DOI: 10.1051/itmconf/20171502002

Classification of user performance in the Ruff Figural Fluency Test based on eye-tracking features

Magdalena Borys[1,*], *Sara* Barakate[2], *Karim* Hachmoud[2], *Małgorzata* Plechawska-Wójcik[1], *Paweł* Krukow[3], and *Marek* Kamiński[1]

[1] Lublin University of Technology, Faculty of Electrical Engineering and Computer Science, Institute of Computer Science, Nadbystrzycka 38D, 20-618 Lublin, Poland

[2] Abdelmalek Essaâdi University, Polydisciplinary Faculty, Martil Road, Tétouan, Morocco

[3] Medical University of Lublin, Department of Clinical Neuropsychiatry, Głuska 1, 20-439 Lublin, Poland

Abstract. Cognitive assessment in neurological diseases represents a relevant topic due to its diagnostic significance in detecting disease, but also in assessing progress of the treatment. Computer-based tests provide objective and accurate cognitive skills and capacity measures. The Ruff Figural Fluency Test (RFFT) provides information about non-verbal capacity for initiation, planning, and divergent reasoning. The traditional paper form of the test was transformed into a computer application and examined. The RFFT was applied in an experiment performed among 70 male students to assess their cognitive performance in the laboratory environment. Each student was examined in three sequential series. Besides the students' performances measured by using in app keylogging, the eye-tracking data obtained by non-invasive video-based oculography were gathered, from which several features were extracted. Eye-tracking features combined with performance measures (a total number of designs and/or error ratio) were applied in machine learning classification. Various classification algorithms were applied, and their accuracy, specificity, sensitivity and performance were compared.

1 Introduction

The Ruff Figural Fluency Test (RFFT) was created by Ronald M. Ruff [1] to be a nonverbal analogue of verbal fluency measures. Fluency itself is understood as the ability to apply some strategies to generate nonverbal responses maximising response efficiency and simultaneously minimising response repetition [2]. Such effective repetition needs processes including planning, inhibition, cognitive flexibility, decision-making and self-monitoring [3, 4]. The test proved to be reliable, and the data support the thesis that the performance depended not on sex, but on age and education [5]. It includes a baseline measurement of cognitive function and executive functioning [6].

The RFFT was proved to be related to performance on the Design Fluency Test (DFT) [7]. It also has good test-retest reliability [5] and interrater reliability [2, 8]. Since figural fluency is related to the functioning of the right frontal lobe [9], the RFFT has also proved successful in distinguishing patients with frontal lobe lesions from a group of healthy control patients [10].

Originally the RFFT was designed as a paper and pencil test used to evaluate nonverbal fluency and executive functioning [2, 11]. In this research a computerised version of the RFFT was used. The test group was homogeneous, composed of seventy healthy students aged 18-31 years.

The work in this paper is novel in recognising eye activity patterns for different groups of cognitive performance during the figural fluency test and classifying those groups of cognitive performance based on eye activity patterns.

2 Related work

The popularity of figural fluency tests has increased in recent years. It is more and more often included in cognitive and neuropsychological test batteries [4]. The RFFT proved to be sensitive to the first changes in cognitive functioning in the domain of executive function [3]. Furthermore, the RFFT is useful in searching for early changes in cognitive function among young and middle-aged persons [12].

Hicks et al. [13] developed an oculomotor-driven version of the trail-making tests (TMT) to assess visual attention, psychomotor processing speed, and task-switching.

Poletti at al. [14] adopted a set of neuropsychological tests for eye-tracking (ET) control. These tests evaluated such issues as language, attentional abilities, executive functions and social cognition. This solution was checked on a sample of healthy participants. The authors applied ET in a neuropsychological battery combined with classical ''paper and pencil'' measures. The set of tests was a cognitive measure of working memory

[*] Corresponding author: m.borys@pollub.pl

abilities and global cognitive efficiency. The researchers found significant correlations between "paper and pencil" screening and ET-based neuropsychological measures.

Keller et al. [15] developed a neuropsychological assessment method for patients with Amyotrophic lateral sclerosis (ALS) based on eye movement. The participants (forty-eight ALS patients and thirty-two healthy controls) were tested by two tests (D2-test and Raven's coloured progressive matrices) adapted for eye-tracking control. The results were compared to the classical paper-pencil version and similar results with high correlation were obtained.

Other studies evaluate cognitive functions using measures based on eye movements and deduce a lack of attention [16] or muscular/neuro-muscular fatigue [17]. Performance of the saccadic eye movement was found to be a sensitive marker for even minor impairment of cerebral function in various physiological and pathological conditions, including sleep deprivation [18], early Alzheimer's [19], extra-motor cerebral pathology [13] and cognitive frontal alterations [20, 21] in amyotrophic lateral sclerosis or mild frontotemporal dementia [22] [Boxer et al. 2006] and even prolonged hypobaric hypoxia monitored in climbers [23].

3 Methodology

3.1 Research questions

A total number of designs (or a total number of unique designs) drawn is the indicator of user performance in the RFFT, while the error ratio indicates correctness of performance taking into consideration a number of perseverative errors in relation to a total number of designs. Those two indices are used to asses an individual figural fluency. On the other hand, the eye-tracking measures can support or substitute some of those test indicators as research in attention and cognitive studies shows.

The primary research question that guided the study is stated as follows: "*Is there any relationship between cognitive performance (measured in a total number of designs and/or error ratio) and eye-tracking measures in the RFFT?*".

If there is any relationship between cognitive performance and eye-tracking measures, the secondary research question is: "*Is it possible to classify users with high and low cognitive performance based on eye measures in the RFFT?*".

3.2 Experiment setting

The experiment was conducted in a special testing laboratory. The room was illuminated with standard fluorescent light and outside light was blocked to ensure stable conditions for the duration of the experiments. The average illumination in the room was about 600 Lux.

Eye activity was recorded using screen-based eye tracker Tobii Pro TX300 (Tobii AB, Sweden). The Tobii TX300 performed binocular tracking with the sampling frequency of 300Hz using the dark pupil and corneal reflection method. The gaze was measured by the eye-tracker with the accuracy of 0.5° and precision of 0.01°.

The experiment was created using Tobii Studio 3.2 and a visual stimulus was presented on a separate monitor equipped with an eye tracker (23'' TFT monitor at 60 Hz). The participants were seated while working with the application. The distance between the screen and the subject was in a range from 50 to 80 cm. The distance depended on individual participant preferences (a comfortable position for working with a computer).

All participants were assessed using the same software and hardware settings (Laptop Asus G750JX-T4191H with Intel Core i7-4700HQ and 8GB of RAM).

The RFFT applied in the study was a computerised version developed on the basis of the original "paper and pencil" test. The application was developed in Java Swing and it is operated using a computer mouse.

The developed version of the RFFT covers three different dot configurations and two distractors added to the first, symmetric dot configuration (beyond the basic plain one). Each trial lasts 60 seconds. When the participant has created single design, he had to press the "Next→" button to clear the board. The accomplished designs were presented in the form of miniatures at the left part of the screen.

The computerised RFFT allows to log all user interactions, and thus to register not only the total number of designs or unique designs, but also the execution time for each individual design or part of the test.

3.3 Procedure

The eye tracker was calibrated using a 9-point built-in calibration procedure at the beginning of each session. Once calibrated, the participants were provided with the experimental instructions on the screen, in which they asked to make as many unique designs as possible by connecting the dots stimuli.

Next, the participants completed five parts of the RFFT using the computer application. Each part had different stimulus presentations (dot configurations and distractors). They covered: one part with the basic dot configuration with five equally spaced points on the white screen (without distractors), two parts with basic dot configurations and distractors added and two parts with different dot configurations on the white background (without distractors).

At the beginning of each part, a short trial, consisting of 3 designs, was run to familiarise the participants with the task. After the trial, the proper test was started.

Each part of the test was repeated three times, then another part started.

3.4 Participants

In the experiment, the exclusion criteria included having a history of significant head injury or brain disfunction and currently experiencing psychological or psychiatric problems, prescription medication use, and problems or pain related to movement. After data recording, the further exclusion criteria included having insufficient eye-tracking ratio (less than 90%) or poor eye-tracking data quality (due to eye-glasses or excessive head movement during study).

Seventy healthy participants (male) were recruited among students at the Lublin University of Technology. From the initial group, 9 participants were excluded (1 individual had a psychiatric medical history, 3 individuals had low eye-tracking ratio and 5 individuals had poor data quality). The final group consisted of 26 Computer Science students and 35 Biomedical Engineering students ranging in age from 18 to 31, mean = 21.4 years, std. dev. =1.96. All were right-handed with normal or corrected-to-normal vision.

The study was approved by the Research Ethics Committee of the Lublin University of Technology (Approval ID 4/2015 from 12 November 2015) and all participants received verbal and written information about the study. All participants signed an informed consent. The participation was voluntary and no compensation was offered.

3.5 Signal pre-processing

The eye activities were analysed off-line using Tobii Studio 3.2 and custom programs written in MATLAB. For each test attempt, the tracking ratio were calculated (based on signal lost), the observations with eye tracking ratio lower than 90% were excluded from further analysis.

Fixations and saccades were exported from the Tobii Studio, they were detected using the Velocity-Threshold Identification (I-VT) fixation classification algorithm [24] for average eye selection. The noise was reduced by using the average moving filter with the window size of 3 samples. Gaps shorter than 75 ms were interpolated. The velocity threshold was set to 30°/s and window length to 20ms. The adjacent fixations were merged and fixations shorter than 60 ms were discarded.

Using 3D eye positions and screen gaze points from the eye-tracker, the amplitude of saccades was calculated in degrees (based on its total distance along the trajectory of the movement between start and end points). The saccades with amplitude greater than 10 degrees were classified as large ones. Also the blinks were identified as zero data (both eyes are not found) embedded in two saccadic events [25]. To avoid any confusion that the signal loss will be identified as a blink, the blinks detected automatically were manually rechecked using the video annotation technique.

The pupil diameter change was extracted for both eyes. The pupil diameter change was normalised with baseline subtraction (subtracting the mean pupil size over the first 100ms of recording) [26].

3.6 Measured feature sets

Based on the log file from the RFFT application the following features were extracted for each part of the test:
- total number of designs – the number of designs drawn by the participant;
- perseverative errors – repetitions of the same design drawn by the participant;
- error ratio – an index for assessing the participant's ability to minimise repetition while maximising unique productions calculated as percentage of errors in the number of all designs.

A variety of ET features inspired by the literature were gathered and some others were proposed. The eye-tracking features were extracted on the basis of the following eye movements and activities:
- fixations (total number of fixations, mean fixation duration, standard deviation of fixation duration, max fixation duration);
- saccades (total number of saccades, total number of small saccades, total number of large saccades, mean saccade duration, standard deviation of saccade duration, max saccade duration, mean saccade amplitude, standard deviation of saccade amplitude, max saccade amplitude);
- blinks (total number of blinks, mean blink duration, standard deviation of blink duration, max blink duration);
- pupillary response (mean left/right pupil size change, standard deviation of left/right pupil size change, max of left/right pupil size change).

All features based on duration are determined in milliseconds and based on amplitude values are in degrees.

3.7 Methods

One-way Analysis of Variance (ANOVA) [27] was applied to analyse potential differences in a continuous random variable between and among the groups. The ANOVA test was, afterwards, followed by the post-hoc test using Scheffe's method. The Kruskal-Wallis test by ranks was used as the equivalent of ANOVA for a variable which does not follow normal distribution. In this work, those methods were used in STATISTICA software.

Hierarchical Agglomerative Clusters (HAC) method [28] was used to identify the set of observations (groups) that are similar and then to differentiate between them. HAC is a bottom-up method that creates a dendrogram in the form of a tree of k-block set partitions where $1 \leq k \leq n$ and n are the number of points to be clustered. The algorithm begins with each point (observation) separately in its own cluster, and the closest pairs of clusters are joined together, resulting in reducing the k number of clusters by 1. The algorithm is repeated until k is equal to 1 and the dendrogram is formed. In this work, the HAC implementation used in Tanagra software [29] was employed, where the clusters are created using a hybrid clustering method [30].

To determine the predictors, two methods were applied: discriminant function analysis and chi-square test. Discriminant Function Analysis (DA) [31] is dedicated to identify the continuous variables best discriminating between the naturally occurring groups. In DA, the groups represent the dependent variables and the predictors are the independent variables. DA might be applied if groups are known a priori. DA is usually applied to determine if a given set of variables could be used in the task of category membership prediction. The chi-square test was also used to determine significant perdictors. This test is commonly used for testing the independence between two categorical variables, so it detects the existence of a relationship between the independent variables and the dependent variables.

Several supervised learning classification methods were used to create a model for predicting high- and low performance in the RFFT using ET features. The best results were obtained for Support Vector Machines, K-Nearest Neighbours and Bagged Decision Trees. To validate the accuracy of models the data set was split into training/testing (80%) and validating (20%) subsets based on the randomise assignment. Firstly, the model was trained and tested using 5 cross-validation on the training/testing subset, then the model was evaluated on the basis of the validating subset. The accuracy and other properties of the models were calculated on their performance on the validating subset.

Support Vector Machines (SVM) [32] is a supervised learning classification method using an appropriately designated discriminant hyperplane. If learning sub-sets are fully separable, the SVM idea is to find two parallel hyperplanes which delimit the wider area and do not contain any probe elements. Those hyperplanes are based on the probe elements called support vectors The idea of the method is to find a maximum-margin hyperplane in a transformed input space, which splits the classes and maximises the distance between them. The method deploys a quadratic programming optimisation problem to get the parameters of the hyperplane.

The K-Nearest Neighbours (KNN) algorithm [33] is also a supervised learning classification method. The algorithm consists only of storing the feature vectors in a multidimensional feature space and class labels of the training samples. To classify a new unlabelled vector it assigns the label which is most frequent among the k training samples (k-nearest neighbours in the feature space) nearest to that query point.

Bagged (or bootstrap aggregation) Decision Trees (Bagged Trees) [34] is a supervised learning classification method that combines the results of many decision trees, which reduces the effects of overfitting (like in an individual decision tree) and improves generalisation. Bagged trees use Breiman's random forests algorithm, which is a set of tree predictors in which each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [35].

4 Results

The data evaluation was performed only for the first part of the RFFT test repeated by each participant three times to preserve the homogeneity of tasks and eliminate the possible influence of participant fatigue. Each repetition was treated as a single observation since eye-activity should correspond to general cognitive performance, without individual/intra-individual properties. After the exclusion based on criteria described in 3.3, there was 177 observations in the analysis.

The analysis consisted of three parts. The first part refers to the division of observations into two groups according to the median of a total number of designs. The second part refers to splitting the observations into two clusters based on the error ratio value. The last part also represents the analysis with clusters based on the total number of designs and the error ratio together. In each part, statistical analysis and classification results are presented.

4.1 A total number of designs

A total number of designs is a simple indicator of the performance in the design fluency test. Two groups based on its median were created. Since the median of a total number of design is equal to 18 designs, group 1 represents 85 observations lower than 18 and group 2 represents another 92 observations. The mean total number of designs in group 1 is 13.6 (with std. dev. = 2.6) and group 2 is 20.7 (with std. dev. = 2.3). One-way ANOVA does not show any statistically significant difference between mean values of age or education (in years) in the groups.

Table 1 shows that there is a statistically significant difference between means of five ET features in two groups created on the basis of the median of a total number of designs. While there are statistically significant difference between means of ET features in groups, the ET features can be used as predictors to classified data into groups.

Table 1. ANOVA results for ET features in two groups based on the median of a total number of designs.

Feature	Group 1 M (SD)	Group 2 M (SD)	F-test F(1,175)	p-value
Total number of fixations	147.1 (19.4)	156.0 (20.0)	9.060	0.003
Mean fixation duration	473.3 (77.8)	442.0 (67.4)	8.238	0.005
Total number of blinks	5.1 (5.1)	3.3 (3.5)	8.125	0.005
Total number of saccades	154.4 (23.7)	162.5 (23.5)	5.202	0.024

The discriminant function analysis (tolerance value = 0.05, forward stepwise procedures) with Wilks' Lambda test of significance were used to investigate the best predictors. Two ET features (total number of blinks with p-value<0.00003 and mean fixation duration with p-value<0.0002) were selected as significant predictors. Moreover, using the chi-squared test, additional features were selected as the best predictors: standard deviation of saccade duration (with p-value<0.002), the total number of small saccades (with p-value<0.0025), of fixations, of large saccades, and of saccades.

Combinations of the identified predictors were used in binary classification. To identify a high performance user in the RFFT, group 1 was seen as the negative class and group 2 as the positive class. The best accuracy (Table 2) was obtained for:

- K-NN classifier with k=15 cosine distance metric and equal distance weight for standardised data using as (2) predictors the total number of blinks and mean fixation duration;
- Bagged Trees classifier with 30 decision tree learners and random forest bag ensemble method using as (5) predictors the standard deviation of the saccade duration, the total number of small saccades, of large saccades, of fixations, and of blinks.

Table 2. The statistical measures of the performance of a binary classification.

Classifier	ACC	Sensitivity	SPC	PPV	MCC
KNN	71%	67%	74%	62%	0.4
Bagged Trees	79%	93%	70%	67%	0.62

The model with the Bagged Trees classifier has the accuracy (ACC) of 79% and detects high performance users correctly, as the sensitivity parameter indicated. The precision (positive predictive value, PPV) in both models is low because the validating subset is unbalanced in favour of the negative class (group 1). However, as the specificity (SPC) parameter shows, the Bagger Trees classifier turned out to be worse than the K-NN in classifying a low performance user. The Matthews correlation coefficient (MCC) implied that the Bagged Tress classifier is in general better in the stated binary classification.

4.2 Error ratio

The median for perseverative errors as well for the error ratio is 0, furthermore the perseverative errors are correlated with the total number of design, therefore the error ratio better describes the correctness of performance.

To find the groups (clusters) of similar observations the HAC method was used. The highest jump in the dendrogram indicated as the best clustering in two groups is presented in Table 3.

Table 3. The groups' characterisation using error ratio as similarity measure.

Feature	Group 1	Group 2	Overall
Number of observations	110 (62%)	67 (38%)	177 (100%)
Error ratio M (SD)	0.00 (0.01)	0.10 (0.06)	0.04 (0.06)

ANOVA does not find any significant differences in variables between the groups. Therefore, no further classification was conducted on the basis of the obtained clustering.

4.3 The total number of designs and error ratio

The additional clustering was proposed on the basis of two features: the total number of designs and the error ratio as indicators of performance and its correctness. The HAC method determined that the optimal number of a cluster is 3. The characteristic of those groups is presented in Table 4. In each group both features were significant and explained the created model. The visualisation of the clusters using on the axis two principal component values of principal component analysis (PCA) for a total number of designs and error ratio values is presented in Figure 1. Moreover, Kruskal-Wallis ANOVA does not show any statistically significant difference between mean values of age or education (in years) in the groups.

The size of the groups varies. Taking into consideration the small size of group 2, it was necessary to check whether the population was normally distributed according to all ET features. For features not following normal distribution the Kruskal-Wallis ANOVA was applied replacing one-way ANOVA. The following features have statistically significant difference between mean (or median) values in at least two of three groups: total number of blinks, mean fixation duration, standard deviation of fixation duration, standard deviation of saccade duration, and standard deviation of left and right pupil size change.

Table 4. The groups' characterisation using a total number of designs and error ratio values as similarity measure.

Feature	Group 1	Group 2	Group 3
Number of observations	66 (37.3%)	12 (6.8%)	99 (55.9%)
Total number of designs M (SD)	12.73 (2.26)	21.33 (3.47)	19.85 (2.48)
Error ratio M (SD)	0.04 (0.05)	0.22 (0.05)	0.02 (0.03)

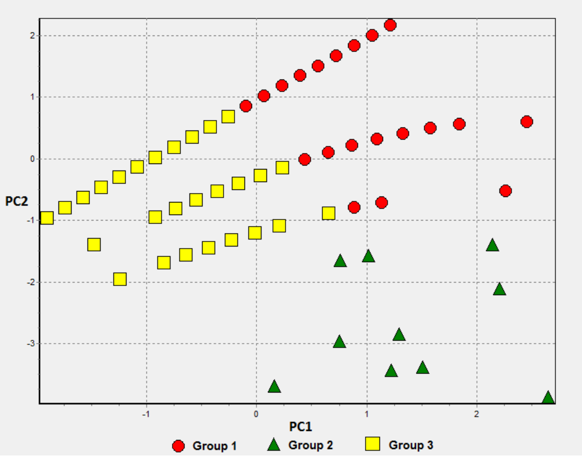

Fig. 1. The visualisation of the clustering groups using two primary components.

The chi-squared test feature selected the following predictors: standard deviation of right pupil size change (with p-value=0.004), standard deviation of saccade duration and a total number of blinks.

The identified predictors from the chi-squared test, features as significantly different between groups from ANOVA and their combinations were used in three-class classification. The group size in the validating subset was unbalanced (11 observations in group 1, 2 in group 2 and 23 in group 3) as results of the HAC cluster size.

The best classifier performances (Table 5) were obtained for:

- SVM (SVM-1) classifier with linear kernel function, kernel scale equals 4, and box constraint level equals 3, the one vs one method and standardised data; with mean fixation duration, standard deviation of saccade duration, standard deviation of the right pupil size change and a total number of blinks as (4) predictors.
- SVM (SVM-2) classifier with linear kernel function, kernel scale and box constraint level equal 4, the one vs one method and standardised data; with mean fixation duration, standard deviation of saccade duration, standard deviation of the left and right pupil size change and a total number of blinks as (5) predictors.

Table 5. The statistical measures of the performance of multiclass classification.

Classifier	ACC	Kappa
SVM-1	63%	0.244
SVM-2	67%	0.308

The classifiers are almost the same, however the additional predictor (standard deviation of left pupil size change) even is strongly correlated with the standard deviation of the right pupil size change allows for a small improvement of the model's accuracy. Both models cannot properly classify observations from group 2 (which are classified as group 3) and show more precision in classifying observations from group 3.

5 Discussion and conclusion

The research questions stated in this work were addressed. The relationship between eye-tracking features and cognitive performance (measured in a total number of designs and/or error ratio) were examined. The extracted features were explored, the best predictors were found and then the classification model was built. The ET features (total number of blinks, mean fixation duration, standard deviation of saccade duration, a total number of small saccades, a total number of fixations, a total number of large saccades, a total number of saccades) proved to be statistically significant and they were used as predictors in order to classify the data into groups.

For binary classification, where groups were created on the basis of the median value of the total number of designs, the accuracy of 79% was obtained for the Bagged Trees classifier with 5 predictors. For multiclass classification, where groups were determined by HAC using the total number of designs and error ratio values, the accuracy was 67% for SVM with linear kernel and 5 predictors. The limit size of the groups was creating a barrier for the proper multiclass training and classification, and mainly for group 2. Among all predictions, the following turned out to be the best: standard deviation of the right pupil size change, standard deviation of saccade duration and a total number of blinks. Further detailed investigation about the multiclass classification will only be achievable after increasing the size of the dataset.

The authors would like to thank all the students who participated in the experiment. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. R. M. Ruff, *Ruff Figural Fluency Test professional manual* (Psychological Assessment Resources Inc., Odessa, FL, 1988)
2. P. S. Foster, J. B. Williamson, D. W. Harrison, *Arch. Clin. Neuropsychol.* **20**, 427–434 (2005)
3. M. E. A. van Eersel, H. Joosten, J. Koerts, R. T. Gansevoort, J. P. J. Slaets, G. J. Izaks, *PLoS One*. **10**, e0121411 (2015)
4. J. S. Kuiper, R. C. O. Voshaar, F. E. A. Verhoeven, S. U. Zuidema, N. Smidt, *BMC Psychol.* **5**, 15 (2017)
5. R. M. Ruff, R. H. Light, R. W. Evans, *Dev. Neuropsychol.* **3**, 37–51 (1987)
6. G. J. Izaks, H. Joosten, J. Koerts, R. T. Gansevoort, J. P. Slaets, *PLoS One*. **6**, e17045 (2011)
7. G. J. Demakis, D. W. Harrison, *Psychol. Rep.* **81**, 443–448 (1997)
8. L. C. Berning, N. C. Weed, M. S. Aloia, *Assessment*. **5**, 181–186 (1998)
9. M. Regard, E. Strauss, P. Knapp, *Percept. Mot. Skills*. **55**, 839–844 (1982)
10. J. V Baldo, A. P. Shimamura, D. C. Delis, J. Kramer, E. Kaplan, *J. Int. Neuropsychol. Soc.* **7**, 586–596 (2001)
11. E. Łojek, J. Stańczak, *Pol. standaryzacja i Norm. Pod. [The Ruff Fig. Fluen. Test. Polish Standarisation Norm. Warszawa Prac. Testów Psychol. Pol. Tow. Psychol.* (2005)
12. P. Krukow, M. Harciarek, J. Morylowska-Topolska, H. Karakuła-Juchnowicz, K. Jonak, *Cogn. Neuropsychiatry*. **22**, 391–406 (2017)
13. S. L. Hicks, R. Sharma, A. N. Khan, C. M. Berna, A. Waldecker, K. Talbot, C. Kennard, M. R. Turner, *PLoS One*. **8**, 8–13 (2013)
14. B. Poletti, L. Carelli, F. Solca, A. Lafronza, E. Pedroli, A. Faini, S. Zago, N. Ticozzi, A. Ciammola, C. Morelli, P. Meriggi, P. Cipresso, D. Lul?, A. C. Ludolph, G. Riva, *et al.*, *Neurol. Sci.* **38**, 595–603 (2017)
15. J. Keller, M. Gorges, H. T. Horn, H. E. A. Aho-Özhan, E. H. Pinkhardt, I. Uttner, J. Kassubek, A. C. Ludolph, D. Lulé, *J. Neurol.* **262**, 1918–1926 (2015)
16. B. Gaymard, C. J. Ploner, S. Rivaud, A. I. Vermersch, C. Pierrot-Deseilligny, *Exp. Brain Res.* **123**, 159–163 (1998)
17. J. J. S. Barton, A. Jama, J. A. Sharpe, *Neurology*. **45**, 2065–2072 (1995)
18. P.-A. Fransson, M. Patel, M. Magnusson, S.

Berg, P. Almbladh, S. Gomez, *J. Vestib. Res.* **18**, 209–222 (2008)

19. Q. Yang, T. Wang, N. Su, S. Xiao, Z. Kapoula, *Age (Omaha).* **35**, 1287–1298 (2013)
20. C. Donaghy, R. Pinnock, S. Abrahams, C. Cardwell, O. Hardiman, V. Patterson, R. C. McGivern, J. M. Gibson, *J. Neurol.* **256**, 420–426 (2009)
21. M. Gorges, H.-P. Müller, D. Lulé, K. Del Tredici, J. Brettschneider, J. Keller, K. Pfandl, A. C. Ludolph, J. Kassubek, E. H. Pinkhardt, *PLoS One*. **10**, e0142546 (2015)
22. A. L. Boxer, S. Garbutt, K. P. Rankin, J. Hellmuth, J. Neuhaus, B. L. Miller, S. G. Lisberger, *J. Neurosci.* **26**, 6354–6363 (2006)
23. T. M. Merz, M. M. Bosch, D. Barthelmes, J. Pichler, U. Hefti, K. U. Schmitt, K. E. Bloch, O. D. Schoch, T. Hess, A. J. Turk, U. Schwarz, *Eur. J. Appl. Physiol.* **113**, 2025–2037 (2013)
24. A. Olsen, "The Tobii I-VT Fixation Filter: Algorithm description" (2012), (available at http://www.tobii.com/)
25. K. Holmqvist, M. Nystrom, R. Andersson, R. Dewhurst, H. Jarodzka, J. van de Weijer, *Eye Tracking. A comprehensive guide to methods and measures* (Oxford University Press, 2011)
26. H. K. Wong, J. Epps, *Comput. Methods Programs Biomed.* **137**, 47–63 (2016)
27. M. G. Larson, *Circulation*. **117**, 115–121 (2008)
28. I. Davidson, S. S. Ravi, *9th Eur. Conf. Princ. Pract. Knowl. Discov. Databases, PKDD 2005*, 59–70 (2005)
29. R. Rakotomalala, *Proc. EGC*. **2**, 697–702 (2005)
30. M. A. Wong, *J. Am. Stat. Assoc.* **77 (380)**, 841–847 (1982)
31. J. Poulsen, A. French, *J. Forensic Sci.* **56**, 297–301 (1996)
32. A. Shmilovici, in *Data Mining and Knowledge Discovery Handbook*, O. Maimon, L. Rokach, Eds. (Springer New York, Second Edi., 2010), pp. 231–247
33. T. M. Mitchell, others, Machine learning. *McGraw Hill Ser. Comput. Sci.* (1997), pp. I--XVII
34. L. Breiman, *Mach. Learn.* **24**, 123–140 (1996)
35. L. Breiman, *Mach. Learn.* **45**, 5–32 (2001)