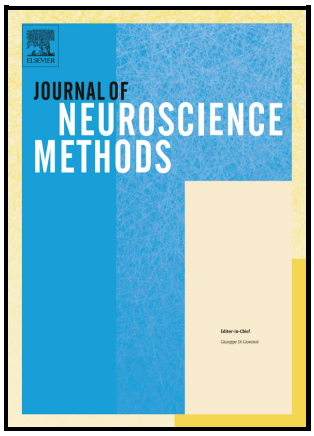
Detection of atypical attentional behaviors in young subjects

Wafa Rekik, S. Le Hégarat-Mascle, Souhir Ezzedini, Giovanni de Marco

Detection of atypical attentional behaviors in young subjects

Wafa Rekik, Sylvie Le Hégarat-Mascle, Souhir Ezzedini, Giovanni de Marco



PII: S0165-0270(24)00086-4

DOI: https://doi.org/10.1016/j.jneumeth.2024.110141

Reference: NSM110141

To appear in: *Journal of Neuroscience Methods*

Received date: 25 January 2024
Revised date: 9 April 2024
Accepted date: 16 April 2024

Please cite this article as: Wafa Rekik, Sylvie Le Hégarat-Mascle, Souhir Ezzedini and Giovanni de Marco, Detection of atypical attentional behaviors in young subjects, *Journal of Neuroscience Methods,* (2024) doi:https://doi.org/10.1016/j.jneumeth.2024.110141

Detection of atypical attentional behaviors in young subjects

Wafa Rekik[1], Sylvie Le Hégarat-Mascle[2], Souhir Ezzedini[3], Giovanni de Marco[3]
[1] Research Laboratory COSIM, Higher School of Communications of Tunis, University of Carthage, Route de Raoued 3.5 Km, Cité El Ghazala, Ariana, 2088, Tunisia.
[2] SATIE, Universitè Paris-Sud, Universitè Paris-Saclay, France.
[3] LINP2 Lab, Paris Nanterre University, UPL Paris, France.
***** Correspondence: Giovanni de Marco demarco.giovanni@gmail.com

Abstract

Background: Vigilance ability refers to the accuracy and speed with which a person performs a cognitive-motor task, either voluntarily (endogenous mode) or following a warning stimulus (exogenous mode). In the context of a force production task, our study focuses on the impact of the states of vigilance by proposing an original approach that allows distinguishing between good (inlier) and poor (outlier) participants. We assume that the use of an external signal and duration of the temporal preparation (foreperiod) increase the speed and the precision of motor responses. Our objective is particularly challenging in the context of a limited dataset with a high level of noise.

New method: Our original methodological approach consists of coupling the RANSAC (RANdom SAmple Consensus) algorithm with a statistical machine learning algorithm to handle noise.

Comparison with existing methods: Our clustering approach, based on the coupling of RANSAC methodology with ensemble classifiers, overcomes the limitations of conventional supervised algorithms that are either not robust to outliers (such as K-Nearest Neighbors) and/or not adapted to few-shot learning (such as Support Vector Machines and Artificial Neural Networks).

Results: The clustering results were validated in terms of reaction time distributions and force error distributions with respect to participant groups. We show that the use of an external signal and duration of the temporal preparation (foreperiod) increase the speed and the precision of motor responses.

Conclusion: Our study has allowed us to detect atypical attentional patterns and succeeds in separating the inliers from the outliers.

Keywords: vigilance, force production, reaction time, precision, attention, RANSAC, machine learning, clustering, outlier, inlier.

1. Introduction

Vigilance is crucial in how we interact with our environment, a trait that is even more pronounced and essential in animals. In humans, vigilance or alertness involves the ability to stay focused and respond quickly to unexpected changes or events. Several approaches, based on psychological questionnaires, cognitive tasks [6, 5, 25, 11] and physiological recordings [1, 18, 14, 28] have been proposed in order to characterize the different states of vigilance:

During a cognitive task (e.g., Go-No-Go and stop-signal tasks), the participants' brains go through phases of neural activation and suppression in response to visual, auditory, or combined stimuli. The assessment of their attentional state is determined by analyzing their average reaction times (RT) [6] and/or their ability to suppress an initiated motor action. In [3], the participant reacts to real and false stimuli, and the RT reflects their level of attention. In our work, in line with [9], we consider two types of alertness tasks: endogenous and exogenous. Alertness is an important component of attention. Endogenous attention occurs when the participant voluntarily (i.e., intentionally) directs his attention towards a goal. This type of alert (top-down) attention is controlled by the participant's expectations as he processes the expected information. It should be maintained over time and reach an optimal value during approximately 500 ms. With exogenous attention, an external signal (alertness) involuntarily (i.e., automatically) directs attention to a sudden source in the environment (e.g., noise, flash). Exogeneous attention should rise and fall very quickly (around 300 ms). Classically, the measures allowing us to precisely evaluate these two types of alertness are: the measurement of the reaction time (ms), the omission rate (%), and the anticipation rate (%) which increase in the presence of the exogenous signal. Another parameter to consider is the accuracy of motor responses (e.g., in the case of a dual task), which may decrease when the participant does not have sufficient time to prepare his motor response. The time interval between the planning and the initiation of the movement following the alert and

the motor response is called the fore-period (FP). The longer the FP, the shorter the RT, resulting in increased response accuracy because the participant has more time to control his response. Thus, performance is often described in the literature as a trade-off between speed and precision of the motor response [17].

In our study, we will focus on the effect of alertness on RT and show that other factors, such as the level of force (low, medium, and high) applied to a hand grip to produce a motor response, also affect the participants' performance. Our hypothesis is that both the force level and the fore-period in the exogenous alertness task influence the participants' ability to process information and, consequently, their motor responses. This hypothesis implies a relationship between the task conditions (force level and fore-period) and the participants' cognitive and motor performance. To test this hypothesis, we proposed an original approach to classify participants according to their behavior and performance during an exogenous attention task of force production. It is based on advances in Machine Learning (ML), which has experienced remarkable growth in popularity over the past few decades due to its excellent performance along with mathematical soundness. Although ML algorithms (more specifically Artificial Neural Networks, ANN) are mainly used in natural language processing and computer vision, where large databases also allow for deep learning models, these methods have been increasingly applied in the medical field (synthesis of drugs [15], understanding of DNA [2], detection of diseases from medical images [3]). However, ML methods are much less used in the evaluation of human behavior, such as with vigilance and attention, due to the scarcity of annotated data. As an example of such works, we mention [10], where the authors developed their own database consisting of EEG and eye tracking signals recorded during attention tasks performed in virtual reality. The EEG signals are assigned an

attention score (between 0 and 100 %) according to the eye tracking signal, which allowed them to apply a recurrent neural network [4].

With this study, we have a double objective (i) to define an ANN able to distinguish between endogenous and exogenous modes, (ii) to use this ANN to detect atypical attentional behaviors. In our work, the data (reaction time, force) were labeled according to the type of task performed (endogenous (endo) or exogenous (exo)), which, as mentioned above, are two different and important components of attention, favoring the application of ANN to model the two types of alertness. We assumed that the vigilance process will induce a different behavior and response in the endogenous mode versus the exogenous one (e.g., globally reduced reaction time) so that an ANN will be able to classify the response mode of a given participant based on his RT and force level values. To do this, we first construct an ANN classifier able to distinguish between endogenous and exogenous modes in the case of standard young participants, and then analyze the performance of this classifier for any given participant and force level. Note that to define a standard expected classification performance as well as predictions of a standard behavior, the network has to be trained despite the high variability of the participants and the presence of some outliers, which is not a trivial task. Indeed, first, the participants do not have equal attentional abilities, which impacts the way the participants use the alert signal to improve their performance i.e. to reduce their RT and improve their force level accuracy in exogeneous attentional mode. Second, our dataset is too small (27 participants) to be robust to the presence of aberrant (non-performing) participants, thus affecting the learning process and the endo/exo classification model. Thus, we also propose a methodological contribution, which consists in the adaptation of the RANdom SAmple Consensus (RANSAC) methodology to statistical machine learning. In our case, this allows us to automatically distinguish between the best performing participants (inliers)

and the anomalously low performing participants (outliers). In other words, RANSAC will help us to detect the ineffective participants, that behave like outliers for our ANN model.

Highlights:

- Atypical attentional patterns are detected in young subjects.
- A score assessing attentional capacity was established to distinguish participants with high (inlier) attentional performance from those with low (outliers) performance.
- This score is based on the coupling of RANSAC methodology with ensemble classifiers, which has been shown to handle outliers in a few-shot learning framework.
- The proposed score-based subject clustering improves the understanding of the duality between reaction time and force accuracy.

2. Participants, experimental paradigm and parameters

2.1. Participants

Twenty-seven healthy participants (14 males) aged between 20 and 35 years with no known history of neurological, psychiatric, or sensorimotor disorders were recruited. All participants were right-handed, as assessed by the Edinburgh inventory for right-handedness. Before the screening process, all participants were provided with a detailed explanation of the study's purpose and procedures. They signed an informed consent form before participating in the study. The institutional research board of the CPP Ile-de-France approved the study protocol.

2.2.Experimental paradigm

In order to study the vigilance states, an experimental protocol was designed as follows: 27 young and healthy people carry out two types of tasks (endogenous and exogenous). Each participant performs 27 trials in endo mode, followed by 27 trials in exo mode. This sequence of 54 trials is repeated four times. Each trial consists in focusing attention in the center of the screen and squeezing a force handle (handgrip) as quickly as possible when the participant hears the "Go" instruction while respecting the level of force requested, among the three levels: low, medium,

strong corresponding to 20 %, 50 % and 80 % of the maximum voluntary force (MVF), respectively. The order of force levels is randomized for both tasks. In the exogenous task, the imperative stimulus (Go) is preceded by an audible alert signal (50 ms duration beep) to alert the participant. The duration between the alert and the imperative stimulus varies according to three foreperiods (FP): 650 ms, 680 ms and 710 ms. During the endogenous task, there is no warning signal, so that the participant has no information about the imminent occurrence of the imperative stimulus. The alertness is synchronized with force signal recorded using a Biopac data acquisition system (MP150); the data is sampled with the Acknowledge software (version 4.2) at the sampling frequency of 10000 Hz.

2.3. Experimental parameters

In our study, we consider two complementary types of data: the reaction time and the force level accuracy. The reaction time is the time taken by a participant to respond to the imperative stimulus (Go signal). We focus on RT since we hypothesize that the vigilance process will correspond to participants who, on average, reduce their TR in exo mode thanks to the alert. It is calculated as the time difference between the imperative stimulus arrival and the time to reach 15% of the force signal deviation from the baseline (cf. Figure 1). Regarding the force error, it indicates how accurately a participant achieved the requested level of force among the three levels: low, medium, and high, corresponding to 20%, 50% and 80% of MVF, respectively. For each trial, it is calculated as the absolute difference between the executed and the requested forces:

$$\zeta_f = |requested\ force - executed\ force| \quad (1)$$

Depending on the type of alerting task, the RT and the force errors are annotated as either endo or exo. Figure 2 shows their evolution over time and the alternation between the two modes.

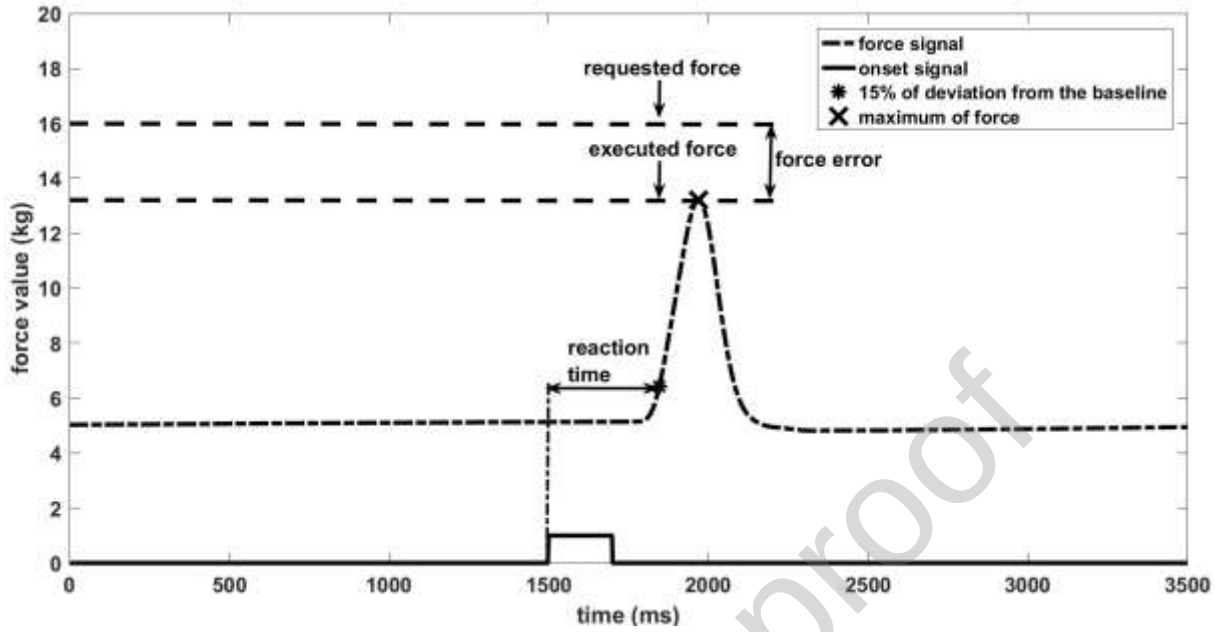


Figure 1. Reaction time and force error calculations using 15% of the deviation from the baseline and the maximum of the signal, respectively.

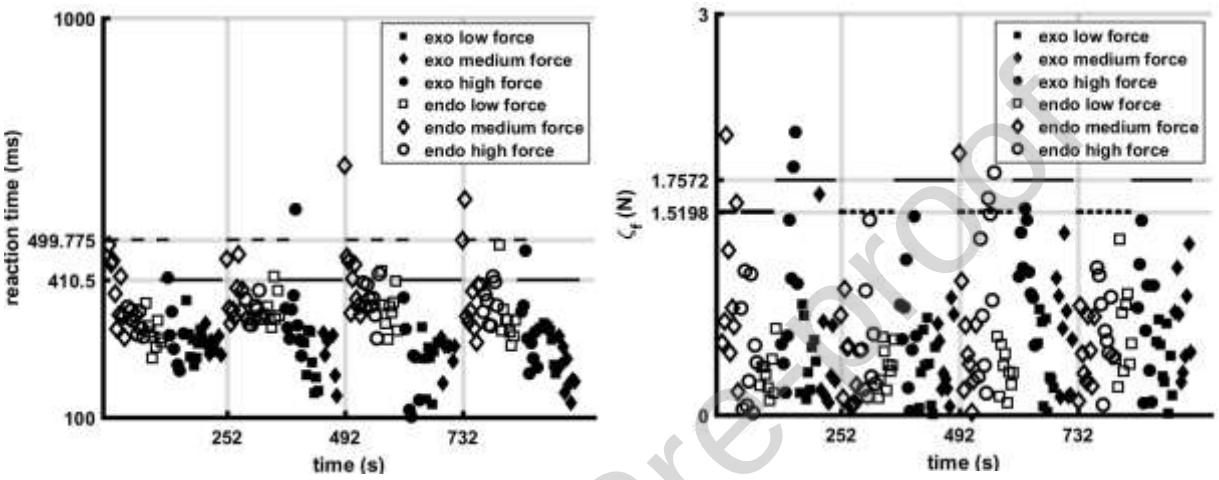
3. Data preprocessing and feature extraction

3.1. Statistical outliers removal

This step consists of removing the outliers from each participant's reaction times. An outlier corresponds to a response that occurs before the onset signal (i.e., an anticipated response) or to a missed response (i.e., with a delay greater than 1000 ms from the onset signal). For these two types of outliers, we rely on the usual statistical criteria, namely boxplots ([26,23]), setting two temporal thresholds equal to 100 ms and 1000 ms, respectively. However, these thresholds are not sufficient for quite focused participants who may still experience occasional lapses of attention during the tasks. Thus, for each mode (endo and exo), the RT values that fall outside the distribution of RT, i.e., greater than 1.5 times the interquartile range *IQR*, especially above the upper third quartile Q_3 (TR> Q_3+1.5 × *IQR*) are also considered as some outliers and then removed. Figure 2a shows two

RT thresholds for outlier detection: the solid and the dashed lines indicate the thresholds for exo and endo modes, respectively.

Similarly, for each participant, the force errors were processed (each mode separately) to remove the outliers that fall outside the force errors distribution in endo and exo modes (Figure 2b).



(a) (b)

Figure 2. Evolution of data over time for a participant: (a) reaction time; (b) force error.

3.2. Normalization of reaction time data

RT is the main variable we consider when comparing the attentional behaviors. However, its dynamic range, which varies between 100 ms and 1000 ms, depends on the participant. In order to put the RT of all participants on the same scale, a `standardization' step is required. Classically, the latter aims at centering the data around a mean of 0 and setting the standard deviation to 1. For any participant indexed i, denoting by ρ_i its RT measurements, with mean μ_i and standard deviation σ_i the standardized RT variable $\overline{\rho_i}$ is obtained as follows: $\overline{\rho_i} = (\rho_i - \mu_i)/\sigma_i$. Note that μ_i and σ_i parameters have been estimated after removing the outliers from the ρ_i distribution.

3.3. Feature extraction

As explained in the Introduction, our database is not large enough to train a Deep Neural Network. Nevertheless, we can benefit from the advances in statistical machine learning in the last decades, if we (i) consider a shallow neural network, (ii) feed it with well-chosen selected features so that the neural layers focus on the decision task.

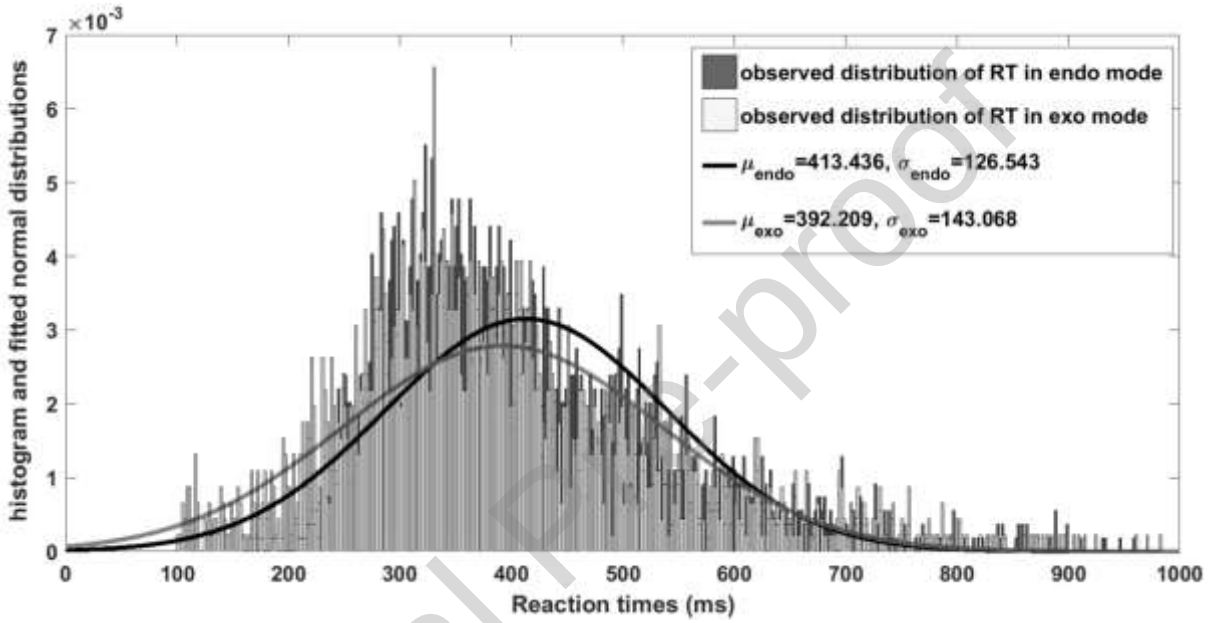


Figure 3. Distributions of RT values in endo and exo modes, without distinction between participants.

Figure 3 shows the histograms of RT values before normalization in endo and exo mode (when all participants are considered together), along with their fitted Normal distributions $N(\mu, \sigma^2)$ (up to a scale factor to fit the sample number). As we can see, the distributions ($p(\rho|endo)$ and $p(\rho|exo)$) heavily overlap, implying the inability to discern both attentional tasks based on the raw RT values. Therefore, we introduce higher level hand-crafted features. Specifically, among the temporal characteristics defined in [27] such as maximum, minimum, energy, variation, activity, complexity, and mobility, we consider those that take into account the specificities of alerting tasks (endo or exo), such as the median and the mean. Secondly, we take into account the

possible training of the participants along the repetitions and their ability to concentrate through the 10^{th} percentile and the minimum. Finally, in order to reduce the measurement noise, the considered features are low-pass filtered, i.e. they are computed based on several RT values gathered into sample groups. Each group consists of nine RT samples corresponding to nine consecutive trials with the same force level (low, medium, or high). This amounts to using 12 sliding windows of samples of size nine on the 108 trials of each alert task (endo or exo). Note that our `prior-informed' feature selection is in agreement with statistical tests, such as ANOVA and Fisher Discriminant Ratio (FDR) according to which a feature is all the more relevant as it minimizes the p-value and maximizes the FDR (Table 1). In the following of the study, we focus on the minimum and the median to feed the classifier.

features	p-value (↓)	FDR(↑)
minimum	$\mathbf{3.9e^{-18}}$	**0.2554**
10^{th} percentile	$\mathbf{4.5e^{-20}}$	**0.2877**
mean	$\mathbf{6e^{-9}}$	**0.1115**
median	$\mathbf{4.2e^{-6}}$	**0.0691**
90^{th} percentile	0.0017	0.0319
extent	0.0059	0.0244
coefficient of variation	0.3229	0.0032

Table 1. Results of ANOVA test and FDR on some features. Retained features are in bold.

4. Proposed approach

4.1. ANN classification

The proposed approach relies on the construction of a classifier able to distinguish endo and exo modes for participants having a standard attentional level. Considering our database, we opted for a shallow network (cf. Figure 4). Specifically, it consists of an input layer, an output layer and two hidden layers with six and four nodes, respectively. The input layer dimensionality is equal to the

number of considered features defined in Section 3.3. The output layer dimensionality is equal to the number of classes, namely two (endo and exo) in this work. The activation functions of the hidden layers are either sigmoid functions or Rectified Linear Unit (ReLU) function, and, for the output layer, a softmax function (since we deal with a classification problem).

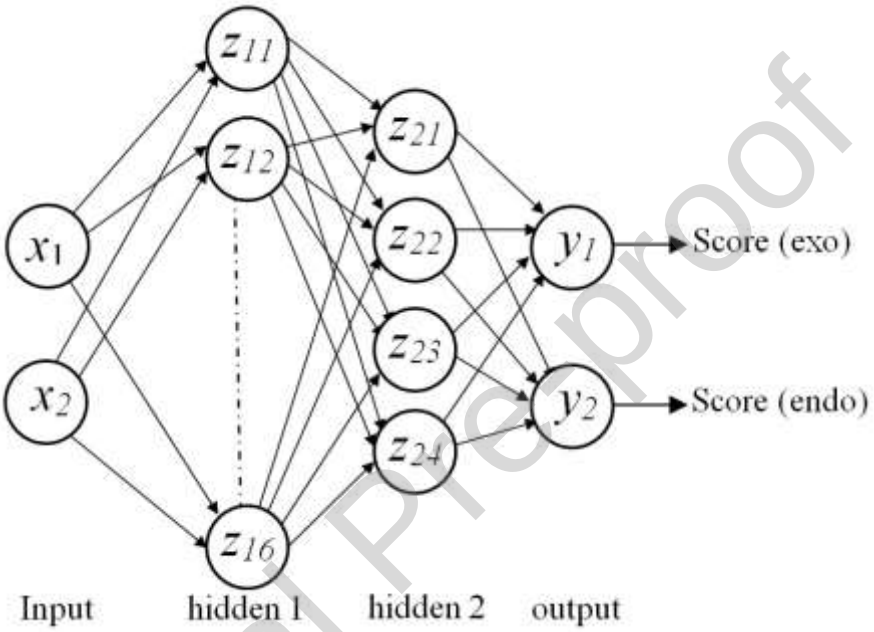


Figure 4. Shallow network architecture: two hidden layers with sigmoid activation functions and an output layer with a sigmoid function too; $X = (x_1, x_2)$ denotes the input feature tensor.

In our case, the attentional ability of the participants is highly variable which, in some cases, boils down to random labels (conversely to the usual machine learning datasets where the data has been carefully labeled). Thus, a first goal of this study is to construct a classifier whose performance applied to a given participant reflects his attentional ability. To do this, we have to (i) train our classifier despite the variability in the attentional abilities of the participants in the database, (ii) group the participants based on their attentional abilities. Next section describes the implemented solution.

4.2.Attentional abilities estimation

The basic idea for classifier learning despite the presence of participants with a very low attentional capacity is to focus on the other participants. However, the categorization of the participants (in terms of attention) is unknown. Then, inspired by the RANdom SAmpling Consensus (RANSAC) method, we estimate many models (classifiers) and select the one that is the most consensual, i.e. has the largest consensus set (also called inlier set).

RANSAC original method [12] was proposed to detect outliers in datasets. Belonging to the field of statistical robust estimation, it allows us to estimate the parameters of a parametric model in the presence of outliers, i.e. data that do not follow the model under consideration. Conversely to robust estimators such as M-estimators, it does not mitigate the weight of some data in the estimation, but it searches a subset of data without any outliers. To do this, it considers subsets of data as small as possible (to minimize the probability of the presence of an outlier) and evaluate the derived model in terms of consistency with the rest of the data. Since original paradigm proposition, many variants of the method have been proposed [7]. For example, the original consistency criterion was simply an error (for the considered data point) lower than a given threshold, but since then many other consistency criteria have been proposed, e.g. to make the model parameter estimation more robust with respect to the inlier set or to the inlier threshold, or to fit to a given application.

In this study, RANSAC is used to train our ANN model. The algorithm consists of estimating the endo/exo classification model that fits to the majority of the participants. Specifically, a given participant *fits* a given model if the performance of the model for that participant's data is greater than a given threshold. The consensus criterion (to be maximized) is then the number of participants that fit a model on test. Algorithm 2 describes the estimation of the subset of the

participants fitting a given model M based on the *accuracy* criterion *acc* calculated from the number of true positives (*tp*), true negatives (*tn*), false positives (*fp*), and false negatives (*fn*):

$$acc = \frac{tp + tn}{tp + tn + fp + fn} \tag{2}$$

Algorithm 1 describes the implemented solution. It is as follows: In the main loop, we randomly select a subset of participants of cardinality n (set as small as possible to avoid atypical participants) to train the ANN. The network is then used as a predictor for each of the remaining participants. For each given participant, based on these predictions and the associated ground truth labels, we can use the performance measure (accuracy denoted *acc* in Algorithm 1) of the ANN classification as an index of the similarity between that participant and the participants used for the classifier training. In order to increase the robustness of participant evaluation, we also take into account model's dependence on its initialization. Specifically, we train the model t_{max} times using different initialization weights and we store the best one (maximizing the evaluated performance).

Finally, note also that, for each ANN found as the best among the t_{max} training sessions, it is definitively kept only if its global consensus is greater than n_{min} after a final training considering its whole set of inliers (rather than the minimal one).

In this study, we set n=3, $t_{max} = 5$, τ_{acc}=0.7, $it_{max} = \left|C^n_{|S|}\right|$, n_{min} equal to 70% of the number of participants. Finally, Algorithm 3 presents the method for clustering the participants into three clusters. For each participant, it counts the number of `good' models (output of Algorithm 1) to which he fits by testing whether the participant belongs to the inlier sets associated to these good models (here the notation $[\![\,.\,]\!]$ refers to the Iverson brackets). In this study, we set τ_{top} and τ_{out} equal to 70% and 30% (of the total number of sets of inliers), respectively.

The algorithms were coded in Matlab, using the pattern recognition network provided by the Deep Learning Toolbox for implementing Artificial Neural Networks. The code is available at 'https://github.com/wafarekik/code-detection-atypical-behaviors.git'

Algorithm 1 Adapted RANSAC.

Require: Set of participants: S, RT dataset for all participants: $D = \{D_k\}_{k\in S}$, cardinality of the subset used for model fit: n, number of iterations: it_{max}, number of training sessions: t_{max}, accuracy threshold : τ_{acc}, minimum number of participants in line with a `good' model: n_{min}.

Ensure: Set of pairs of a `good' model and its associated subset of inlier participants : $\boldsymbol{\Theta_{M,I}}$.

Initialization: set $\Theta \leftarrow \emptyset,\ I \leftarrow \emptyset,\ it \leftarrow 0$

for it=1 to it_{max} **do**

set $S\leftarrow$ n participants randomly selected and $\bar{S}$ the set of the other participants.

$D_S \leftarrow RT$ data associated to S

$n_{it} = 0$

for t=1 to t_{max} do

$M \leftarrow$classifier trained on D_S using t^{th} random initialization of weights w.

Run Algorithm 2 with entries (M, S, D, τ_{acc}) to derive I the subset of participants in line with M.

if $|I| > n_{it}$ then

$n_{it} \leftarrow |I|$

Save data associated to inlier participants and initialization parameters in

$\Theta_{it} \leftarrow (\{D_k\}_{k\in I}, w)$

end if

end for

$M_{it} \leftarrow$ model trained using Θ_{it}'s data and initial weights.

Run Algorithm 2 with entries $(M_{it}, S, D, \tau_{acc})$ to derive I_{it} the subset of participants in line with M_{it}.

if $|I_{it}| \geq n_{min}$ **then**

$\Theta_{Models} \leftarrow \Theta_{Models} \cup \{(M_{it}, I_{it})\}$

end if

$it \leftarrow it + 1$

end for

Algorithm 2 Subset of participants in line with a model.

Require: Model M, set of participants to test: S, RT dataset for all participants: $D = \{D_k\}_{k\in S}$,

accuracy threshold: τ_{acc}

Ensure: inlier set: I

$I \leftarrow \emptyset$

for $k \in S$ **do**

$D_k \leftarrow$ RT data associated to participant k

Compute accuracy acc of model M on D_k

if $acc > \tau_{acc}$ **then**

$I \leftarrow I \cup \{k\}$

end if

end for

Algorithm 3 Participant clustering.

Require: Set of subsets of inlier participants (associated to a `good' model): $S_I = \{I_j\}$, top threshold: τ_{top} , outlier threshold: τ_{out}.

Ensure: Top inlier set: I_{top}, intermediate inlier set: I_{inter}, outlier set: I_{out}

Initializations: $I_{top} \leftarrow \emptyset$, $I_{inter} \leftarrow \emptyset$, $I_{out} \leftarrow \emptyset$

$N \leftarrow |S_I|$

for i=1 **to** $|S|$ **do**

$score_i = 0$

for j=1 to N **do**

$score_i \leftarrow score_i + [\![i \in I_j]\!]$

end for

$score_i \leftarrow \frac{1}{N} score_i$

if $score_i > \tau_{top}$ **then**

$I_{top} \leftarrow I_{top} \cup \{i\}$

else if $score_i < \tau_{out}$ **then**

$I_{out} \leftarrow I_{out} \cup \{i\}$

else

$I_{inter} \leftarrow I_{inter} \cup \{i\}$

end if

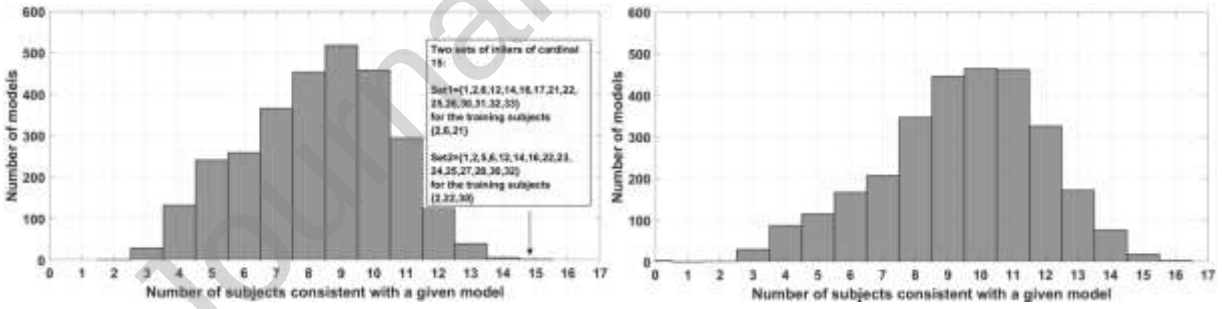
end for

5. Results

In this section, we present the main results obtained on our database.

5.1.Participant categorization

Figure 5 shows some results of Algorithm 1, namely the histograms of the number of models that fit a given number of participants, after first estimation (i.e., with training based on the random triplet of participants) and after final step (i.e., with training based on the whole inlier participant set). According to Figure 5a, only two models (derived from participant triplets {2,6,21} and {2,22,30}) have been fitted to a total of 15 participants. Each triplet refers to a subset of three participants randomly selected as input for Algorithm 1. However, according to Figure 5b, the maximum number of participants fitted to a model is 16, achieved by 3 models, and 18 other models fit groups of 15 participants. In the following, we focus on the corresponding inlier groups of cardinality equal to 15 or 16 (3 sets of cardinality 16 and 18 sets of cardinality 15). According to Figure 5, these 21 sets of participants associated to the best models (in terms of number of participants) form the set of subsets of inlier participants S_I at the entry of Algo. 3.



(a) (b)

Figure 5. Selection of the best classification models (case of sigmoid activation function for hidden layers) in terms of the number of participants that are fitted to it: (a) first training; (b) final training.

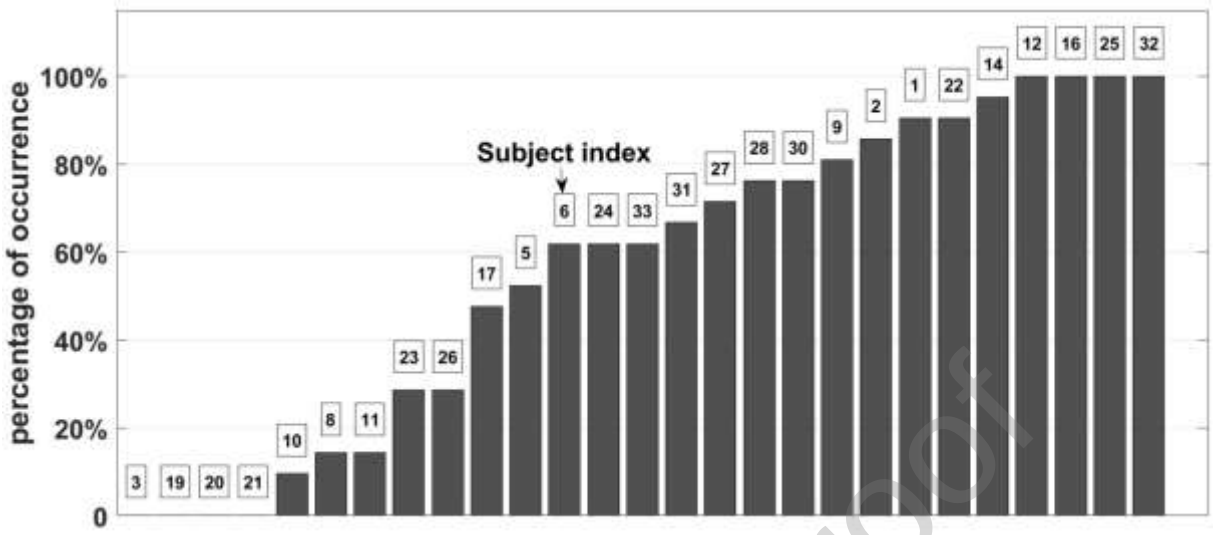


Figure 6. Participant evaluation according to the percentage of their occurrence in the inlier sets of cardinalities 15 and 16; case of sigmoid activation function for hidden layers.

Figure 6 plots the score of each participant calculated in Algorithm 3. Remember that the participant score reflects how much a participant shares similar behavior in terms of endo/exo response separation with other participants (represented through an ANN model). Then, the three groups of participants are: (i) the most conformal participants whose occurrences are greater than 70%; (ii) the outliers (atypical participants) with occurrences less than 30%; (iii) participants with intermediate conformity, i.e. between 30% and 70%. However, to increase the robustness of the derived group of participants, we also run the algorithm with ReLU (Rectified Linear Unit, [13]). Indeed, ReLU activation is very popular in Deep neural Networks (DNN) since the latter require simplest activation functions (less complex to compute than a sigmoid). For example, Krizhevsky et al., in their very famous article [16] often referred to as the beginning of deep learning, proposed to replace the hyperbolic tangent by ReLU activation to reduce the computation time by a factor of six. Obtained experimental results are very close (only three participants close to group thresholds switch), but considering both results allows us to get more robust participant groups (in the spirit of ensemble networks). The top inliers are now the participants who show complete

agreement between the ReLU and sigmoid-based networks, with their frequencies greater than 0.7 in both cases. As a result, the top inlier group includes participants with index in {1,12,14,16,22,25,32,28}, the outlier group includes those with index in {3,8,10,11,19,20,21,23,26}, and the intermediate inlier group consists of participants with index in {2,5,6,9,24,27,17,30,31,33}.

5.2.Robustness with respect to feature selection

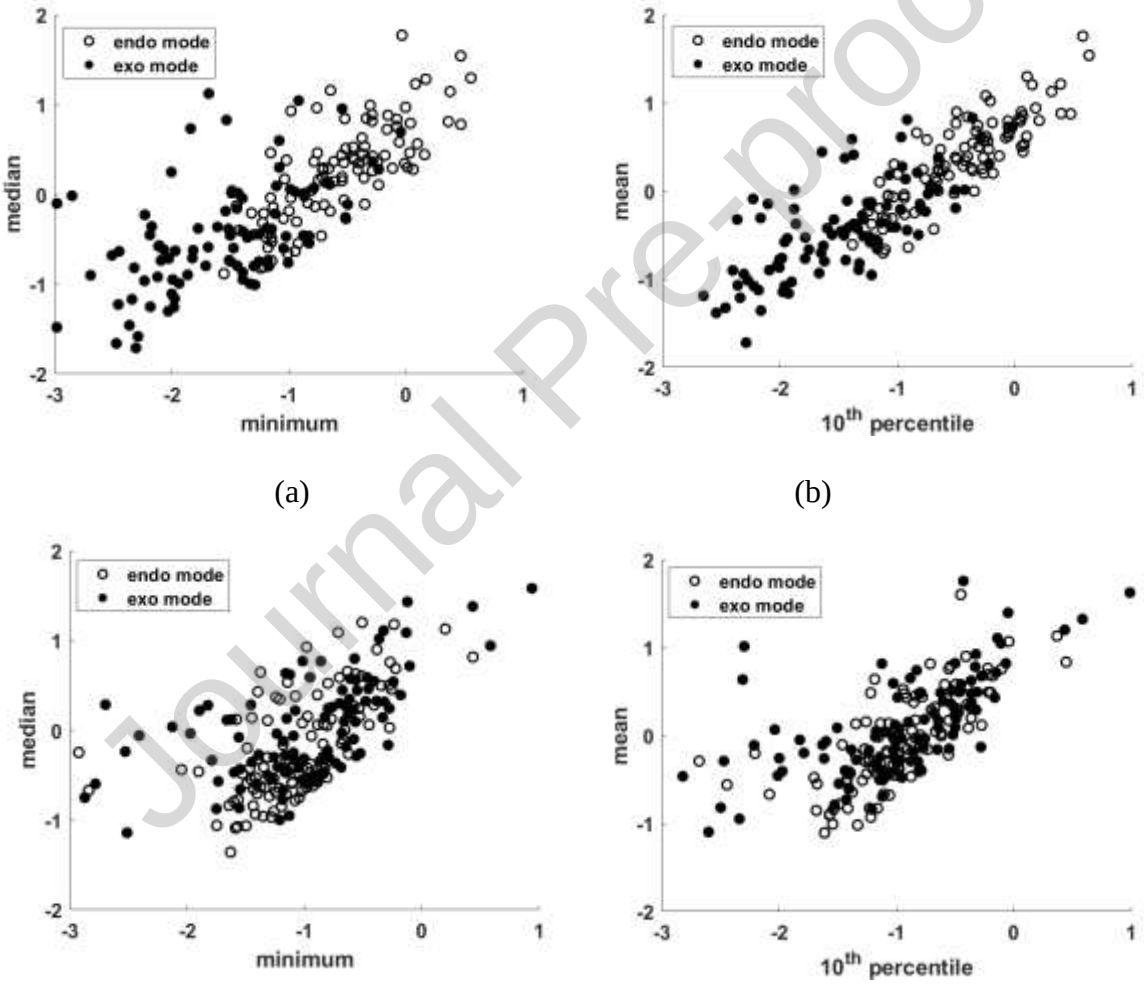


Figure 7. Distributions of standardized RT values in the feature space considering two subsets of participants: (a,b) the most conformal participants; (c,d) outliers.

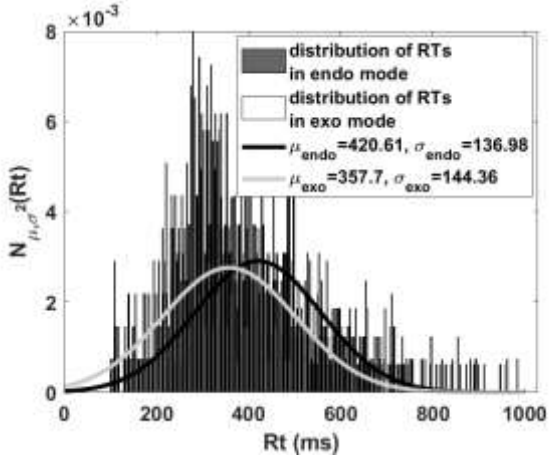
The feature pair distributions in endo and exo modes for the most conforming participants and for outliers are plotted as 2D scatterplots in Figure 7. While the feature distributions overlap strongly for outlier participants (cf. Figure 7c and Figure 7d), for the most conforming participants (Figure 7a and Figure 7b), the two groups of endo and exo points overlap only slightly, regardless of the feature pairs considered (minimum , median) or (10^{th} percentile, mean). Thus, to evaluate the robustness of our approach with respect to the endo/exo classification features, the 10^{th} percentile and the mean have been considered rather than the minimum and the median. As a result of Algorithm 1, 17, 16 and 15 participants were fitted to 2, 4 and 28 models, respectively. Comparing the clustering results derived using the respective pairs of features, we observe that although the ranking of the participants according to their scores can vary slightly, there is a strong agreement regarding the set of the top inliers (one participant out of eight) and the set of outliers (exactly the same group in both cases).

5.3. RT distributions versus participant group

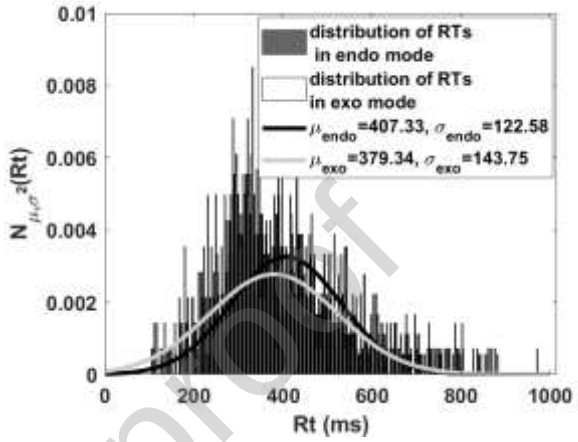
So far, we have proposed an ANN classification model to distinguish the RT measurements performed in either endogenous mode or exogenous mode, and we have used the accuracies of this model to score the participants and cluster them into three groups (top-inliers, intermediate inliers, and outliers). Before using this partition for further analysis, let us specify the benefit of the ANN model. A neural network can be viewed as a function that projects the input values (RT measurements in our case) into the score space where the classes (endogenous and exogenous modes) should be more separable. To evaluate this "more separable" assertion, we analyzed the scores provided by the ANN using either the sigmoid function (called "sigmoid scores") or the ReLU function ("ReLU scores"), with respect to the original RT measurements. In terms of p-values, the three datasets have a p-value less than 0.001, which means that the endogenous and

exogenous modes have statistically significantly different distributions. However, a low p-value only reflects the fact that the hypothesis H0 (homogeneity of the endogenous and exogenous distributions in our case) has to be rejected (statistical significance) and not the magnitude of the difference between the distributions (substantive significance, based on the relative overlap between the two distributions) , which is usually measured by the effect size η^2. Considering the RT values of all participants, the obtained η^2 value is equal to 0.012. Since, according to [8,21], when the effect size is less than 0.2, the difference between the two groups is negligible (even if it is statistically significant), we conclude that the RT distributions were barely separable. However, thanks to the proposed methodology, this η^2 can be greatly improved: Considering the "sigmoid scores" and "ReLU scores" of the inliers, the obtained η^2 values are equal to 0.357 and 0.293, respectively. Then, the η^2 value increases up to a factor of 30 with respect to the all participant RT values. Finally, comparing the "sigmoid scores" and the "ReLU scores", still based on the effect size criterion, we notice that the sigmoid scores allow for a slightly better distinction between the two distributions (endo or exo mode). The proposed algorithm outputs the participant groups, allowing us to categorize participants according to their ability to use the alertness signal to reduce their reaction time. To see if this RT reduction is effective, Figure 8 shows the histograms and fitted normal distributions of RT values in endogenous and exogenous modes for different groups of participants: (a) the most conformal participants, (b) remaining inliers (c) all inliers, (d) outliers. From Figure 8, we clearly notice that the overlap between the two distribution of RT values in exo and in endo modes depends on the group of participants: from Figure 8a to Figure 8d, the difference between the means of RT values in endo and exo modes are equal to 62.9 ms, 28 ms, 45 ms and 15.9 ms, respectively. We recall that the means difference was equal to 20 ms when considering

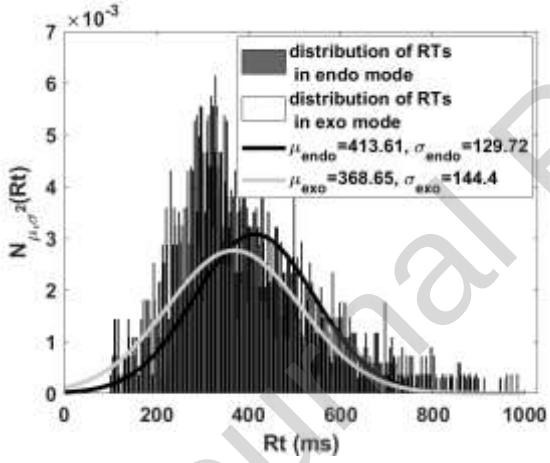
all participants (Figure 3). In addition, we note that the standard deviations difference decreases from 17 ms for all participants to 14.7 ms for all inliers to 7.3 for the most conformal participants.



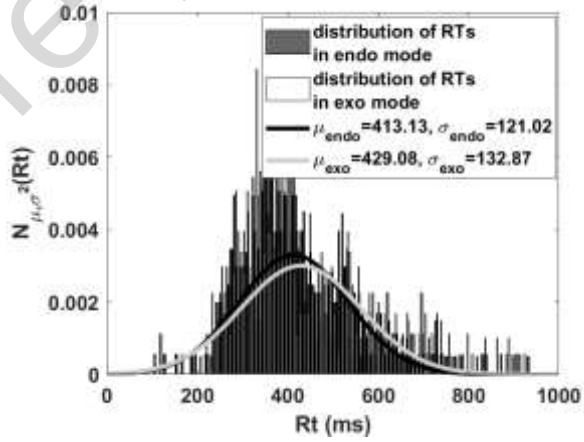
(a) top-inliers; $p<0.001$, $\eta^2=0.024$



(b) intermediate inliers; $p<0.001$, $\eta^2=0.022$



(c) inlier group; $p<0.001$, $\eta^2=0.041$



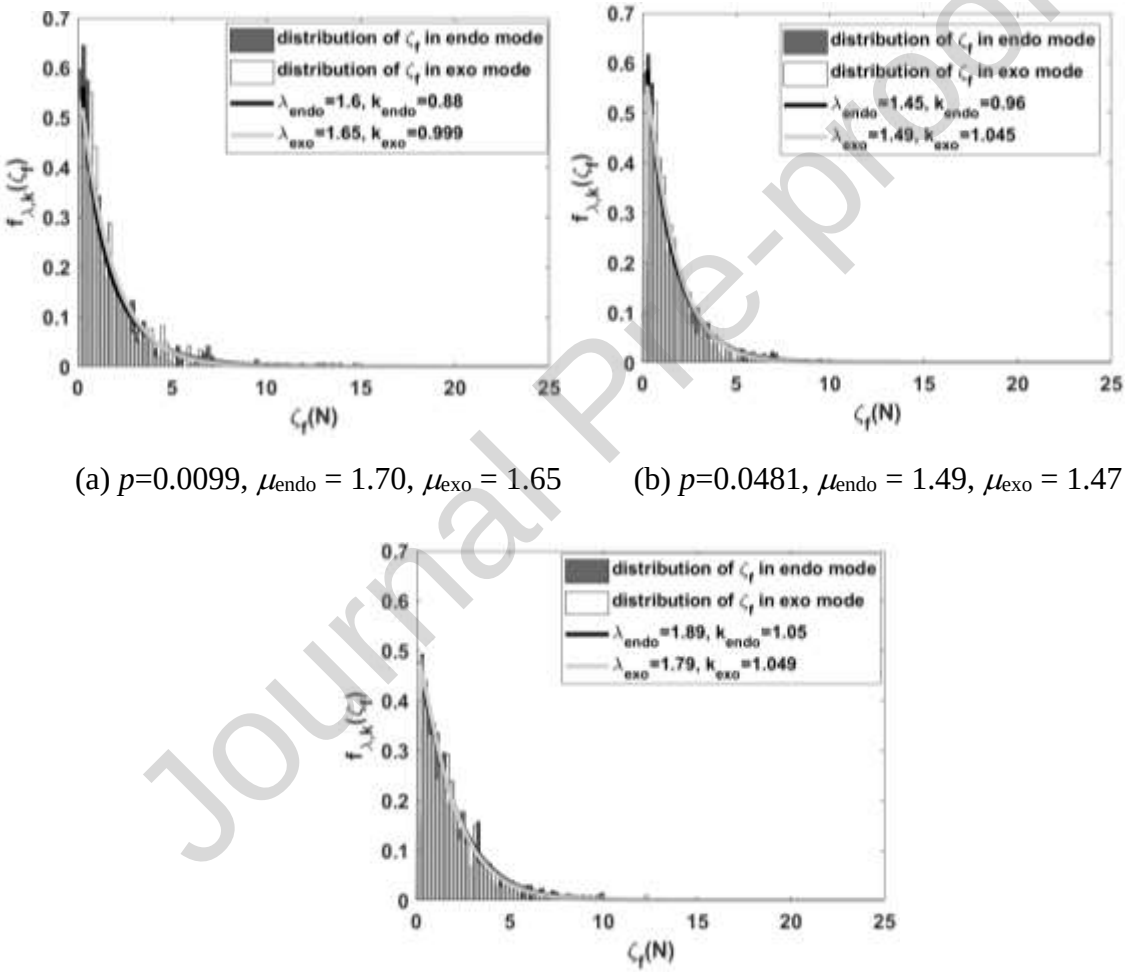
(d) oulier group; $p<0.001$, $\eta^2=0.007$

Figure 8. Distributions of reaction time values for different subsets of participants in endo and exo modes: (a) the most conformal participants {1,12,14,16,22,25,32,28}, (b) remaining inliers {2,5,6,9,24,27,17,30,31,33}, (c) all inliers {1,…,33}, (d) outliers {3,8,10,11,19,20,21,23,26}. For each subset, p-value (p) and effect size (η^2) are given.

5.4. Force error distributions versus participant group

The objective of the force error analysis is as follows. On the basis of Reaction Times (and the proposed RANSAC-ANN method), we were able to distinguish three groups of participants according to their ability to take advantage of the alertness signal (exogenous attention) to reduce

their RT values. Now, we wonder whether the same subjects (participant group) who were able to focus their attention to reduce RT, were also able to reduce the force errors. Depending on the answer to this question, we could either presume that reaction rapidity and force accuracy are two positively related tasks (improving the first will also improve the second) or that they are independent or competitive (improving the first will not affect or even will penalize the second one).



(a) p=0.0099, μ_{endo} = 1.70, μ_{exo} = 1.65 (b) p=0.0481, μ_{endo} = 1.49, μ_{exo} = 1.47

(c) p=0.6275, μ_{endo} = 1.83, μ_{exo} = 1.72

Figure 9. Observed distributions (fitted by a Weibul distribution) of force errors of subsets of participants: (a) the most conformal participants {1,14,16,22,25,32,28}, (b) all inliers {1,…,33}, (c) outliers {3,8,10,11,19,20,21,23,26}.

In this study, the methodology used to answer to the previous question is to examine whether the difference between the force errors observed in the endogenous and exogenous modes depends on the participant group. Specifically, we plotted (cf. Figure 9) the two force error distributions corresponding to the endo and to exo modes, for each of the three participant groups defined on the basis of their ability to reduce RT in the exogenous mode.

Statistical tests were performed on the distributions of the force errors. The hypothesis H0 corresponds to the homogeneity assumption between the distributions of force errors in endogenous and exogenous modes. According to the Kolmogorov-Smirnov test, the hypothesis H0 tested for the top inliers can be rejected with a Type I risk α=1%, since p-value is equal to 0.0099. When considering the whole inlier set, the p-value is 0.048, which is less than 5% and can therefore still be rejected with a reasonable risk. Conversely, in the case of the outlier set, the p-value is 0.6275, indicating that H0 cannot be rejected, i.e. the exo and endo distributions of force errors cannot be considered as inhomogeneous.

6. Discussion

Our study focuses on vigilance, which is essential to our interactions with the environment. We investigate how vigilance or alertness, involves the ability to stay focused and respond quickly to changes or unexpected events. Our study proposes an artificial intelligence-based approach to classify participants according to their behavior and performance during an attentional task of force production. Specifically, our study used an artificial neural network (ANN) model and RANSAC algorithm to identify the participants who behaved as outliers, indicating atypical attentional behaviors. Our approach highlights the promising utility of combining these techniques to explore and quantify human vigilance as well as identifying atypical attentional patterns. Based on our study, we outline the main results below.

6.1.Anticipation hypothesis validation

Firstly, based on Figure 8, we note that the ranking of the participants provided by Algorithm 3 is effective since the "most conformal" participants significantly reduced their RT in exogeneous mode with respect to endogeneous mode which is an expected outcome, in line with the literature ([20]). Then, the hypothesis regarding anticipation appears as a valid hypothesis provided that the participants of the group show a consistent behavior (inlier participants). Specifically, our study distinguishes participants who are able to use the alert signal to enhance their vigilance and improve their reaction times. Additionally, our methodology enabled the isolation of some participants who disregarded the alert signal and exhibited suboptimal attentional behavior, as evidenced by their RT values being higher in exogenous mode than in endogenous mode. For this specific group of participants, we assumed that they either did not rigorously follow the instructions, or the repetition of trials in the cognitive task might have induced a state of mind wandering, consequently leading to a decrease level of attention. Finally, we propose to interpret the decrease of the standard deviations difference as an index of performance or vigilance stability; this metric is particularly more efficient among groups with higher performance levels.

6.2. Error force analysis

Concerning the distributions of the force errors, we assume that there is a part of the force error that is due to imprecision in the control of the force, and that is common to most of the measurements. Subsequently, the interesting part (for our study) occurs when the force error can be attributed to an attention deficit, probably due to a focus on the answer speed, i.e. when the force error is relatively important. To validate this result, we have performed Kolmogorov statistical test on force error distributions. Firstly, as with the RT analysis, Figure 9 confirms the relevance of the participant clustering for the analysis of attentional behavior, as it reveals varied

behaviors across different groups. However, in contrast to the RT case where such a relevance could be expected since the clustering criterion involves RT values, the force error data were not considered in the clustering process at all, rendering our conclusions even more intriguing.

Then, the inlier group behavior (Figure 9a and Figure 9b) shows that the distributions of force errors in endogenous and exogenous modes are significantly different (p-value lower than 1% or 5% depending if we consider the top-inlier or inlier group) when the participants are able to consider the alertness signal. Concerning the outlier participants (Figure 9c), they use neither the alert signal nor the fore-period to improve their precision. Their force errors are similar in both endogenous and exogenous modes (p-value > 0.05).

Now, upon examining the Weibul distribution parameters and the means values (μ_{endo} and μ_{exo}), we observe that all groups exhibit a reduction in force error in the exogenous mode compared to the endogenous mode, that shows that the alertness signal is used efficiently. However, the reduction in the (top) inlier group is lower than in the outlier group. Therefore, we infer that when the two modes significantly differ in terms of force errors, i.e. in the case of the top-inlier and inlier groups, the two tasks are competitive since the reduction in RT comes at the cost of reducing force error. In other words, the motor response and its accuracy are optimized when subjects prioritize the speed (Reaction Time) at the expense of the force level, as already observed by [19].

7. Conclusion

In this work, we proposed a methodological approach to estimate the vigilance ability of participants during a cognitive motor task that is either voluntary (endogenous mode) or guided by an external signal (exogenous mode). It consists of distinguishing between participants who use the alertness signal to reduce their reaction time and participants who react similarly in the presence or absence of the alertness. Given a small dataset, our basic idea was to couple the

RANSAC algorithm with a shallow neural network to handle noisy labels that impact the training process for endogeneous/exogenous classification. Although our clustering results were validated through the distributions of reaction-times and error forces with respect to clusters of participants, future work will focus on more sophisticated testing of the approach on simulated data presenting noisy labels. Another future perspective of our research involves fusing reaction time data with functional imaging data to characterize the influence of vigilance states on the endogenous and exogenous neural networks of attention.

References

[1] A. E. Alchalabi, S. Shirmohammadi, A. N. Eddin, M. Elsharnouby, Focus: Detecting ADHD patients by an eeg-based serious game, IEEE Transactions on Instrumentation and Measurement 67 (7) (2018) 1512–1520.

[2] B. Alipanahi, A. Delong, M. T. Weirauch, B. J. Frey, Predicting the sequence specificities of dna-and rna-binding proteins by deep learning, Nature biotech- nology 33 (8) (2015) 831–838.

[3] V. Andrearczyk, V. Oreiller, A. Depeursinge, Head and Neck Tumor Segmen- tation: First Challenge, HECKTOR 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings, vol. 12603, Springer Nature, 2021.

[4] P. Bashivan, I. Rish, M. Yeasin, N. Codella, Learning representations from eeg with deep recurrent-convolutional neural networks, arXiv preprint arXiv:1511.06448 (2015).

[5] S. Bioulac, D. Purper-Ouakil, T. Ros, H. Blasco-Fontecilla, M. Prats, L. Mayaud, D. Brandeis, Personalized at-home neurofeedback compared with long-acting methylphenidate in an european non-inferiority randomized trial in children with ADHD, BMC psychiatry 19 (2019) 1–13.

[6] F. Blume, J. Hudak, T. Dresler, A.-C. Ehlis, J. Kuhnhausen, T. J. Renner, C. Gawrilow, Nirs-based neurofeedback training in a virtual reality classroom for children with attention-deficit/hyperactivity disorder: study protocol for a randomized controlled trial, Trials 18 (1) (2017) 1–16.

[7] S. Choi, T. Kim, W. Yu, Performance evaluation of ransac family, Journal of Computer Vision 24 (3) (1997) 271–300.

[8] J. Cohen, Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum, 1988.

[9] S. Daly, J. Thai, C. Belkhiria, C. Langley, A. Le Blanche, G. de Marco, Temporal deployment of attention by mental training: an fmri study, Cognitive, Affective, & Behavioral Neuroscience 20 (2020) 669–683.

[10] V. Delvigne, H. Wannous, T. Dutoit, L. Ris, J.-P. Vandeborre, Phydaa: Physiological dataset assessing attention, IEEE Transactions on Circuits and Systems for Video Technology 32 (5) (2021) 2612–2623.

[11] H. Eom, K. Kim, S. Lee, Y.-J. Hong, J. Heo, J.-J. Kim, E. Kim, Development of virtual reality continuous performance test utilizing social cues for children and adolescents with attention-deficit/hyperactivity disorder, Cyberpsychology, Behavior, and Social Networking 22 (3) (2019) 198–204.

[12] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (6) (1981) 381–395.

[13] K. Fukushima, Visual feature extraction by a multilayered network of analog threshold elements, IEEE Transactions on Systems Science and Cybernetics 5 (4) (1969) 322–333, doi: 10.1109/TSSC.1969.300225.

[14] Z. Gao, X. Wang, Y. Yang, C. Mu, Q. Cai, W. Dang, S. Zuo, EEG-based spatio– temporal convolutional neural network for driver fatigue evaluation, IEEE transactions on neural networks and learning systems 30 (9) (2019) 2755–2763.

[15] E. Gawehn, J. A. Hiss, G. Schneider, Deep learning in drug discovery, Molecular informatics 35 (1) (2016) 3–14.

[16] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems 25 (2012).

[17] M. A. Lawrence, R. M. Klein, Isolating exogenous and endogenous modes of temporal attention., Journal of Experimental Psychology: General 142 (2) (2013) 560.

[18] M. D. Liechti, L. Valko, U. C. Muller, M. Dohnert, R. Drechsler, H.-C. Steinhausen, D. Brandeis, Diagnostic value of resting electroencephalogram in attention-deficit/hyperactivity disorder across the lifespan, Brain topography 26 (2013) 135–151.

[19] F. Maquestiaux, M. Lagüe-Beauvais, E. Ruthruff, A. Hartley, L. Bherer, Learning to bypass the central bottleneck: declining automaticity with advancing age., Psychology and Aging 25 (1) (2010) 177.

[20] C. McCormick, R. Redden, A. Hurst, R. Klein, On the selection of endogenous and exogenous signals, Royal Society Open Science 6 (11) (2019) 190134.

[21] J. Cohen, A. P. Primer, Quantitative methods in psychology, Psychological bulletin 112 (1) (1992) 155–159.

[22] S. Ruder, An overview of multi-task learning in deep neural networks, arXiv preprint arXiv:1706.05098.

[23] N. C. Schwertman, M. A. Owens, R. Adnan, A simple more general boxplot method for identifying outliers, Computational statistics & data analysis 47 (1) (2004) 165–174.

[24] O. Sener, V. Koltun, Multi-task learning as multi-objective optimization, Advances in neural information processing systems 31.

[25] Y. Tan, D. Zhu, H. Gao, T.-W. Lin, H.-K. Wu, S.-C. Yeh, T.-Y. Hsu, Virtual classroom: An adhd assessment and diagnosis system based on virtual reality, in: 2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS), IEEE, 2019, pp. 203–208.

[26] J. W. Tukey, et al., Exploratory data analysis, vol. 2, Reading, MA, 1977.

[27] C. Vidaurre, N. Kr¨amer, B. Blankertz, A. Schl¨ogl, Time domain parameters as a feature for eeg-based brain–computer interfaces, Neural Networks 22 (9) (2009) 1313–1319.

[28] W.-L. Zheng, K. Gao, G. Li, W. Liu, C. Liu, J.-Q. Liu, G. Wang, B.-L. Lu, Vigilance estimation using a wearable eog device in real driving environment, IEEE transactions on intelligent transportation systems 21 (1) (2019) 170–184.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Highlights

- Atypical attentional patterns are detected in young subjects.

- Score disentangles high (inliers) and low (outliers) attention performance.
- Score combines RANSAC and classifiers for outlier handling in few-shot learning.
- Score-based subject clustering clarifies RT and force accuracy duality.