 sequences, after an arbitrary rest period. All images were presented to the subject for the same amount of time, from now on referred to as 'Image Display Period' (IDP), after which a neutral background appeared for the same duration, named 'Inter-Image Period' (IIP). The number of images was adjusted in each case so that the total length of one sequence was maintained constant and equal to 40 seconds. Likewise, the number of oddball images in each sequence was adjusted so that the ratio with respect to the number of non-oddball images was kept uniform (10%). Oddball images were placed randomly in the sequences. The parameters of the first experiment protocols, further referred to as the 'Reliability vs. Speed' experiment, are summarized in Table 1.

3.2 Experiment 2 - 'Fuzzy' classes: Beyond the oddball paradigm

Our second experiment aims to assess the potentiality to go beyond the oddball paradigm (in which classification is commonly restricted to images that belong to clearly pre-defined binary classes). Here, we extend this concept to three classes in which one of them is a 'non-obvious oddball', i.e. one for which scientific expertise is required to be recognised. Note that this implies that when confronted with these images, a mere visual pattern-matching process (e.g., as one expects when recognising a face) is insufficient to ellicit a clear response.

The visual stimuli used in this experiment were taken from the European Space Agency's database of 'multilayer coatings for thermal applications'.[2] This database contains images obtained during the process of designing a multilayered material that exhibits pre-defined thermal emissivity profiles (we call *target* the ideal profile). In these pictures, the spectral directional properties of the material are presented as 2-dimensional contour plots, with axes showing angle and wavelength parameters; the colour of the point represents the magnitude of the target parameter (such as emittance for example). The image set used was taken from different optimization experiments for different desired properties of the material, and their contours plotted in a normalized range of parameter values (stripped from the axes and the legend). However since a material matching exactly the desired properties of the target is often not obtainable, the best solution can therefore only be similar to the ideal target solution: 'clear oddballs' are chosen among the profiles that look very similar to targets, whereas 'non-obvious' ones are fuzzy and more hardly classifiable based on simple color pattern-matching rules.

Whether 'non-obvious oddballs' are distinguished from obvious ones or not hence requires knowledge about the optimisation process and its implications, to be consistently ranked appropriately; so one would expect the three distinct classes to be apparent only for experts.

In each session, subjects were initially shown what target images look like, and instructed to "look for similar profiles". Participants were also shown examples of images considered as obvious oddballs, thus providing them with an idea of the range of differences considered as acceptable between target solutions and 'good' solutions. Then, a sequence of images was displayed containing either very different properties from the ideal target (background images), very similar to the target (obvious oddballs) or slightly similar to the target (non-obvious oddballs). Examples of such images are shown in Figure 3.2, and the parameters of the experimental protocol are presented in Table 2. The presentation rate is maintained identical to that of the first *Reliability vs. Speed* experiment, in order to allow for a reliable classification of the oddballs. A notable difference however is that no Inter-Image Period is used whatsoever in this case (IIP=0ms), in order to increase the number of processed images per minute.

Six subjects took part in the experiment, one was an expert scientist from the European Space Agency familiar with the significance of the second image set. Two different target images were used, and for each one of them 5 sequences of images containing 3 obvious and 3 non-obvious oddballs. As in previous experi-

No. subjects	*No. targets*	*No. of seqs per target*	*Images in seq.*	*Oddballs in seq.*	*Reps*	*IDP/IIP (ms)*	*T (s)*
6	2	5	50	3+3	2	500/0	25

Table 2 Parameters of the three classes experiment

[2] The database can be visited at the link `www.esa.int/gsp/ACT/nan/op/bigrunresults.htm`

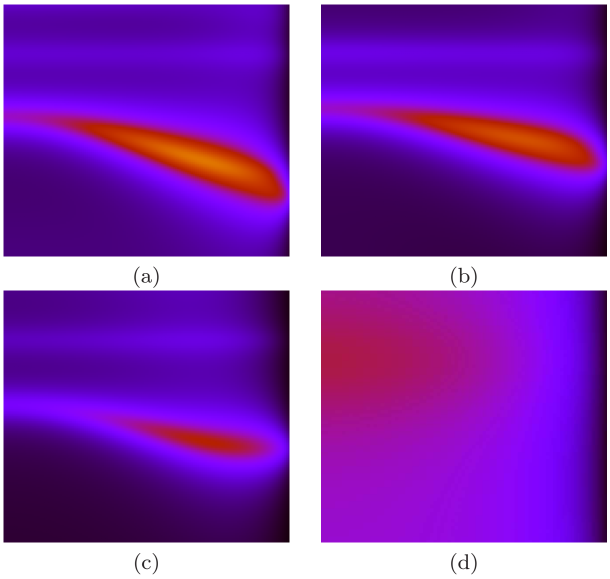


Fig. 2 Examples of (a) a target, (b) obvious oddball, (c) non-obvious oddball and (d) background images used in Experiment 2.

ments, measurements were conducted twice. A moderate speed (IDP=500ms) was used for the presentation of images (sequences of 25 seconds in length).

4 Signal Processing and Classification

EEG data recorded in both described experiments were processed in an attempt to assess if the signal carries sufficient information for automatic recognition of the subject's scientific interest in the presented images. This section describes adopted signal processing procedure and the methodology of the signal clasifiers construction.

4.1 Analysis of the Average Signals

The first step in signal analysis was a study of the averaged waveforms, aimed to visual demonstration of the different signatures embedded within the subjects' brainwaves (when confronted with 'oddball' and 'non-oddball' images in the case of the first experiment, and with 'non-oddball', 'obvious' and 'non-obvious oddball' in the second experiment). These were constructed out of the raw measurements from the Pz electrode, which has been shown to yield the strongest differentiation between stimuli and non-stimuli, and then low-pass filtered at a cut-off frequency of 14Hz – established after empirical observation of the spectral distribution. Averaged signals were then obtained across the stimuli classes and/or the participants.

4.2 Signal Processing and Feature Extraction

While the analysis on the averaged waveforms holds interesting information, it is only through the correct classification of the recorded data that we can assert the utility of our approach. In order to construct an artificial classifier for the recorded signals, one has to first pre-process the raw data, and then perform feature extraction to pinpoint those characteristics of the signal which carry information relevant to the classification task.

Initial empirical experimentation pointed out that the ideal pre-processing technique differs depending on the subject and experiment considered. A generalized approach was hence adopted which achieved good results across all subjects, and which exploits both time and frequency information at varying degrees of resolution in the time-frequency domain. For each channel, the following features were extracted (the temporal offset of extraction is the presentation time of the stimulus to the subject):

– 14 samples were extracted from the signal for the time-window between 220ms and 810ms, low-band filtered at a cut-off frequency of 14Hz. A time resolution of 40ms (inferior to any IDP) was thus here obtained. It is intended to encode the main structural differences in time between oddballs and non-oddballs.
– Spectral information – as obtained from the Fast Fourier Transform (FFT) – of the raw signal (the DC component is previously removed) during the time-window ranging from 220ms to 620ms in which the P300 is expected. 5 features were extracted for frequencies from 1hz to 15hz at a spectral resolution of 3Hz, which attempt to point out differences in the high frequencies over a short time-frame.
– Additional spectral information of the low frequencies between 1Hz and 5Hz for the whole signal (time window between 220ms and 1000ms). 5 attributes were chosen, which thus encode changes at a resolution of 1Hz.

Note that this methodology intends to take advantage of both variations in time and frequency, and that it targets the specific features where changes are expected. Moreover, two degrees of resolution were combined to improve the accuracy of the model.

Let us emphasize nevertheless that in practice there is a great degree of variation in individual samples, and that a P300 peak may well reach its maximum amplitude at 450ms rather than at 300ms, as documented in [6]. Our approach attempts to account for such variability through the choice of carefully selected time-windowing models. Additional signal processing algorithms were experimented with, namely Principle Component Analysis (PCA) and Haar Wavelet Coefficients, yet these demonstrated no additional performance gain to classification accuracy when combined with the methodology outlined.

Before classification, samples were normalized into the range [-1,1]. Finally, for each stimulus 24 attributes were extracted form each dataset. As mentioned in section 3, in this paper we only detail the results obtained by recording with the 4-channel device. Thus, since the EEG setup consists of 4 channels, an overall feature vector of 96 features per stimulus wa gathered.

4.3 Classification

For the experiment 1, the classification task we were attempting is applied on binary data, where samples belong to one of two classes. We therefore selected a Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel [28] as our main classification technique, already shown in [14] to be suited to the task of classifying ERP signals. Our implementation made use of both the WEKA toolkit [29] and the LibSVM library [5].

As derived from the description of the *'Reliability vs. Speed'* experiment, two fundamental machine learning challenges are here encountered, namely that of the class imbalance problem, and the curse of dimensionality [1]. The first manifests itself in the probability of an oddball image, which is initially set to only 10%, but for which we intend to achieve the highest precision in detection. The curse of dimensionality, on the other hand, results from the relationship between the number of oddball samples available for training and the length of the feature vectors. As specified in the experimental protocol, the duration of each trial is to remain constant (i.e. 40s) therefore entailing important variations in the number of oddballs in each trial (e.g. only 30 in the slowest case, and up to 382 for the fastest, as further detailed in Table 3). Whilst this enabled a degree of cross-comparability between subjects and experimental parameters, it certainly has an impact upon the classification task too.

IDP / IIP	*# Non-Oddballs*	*# Oddballs*	*Ratio NO/O*
500ms	400	30	13.3
300ms	670	61	11
150ms	1330	164	8.1
100ms	2000	230	8.6
50ms	4000	382	10.4

Table 3 Distribution of Oddball and Non-Oddball Stimulus in Experiment 1

To address this issue, we pruned the feature vectors from their original length of 96 attributes to 35 attributes via an SVM attribute evaluator (as implemented in the Weka toolkit). This process was conducted on a per-subject basis, primarily motivated by the significant variations existing between participants. We arrived at the figure of 35 attributes through empirical testing, for which little performance degradation was recorded. Further pruning did however lead to a pronounced drop-off in performance. Note that although this capacity to discard nearly two-thirds of our feature vector translates the presence of highly redundant or non-discriminative attributes, an automated methodology would turn out to be inadequate under the current conditions because of the strong fluctuations across subjects.

An in-depth analysis was also performed to quantify the influence of each single channel during the classification procedure, in an attempt to assess whether all four channels provide meaningful information and should therefore be maintained, or whether – on the contrary – some of them should be discarded. The 35 most relevant attributes (as reported by the SVM attribute selection algorithm) were iteratively selected among all considered features for each dataset, and they were given a score according to their contribution. A score of 35 was given to the most relevant attribute and 1 to the least. Scores were then added up for each channel and compared, hence giving an insight into their individual impact on the overall classification. For all experiments and subjects, it was observed that two channels (Pz and P3, with 27% and 28% impact respectively) had consistently more effect on the overall classification than Cz and P4. Yet it was shown that all four channels do add relevant information, later employed by the classifier, and should thus be maintained for our recognition purposes.

As for the class-imbalance problem, we implemented a modified bagging approach, similar to that of Natsev et al. [15] where a balanced training set was iteratively constructed by considering on the one hand, a set of oddball samples – identical for all evaluations – and on the other, an equal number of randomly-selected non-oddballs. Note that this differs from traditional bagging approaches as only one of the classes is randomly selected.

Random sub-sampling cross-validation was then performed to iteratively build the classifier, whereby we instituted an approximate 66/33 split between training and test samples based upon the number of oddballs. Training was undertaken on a balanced dataset as aforementioned. Testing, on the other hand, was carried out on a set of instances that preserved the initial ra-

IDP / IIP	*Training Set (O/NO)*	*Test Set (O/NO)*
500ms	20/20	10/133
300ms	40/40	16/176
150ms	100/100	64/158
100ms	160/160	72/602
50ms	300/300	82/853

Table 4 Training and Test Set Distribution of Oddball (O) and Non-Oddball(NO) samples.

tio between oddballs and non-oddballs, hence faithfully translating the classification performance that would be expected in a real situation. The test-set thus comprised all remaining oddball images, along with randomly chosen images among the remaining non-oddballs so as to keep a 10% ratio between them (Table 4 details the number of training and test samples used in each case). The cross-validation methodology was constructed out of 30-folds, and for each fold a grid-search optimization was run to determine the best parameters (C,γ) for the SVM. To evaluate classifier performance both ROC graphs and the AUC (Area Under the Curve) metric were used.

For the experiment 2, the approach adopted has to be slightly different, as instead of one binary classification task, there are now three image classes ('background', 'obvious' and 'non-obvious' oddballs). In order to handle this situation, a two-step analysis has been performed. First, our approach addressed the problem of discerning between oddballs and non-oddballs, and then tackled the question of whether 'obvious' and 'non-obvious' oddballs may also be distinguished. Note that these steps are intended to complement each other. The first aims to confirm that the performance reached under the conditions of experiment 2 is comparable to that of experiment 1, hence asserting that the properties previously observed are retained. The second 'obvious vs. non-obvious' analysis embodies a deeper evaluation which focuses on whether it may be possible to push the classification beyond the initial two-class description. It attempts to recognize patterns that respond to a higher-level cognitive process related to scientific interest, even for images that are non salient, i.e. not clearly defined in terms of what constitutes an interesting one.

For the 'oddball vs. non-oddball' classification, all obvious and non-obvious stimuli were gathered in a unique class (120 instances) and compared to non-oddball images which comprised 880 samples. In the second analysis, only the oddball dataset was considered, split into obvious and non-obvious images, which comprised 60 instances each. Random sub-sampling validation was performed on balanced training sets as aforementioned. These were built out of 80/80 samples in the first case and 45/45 instances in the second. The parameters of the classifier (C, γ) were optimized for each fold through grid-search, and testing was then iteratively undertaken on 40 oddballs/290 non-oddballs for the first analysis (hence maintaining the ratio between both classes), and 15 obvious/15 non-obvious for the second. This procedure was repeated 50 times with randomly-chosen initial seeds, and results were averaged across seeds.

5 Results

We present hereafter the analysis of the recordings obtained throughout the experimental sessions, and results obtained by the created signal classifiers.

5.1 Experiment 1 - Averaged Signal Analysis

For all combinations of IIP=IDP values in experiment 1, figure 3 depicts the averaged results for one particular subject (similar trends were encountered for all subjects in all experiments). Note that all curves present a "positive deflection in voltage at a latency of roughly 300 ms" corresponding to the event related potential (ERP), the P300. This peak embodies a cognitive function which, in this particular experiment, relates to the decision-making process experienced by the participants when deciding whether the picture is an oddball or not. Interestingly, the magnitude of the P300 clearly decays for faster visual stimuli rates. In addition, for small IDP and IIPs, a visually evoked potential at the same frequency as the IDP can also be clearly observed, hence translating the effect in perception of each single image (see for instance the averaged non target signal for IDP=100 and IDP=50ms). Even for these two extreme cases, a faint footprint of the P300 wave is still distinguishable from the augmented amplitude of the oscillations at around 300 ms. This preliminary analysis shows that the experimental protocol presented provokes a response which is recorded in the EEG signal, and hence suggests to search for an automated procedure to discriminate oddball from non-oddball images using the single signals recorded after each image is presented.

5.2 Experiment 1 - Signal Classification

Figure 4 and table 5 present the ROC graphs and the AUC metric values respectively for the classifiers created from the signals gathered in experiment 1. The results readily demonstrate that it is indeed possible to construct a discriminative classifer, even for the fastest

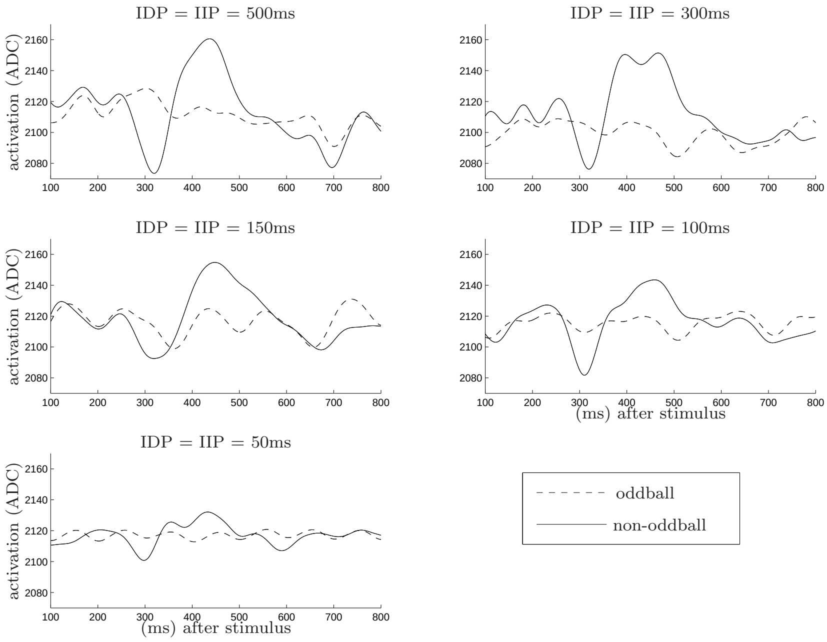


Fig. 3 Averaged EEG reading during the 800ms following the image presentation during the *Reliability vs. Speed* experiment. Different IDP/IIPs are shown.

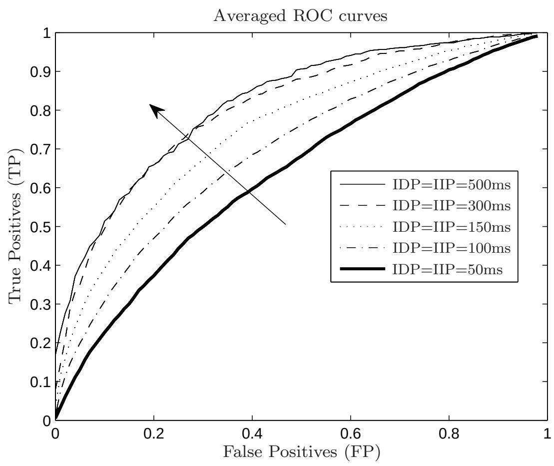


Fig. 4 Averaged ROC Curves for Experiment 1. The arrow underlines the improvement, as measured by the AUC, as the speed is reduced.

	500ms	*300ms*	*150ms*	*100ms*	*50ms*
Subject 1	0.8254	0.7997	0.7291	0.6702	0.6276
Subject 2	0.8297	0.8164	0.8012	0.7492	0.6114
Subject 3	0.9043	0.7844	0.6593	0.6282	0.6362
Subject 4	0.6946	0.8072	0.7948	0.7207	0.6524
Average	0.8135	0.8019	0.7461	0.6821	0.6319

Table 5 AUC Values across subjects for Experiment 1.

case considered, albeit with far reduced accuracy. Overall it is emphasized that a gradual linear decline underlies the change in performance from the 500ms experiment to the 50ms one, going from 0.81 to 0.62 AUC average values. Note nevertheless that for the fast cases (i.e. 100ms and 50ms) there are up to 8 stimulus events occurring in the sampling window of 220ms-1000ms for feature extraction. As such, significant overlap in the ERP's being generated is expected. Under such circumstances the extraction and classification of any signal from the very fast experiments is a very positive development.

It is important to note note that further specific user-dependent patterns could also be considered, which would obviously help improve the presented classification techniques. We remind nevertheless our concern about achieving a *global* methodology, even for the case of such a simple setup – cheap EEG device of only 4-nodes – and under visual stimulus at very fast presentation speeds. The conclusion of this experiment is that this is indeed possible; a compromise may also be derived from our results that accounts for both speed, reliability and usability.

5.3 Experiment 2 - Averaged Signal Analysis

The dataset used for the experiment 2 includes images that can be classified in three different ways: 'non-

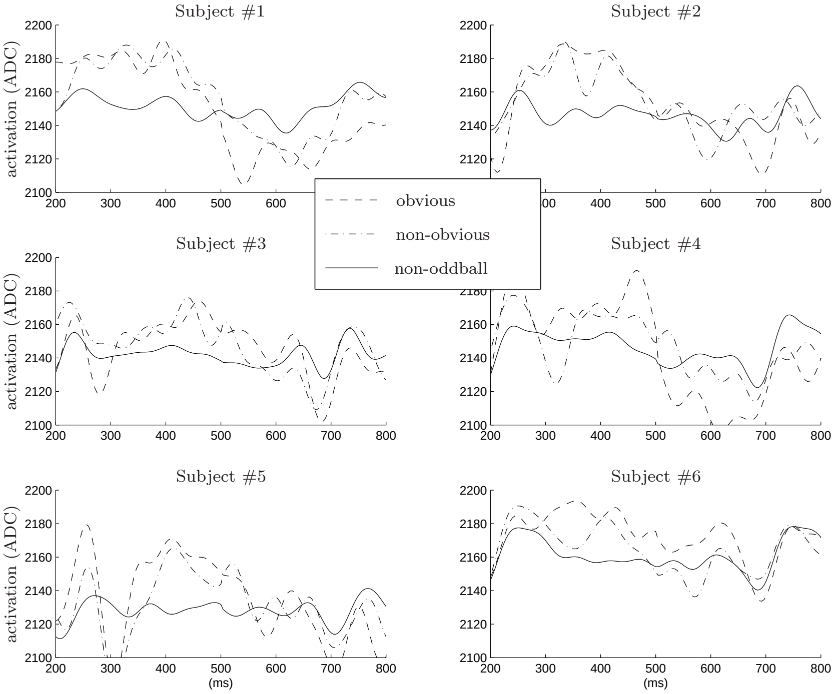


Fig. 5 Averaged EEG reading during the 800ms following the image presentation during experiment 2.

oddball', 'obvious oddball' and 'non-obvious' oddball. Here we evaluate the signal averages after the presentation of images belonging to each one of the three classes. In figure 5 the results of the averaging process for the six participating subjects are shown. For all of them, the averaged signal of the oddball classes is clearly distinguishable from the averaged signal of the non-oddball class. For some subjects the difference can be summarized in the detection of a positive peak around 300 ms (the P300 wave), and for some subjects this is preceded by a sharp negative peak (e.g., Subject #3,#4,#5). The EEG readings from the Pz channel are, also in this experimental protocol (IDP=500ms, IIP=0), discriminating between oddball and non oddball, but the P300 wave is somehow provoked less neatly and other features of the signal are emerging (the absence of an inter-image period is likely to be the cause together with the particular image set used).

5.4 Experiment 2 - Signal Classification

The AUC values for the constructed classifiers are reported In table 6. The degree of classification in the 'oddball' vs. 'non-oddball' case is proven to be in the range of experiment 1, with AUC values around 0.80 for most subjects. The recognition between background and stimuli is thus not significantly affected by the current experimental conditions, which may therefore be used as a testbed to perform further studies beyond the oddball paradigm.

	oddball vs. non-oddball	*obvious vs. non-obvious*
subject	*AUC*	*AUC*
1	0.80	0.50
2	0.84	0.55
3	0.78	0.64
4	0.85	0.61
5	0.81	0.44
6	0.68	0.63

Table 6 AUC results for the second experiment

An overall evaluation of the recognition between obvious and non-obvious oddballs already points out that this issue is somewhat more delicate. The performance is shown indeed to yield less accurate results, with most AUC values below 0.65, mainly as a consequence of the similarities between the two signals considered. Yet some interesting patterns may already be pointed out, which confirm our initial observations *(which were? ...)*. Note indeed that the degree of accuracy is distributed as mentioned in the previous section, namely recordings from subjects #1, #2 #5 cannot be classified with a higher accuracy than that of random guess, whereas in the case of the other subjects a certain degree of differentiation is possible. Subjects #3, #4 and #6 present an AUC well above chance (0.64, 0.61 and 0.63 respectively).

These results represent as such a proof of concept on how the initial oddball paradigm may be extended to allow the automated extraction of more than two classes from EEG signals, even for classes that are 'fuzzy' (and which may thereby convey signs of personal interest).

A careful reader will note that the perspective adopted here is mainly computational, and our results only derived from the classification capabilities. We do not claim to conclude here that fundamental neurophysiological differences underlie the mechanisms that generate the 'obvious' and 'non-obvious' signals. Our classification is performed regardless of whether the neural mechanism at their origin is actually different or not. Note for instance that, even though subjects #2,#3 and #6 present similar performance, in the case of the latter, the differentiation seems to be related to a positive peak after 300ms,[3] whereas for the first ones it results from an important negative deflection at 250ms. In the framework of our study, such differences do not provide any further insight into the classification problem and are hence treated identically.

6 Conclusion

In this article, we propose a novel approach to gain insight on subjects interest on given images. We build large data sets (interesting, non-interesting) automatically and directly from EEG signals classified during a rapid serial visualization experiment. This allows to avoid the use of long and tiring interviewing sessions, and to access directly the expert implicit visual processing.

In this context, we analyze the classification accuracy that can be achieved by employing a simple, wearable 4-node EEG setup. Using standard machine learning techniques we find that satisfactory results (around 80% classification accuracy) can be obtained while the subject is processing 100 images per minute (and up to 90% in slower cases). We also find indications that some subjects, show the ability to discriminate among more than two classes in their recorded EEG signals, and we speculate this ability is related to their scientific expertise on the displayed images.

Acknowledgments

The authors wish to acknowledge the ESA Ariadna scheme (`www.esa.int/gsp/ACT/ariadna/index.htm`) for having initiated and supported this research. The authors from DCU additionally wish to acknowledge Science Foundation Ireland under grant 07/CE/I1147 for partial support of this work.

[3] Interestingly, subject #6 was the only subject to be a scientist and to be very familiar with the scientific meaning of the images.

References

1. R. Akbani, S. Kwek, and N. Japkowicz. Applying Support Vector Machines to Imbalanced Datasets. In *Proceedings of the 15th European Conference on Machine Learning (ECML 2004)*, pages 39–50, Pisa, Italy, 2004.
2. B.Z. Allison and J.A. Pineda. Erps evoked by different matrix sizes: implications for a brain computer interface (bci) system. *IEEE Trans Neural Syst Rehabil Eng.*, 11(2):110–3, 2003.
3. ITU Radiocommunication Assembly. Methodology for the subjective assessment of the quality of television pictures. Technical report, 1974-2002.
4. A.G. Barto. Intrinsically motivated learning of hierarchical collections of skills. pages 112–119, 2004.
5. C. Chang and C. Lin. *LIBSVM: a library for support vector machines*, 2001.
6. M. Comerchero and J. Polich. P3a and p3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, 110(1):24 – 30, 1999.
7. E. Donchin, K.M. Spencer, and R. Wijesinghe. The mental prosthesis: Assessing the speed of a p300-based brain-computer interface. *IEEE Trans. Rehab. Eng*, 8:174–179, 2000.
8. L.A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70(6):510–523, December 1988.
9. D. Friedman, Y.M. Cycowicz, and H. Gaeta. The novelty P3: an event-related brain potential (erp) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, 25(4):355 – 373, 2001.
10. A. Gerson, L. Parra, and P. Sajda. Cortically coupled computer vision for rapid image search. *IEEE Transactions on neural systems and rehabilitation engineering*, 14(2):174–179, 2006.
11. D. Hayden, S. Chien, D. Thompson, and R. Castano. Onboard clustering of aerial data for improved science return. In *Proceedings of the IJCAI-09 Workshop on Artificial Intelligence in Space*, 2009.
12. T. Hruby and P. Marsalek. Event-related potentials-the P3 wave. *Acta Neurobiologiae Experimentalis*, 63(1):55–63, 2002.
13. L. Itti and P. F. Baldi. Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems, Vol. 19 (NIPS*2005)*, pages 547–554, Cambridge, MA, 2006. MIT Press.
14. F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of Neural Engineering*, 4(2):R1–R13, 2007.
15. A. Natsev, M. Naphade, and J. Tesic. Learning the semantics of multimedia queries and concepts from a small number of examples. In *Proceedings of the 13th annual ACM international conference on Multimedia (MULTIMEDIA '05)*, pages 598–607, Singapore, Singapore, 2005.
16. P.-Y. Oudeyer and F. Kaplan. What is intrinsic motivation? a typology of computational approaches. *Frontiers in Neurorobotics*, 1, 2007.

17. C. Ranganath and G. Rainer. Neural mechanisms for detecting and remembering novel events. *Nat Rev Neurosci*, 4:193–202, March 2003.
18. M. Rucinski. Curiosity cloning image viewer user's manual. Technical Report CCIVUM01, European Space Agency, the Advanced Concepts Team, 2008. Available on line at http://www.esa.int/act.
19. P. Sajda, E. Pohlmeyer, Jun Wang, L.C. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M.K. Singh, and Shih-Fu Chang. In a blink of an eye and a switch of a transistor: Cortically coupled computer vision. *Proceedings of the IEEE*, 98(3):462 –478, mar. 2010.
20. J. Schmidhuber. Simple algorithmic principles of discovery, subjective beauty, selective attention, curiosity and creativity. In *ALT*, pages 32–33, 2007.
21. W. Schultz, P. Dayan, and P. Read Montague. A neural substrate of prediction and reward. *Science*, 275:1593–1599, 1997.
22. E.W. Sellers and E. Donchin. A p300-based braincomputer interface: Initial tests by als patients. *Clinical Neurophysiology*, 117(3):538–548, March 2006.
23. P. Shenoy and D.S. Tan. Human-aided computing: utilizing implicit human processing to classify images. In *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, pages 845–854, New York, NY, USA, 2008. ACM.
24. T. Smith, S. Niekum, D.R. Thompson, and D.S. Wettergreen. Concepts for science autonomy during robotic traverse and survey. In *Proc. IEEE Aerospace Conf.*, 2005.
25. D.R. Thompson. Domain-guided novelty detection for autonomous exploration. In *IJCAI'09: Proceedings of the 21st international jont conference on Artifical intelligence*, pages 1864–1869, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.
26. S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381(6582):520–522, June 1996.
27. R. Vanrullen and S.J. Thorpe. The time course of visual processing: From early perception to decision-making. *J. Cognitive Neuroscience*, 13(4):454–461, 2001.
28. V. Vapnik. *The nature of statistical learning theory.* Springer-Verlag New York, Inc., New York, NY, USA, 1995.
29. I. Whitten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques; (2nd ed.).* Morgan Kaufmann Publishers, San Francisco, CA, USA, 2005.