Performance analysis of chess players comparing traditional and novel cognitive perception ranking measures

Laercio Silva Junior
Department of Electrical Engineering
Centro Universitário FEI
São Bernardo do Campo, Brazil
Email: laercio.silva@fei.edu.br

Wilson da Silva
Centro Universitário UniDomBosco
Curitiba, Brazil
Email: wilsilva@icloud.com

Carlos Thomaz
Department of Electrical Engineering
Centro Universitário FEI
São Bernardo do Campo, Brazil
Email: cet@fei.edu.br

***Abstract*—Traditionally, proficiency in chess has been measured by metrics based on accuracy and response time or performance in tournaments, but not considering how cognitive signals influence in the decision making in this complex game. In this work, we have carried out a performance analysis of chess players comparing a standard ranking measure with a novel one proposed here. Using the idea of treating participants eye movements and brain signals, when answering several on-screen valid chess questions of distinguished complexities, as high-dimensional data we have shown that expertise is consistently associated with the ability to process visual information holistically using fewer fixations rather than locally focusing on individual pieces. Results show that traditional metric to quantify proficiency presented accuracy up to 73,3%, while the proposed cognitive one reached accuracy up to 87,5% and 98,9% for the electroencephalography and eye movements, respectively. These findings might disclose new insights for teaching and predicting chess skills.**

I. Introduction

CHESS game has attracted the interest of many academic works in distinct areas of the scientific knowledge, due to its fundamental questioning about human reasoning. [1], [2]. In the last years, biosignals like electroencephalography [3]–[5] and eye movements [6]–[8] have been used to describe and explain differences between chess experts and novices depending on their responses to chessboard configurations presented on a computer screen.

In this work, following these earlier investigations, we have moved one step further and explored the potential role of such biosignals' measurements as ranking metrics de facto to analyze the performance of chess players. More specifically, we have implemented a novel chess ranking measure based on eye movements and brain signals in order to compare their results with the traditional metric based on high accuracy and short reaction time. In all experiments carried out, the visual perception measure has ensured the best classification of the participants' expertise on a question-by-question basis.

The remainder of this paper is structured as follows. Section II presents equipment used in the experiments and volunteers who participated in this study. Section III explains the procedure followed here and how brain signals and eye movements were processed. In Section IV is explained the metrics used here to measure proficiency and how to compare them. In sequence, Section V presents example of questions and discuss it. Finally, in Section VI, we conclude this article and open new possibilities for future works.

II. Apparatus and participants

Eye movements were recorded with a Tobii TX300 equipment that comprises an eye tracker unit integrated to the lower part of a 23in TFT monitor. The eye tracker performs binocular tracking at the data sampling rate of 300Hz, and has minimum fixation duration of 60ms and maximum dispersion threshold of 0.5 degrees. These are the eye tracker defaults for cognitive research. Data collection were performed with the use of the Tobii Studio programming package running on an attached notebook (Core i7, 16Gb RAM and Windows 7).

Brain electrical signals were obtained through an EEG device, OpenBCI. This is an open-source equipment, it has sampling frequency of 125Hz and resolution of 32 bits per channel. In this experiment, we have used an EEG headcap to acquire 16 channels according to the 10-20 conventional system, which was used as reference for the positioning of the electrodes **??**. In this study, electrodes Fp1, Fp2, F3, F4, F7, F8, T3, T4, T5, T6, C3, C4, P3, P4, O1 e O2 were chosen to cover the maximum scalp area. Figure shows the electrodes' positioning.

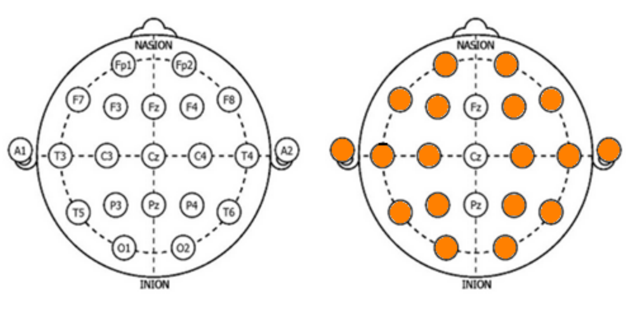

Fig. 1. Electrodes' positioning.

A total number of 32 participants, aged from 6 to 54 years, were involved in this study, with different qualifications: 4 professional chess players with ELO rating [11], 4 chess teachers, 4 school children who commonly compete in national chess championships, and 20 participants who know chess but not practice it regularly. All subjects participated on a voluntary basis and written informed consent was obtained from all participants.

III. Experimental design and data processing

The experimental design consists of 50 on-screen questions related to full size (8x8) and valid chessboard configurations divided into 5 categories, composed of 10 questions each, of distinguished complexities [3], [12]: Object recognition (C1), where participants had to simply respond whether or not, for example, a specified chess piece was on the board; Check detection (C2), where participants should respond whether or not a specified king was in check; Checkmate judgment (C3), where participants should determine whether or not a specified king was checkmate; Checkmate in one move (C4), in which the decision was whether or not exists a piece on the board that could checkmate a specified king in the next move; Rule retrieval (C5), in which participants analyzed whether or not a simple, single move of a specific piece is valid. The time of presentation of each stimulus was controlled by the participants themselves, but they were instructed to produce their responses as correctly and quickly as possible.

The experiments began with a calibration procedure as implemented in the Tobii apparatus and positioning EEG headcap on volunteers. On each trial, the question to be solved was shown only once before its chessboard diagram, presented afterwards on a new screen. All the stimuli were presented centralized on a black background using the 23in TFT monitor with a screen resolution of 1280x1024 pixels. All chessboard configurations were visualized on the TFT monitor with 800 pixels wide and 800 pixels high. Each question followed by its corresponding chess diagram preceded the next pair of question and chess diagram stimuli until all the 50 trials were presented. Figure **??** presents how the test was performed.

A. EEG data processing

EEG data collected by OpenBCI does not have any prior processing. Due to this, a high pass filter with cut-off frequency of 0.5Hz, a low pass filter with cut-off frequency of 50Hz and a band-stop filter with cut-off frequency of 60Hz were used to pre-process the signal. After the electroencephalogram signals have been filtered, we have processed according to the method proposed by Rocha et al. [14], which investigates the communication among the specialized neural agents involved in cognition, using the EEG data 2 seconds prior to decision-making.

The first step of this summarization is to calculate the linear correlation coefficients among the electrical amplitude values registered by each electrode and all the others. It was adopted here the Pearson correlation for being a parametric

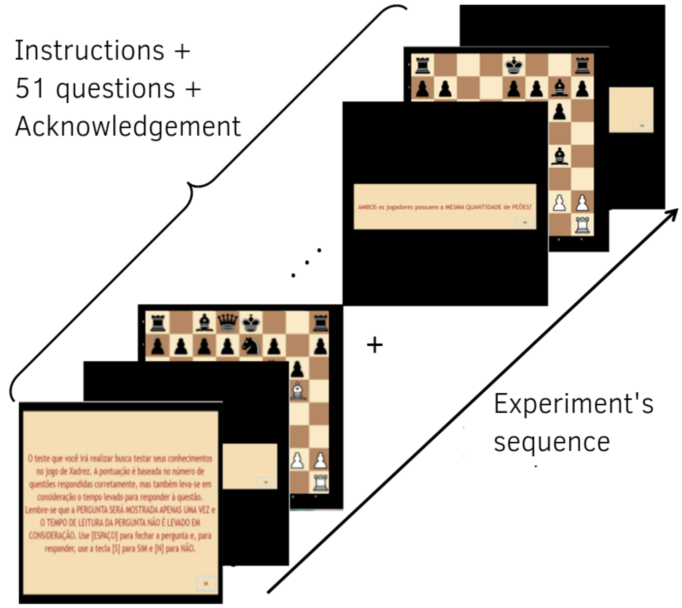

Fig. 2. Question's experiment sequence.

test [15]. The next step was the entropy calculation based on the Shannon entropy formula [16]:

$$h(c_{i,j}) = -c_{i,j}\log_2 c_{i,j} - (1-c_{i,j})\log_2(1-c_{i,j}), \quad (1)$$

where $c_{i,j}$ is the correlation of two distinct channels.

Analogously the entropy of the average correlation of each electrode can be calculated according to the equation (2),

$$h(\overline{c}_i) = -\overline{c}_i\log_2 \overline{c}_i - (1-\overline{c}_i)\log_2(1-\overline{c}_i), \quad (2)$$

where,

$$\overline{c}_i = \frac{1}{(n-1)}\sum_{j=1}^{n-1} c_{i,j}, \quad (3)$$

and n is the number of electrodes. In this case, $n = 16$.

The information provided by each electrode is given by the sum of the differences between the average entropy correlation and the entropy of the electrode with the other channels, that is,

$$h(c_i) = \sum_{j=1}^{n}(h(\overline{c}_i) - h(c_{i,j})). \quad (4)$$

The information provided by each electrode can now be composed as a data matrix $\mathbf{X}$ in a way that the corresponding sampled input data can be treated as a high-dimensional point in a multivariate space.

B. Eye movement data processing

As the output of the software is the heatmap and diagram together in one image, it is necessary to remove the background of each image to analyze only the focus of attention of volunteers. For this, the diagram is subtracted to the heatmap obtaining, then, an image 800x800 with only the focus of attention of the volunteers in RGB. It is not necessary that the information from the heatmap has three dimensions, then this image was transformed to gray-scale to become a two-dimensional image. For this transformation, it was assumed that the points of greatest fixation in red would be transformed to the value of 255 in the gray-scale scale, while the lowest fixation in green color was transformed to the value 64. The other colors varied within the lower and upper limits according to the scale of color adopted for the construction of heatmaps. Figure 3 shows this transformation.

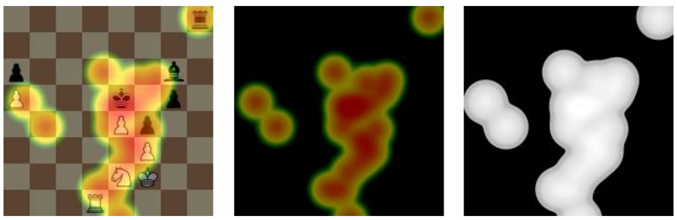

Fig. 3. Heatmap in RGB to grayscale.

IV. Metrics of proficiency evaluation

A. Traditional metric

To quantify a traditional discriminant measure, we have used the general Volke's strategy of high accuracy and short reaction time calculated as [3]:

$$V_s = \left(K_c - \frac{N}{2}\right) \cdot \frac{RT}{RT_s}, \tag{5}$$

where V_s is the performance value of participant s, K_c the participant's number of correct responses, N the total number of pre-defined trials, RT the mean reaction time across all participants and pre-defined trials, and RT_s the participant's mean reaction time across all pre-defined trials.

These performance values served as a discriminant measure of the participants' chess skills on a question-by-question basis, considering only the subjects classified here as belonging to the sample group of novices or experts. We have considered novices and experts all the subjects whose overall Volke's performance values arranged from smallest to largest correspond to the first and fourth statistical quartiles respectively.

B. Cognitive metric

In this work, we want to separate expert and novice groups. For that, we have used the Linear Discriminant Analysis (LDA). To quantify the cognitive characteristics among groups of experts chess players and novices, it is proposed here to treat each vector of brain signal characteristics and each gray-scale heatmap of eye movement characteristics as a pattern in a n-dimensional space, making this a problem of multivariate statistics.

To analyze brain signals, a data matrix is then created with the information from volunteers classified as proficient and novice. The gray-scale heatmap can now be presented as a matrix with two dimensions. For each heatmap of each volunteer to be represented as information each of the lines were sequentially concatenated into a single vector, represented by x = $[x_1,\ x_2,\ ...,\ x_k]$. This vector represents the extraction of characteristics of the eye movements. In this concatenation, it becomes mathematically unfeasible to process due to the amount of data. To decrease the dimensionality, PCA (Principal Component Analysis) was used in which the dimensionality has been reduced to the number of main components, whose self values are non-null [17]. After the application of the PCA, just as it was done with the EEG signals, these data will be treated as a matrix in a multi-dimensional space. as such:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \\ \vdots \\ \mathbf{x_N} \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1n} \\ x_{21} & \dots & x_{2n} \\ \vdots & \vdots & \vdots \\ x_{N1} & \dots & x_{Nn} \end{bmatrix}, \tag{6}$$

where N is the number of volunteers and n is the number of electrodes ($n = 16$) when EEG is analyzed or n is the number main components when eye movements are analyzed.

Since the goal is to create a metric that can measure chess players proficiency through their cognitive signals and obtaining a numerical value as a result, as well as the traditional metric, it is proposed here to use the well-known LDA technique (Linear Discriminant Analysis) to achieve this goal. Let the between-class scatter matrix $\mathbf{S}_b$ be defined as

$$\mathbf{S}_b = \sum_{i=1}^{g} N_i(\overline{\mathbf{x}}_i - \overline{\mathbf{x}})^T(\overline{\mathbf{x}}_i - \overline{\mathbf{x}}), \tag{7}$$

where $\overline{\mathbf{x}}_i$ is the unbiased sample mean, g is the total number of groups (here, $g = 2$), N_i is the number of training patterns of each group i and the grand mean vector $\overline{\mathbf{x}}$ is given by

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{g} N_i\overline{\mathbf{x}}_i = \frac{1}{N}\sum_{i=1}^{g}\sum_{j=1}^{N_i} x_{ij}. \tag{8}$$

Now, let the within-class scatter matrix $\mathbf{S}_w$ be defined as

$$\mathbf{S}_w = \sum_{i=1}^{g}\sum_{j=1}^{N_i}(\mathbf{x}_j - \overline{\mathbf{x}}_i)^T(\mathbf{x}_j - \overline{\mathbf{x}}_i). \tag{9}$$

The main objective of LDA is to find a projection vector L that maximizes the ratio of the determinant of the within-class scatter matrix, described as

$$\mathbf{L} = argmax\frac{\mid \mathbf{P}^T\mathbf{S}_b\mathbf{P} \mid}{\mid \mathbf{P}^T\mathbf{S}_w\mathbf{P} \mid}. \tag{10}$$

where $\mathbf{P}$ is the eigenvector matrix.

However, in limited sample and high dimensional problems, S_w is either singular or mathematically unstable and the standard LDA can not bee used to perform the separating task. [18]. In order to avoid this problem, we have used a maximum uncertainty LDA-based (MLDA) that considers the issue of stabilizing the S_w in Fischer's criterion formula described in Equation (10) with its regularization version [18].

To project all volunteers in the most discriminant EEG and eye-tracking hyperplane, it is subtracted the training data with zero mean from intermediate data matrix and multiplied by the result of MLDA, as it follows:

$$H_s = (\mathbf{x}_i - \overline{\mathbf{x}})\mathbf{L}, \tag{11}$$

where H_s is the most discriminant feature of the information provided by brain signals or eye movements on MLDA space for each volunteer (subject).

C. Metrics comparison

To assess the degree of monotonicity between both discriminant measures, we calculated the standard Spearman's correlation given by

$$\rho = \frac{\sum_{j=1}^{N}(R_j - \bar{R})(Q_j - \bar{Q})}{\sqrt{\sum_{j=1}^{N}(R_j - \bar{R})^2 \sum_{j=1}^{N}(Q_j - \bar{Q})^2}}, \tag{12}$$

where R_j and Q_j represent, to a given on-screen question and chessboard diagram, the rankings of the same set of N paired participants determined respectively by their V_s and H_s discriminant measures described in equations (5) and (11).

V. Results

We adopted a leave-one-out validation method to estimate the classification accuracy of the visual perception discriminant measure. The standard sample group mean of each class has been calculated from the corresponding discriminant scores and the minimum Euclidean distance from each class mean has been used to assign a test sample to either the novices or experts classes. Table I shows the average values of visual perception and EEG accuracy (acc) and shows Spearman correlation (corr) of each category for both cognitive metrics.

TABLE I
Average accuracy of each category and Spearman correlation between the tradition and cognitive metrics.

Cate gory	Std Acc	Visual Perception Acc	Visual Perception Corr	EEG Acc	EEG Corr
C1	0.633	0.62-0.98	0.03	0.47-0.90	0.56
C2	0.800	0.67-0.98	-0.46	0.42-0.88	0.46
C3	0.733	0.61-0.99	-0.22	0.57-0.86	0.62
C4	0.600	0.48-0.99	-0.09	0.48-0.89	0.54
C5	0.900	0.56-0.98	0.04	0.43-0.82	0.42
Mean	0.73	0.58-0.98	-0.141	0.47-0.87	0.518

According to Table I, the visual information that volunteers have used when making their on-screen decisions are highly discriminant (greater than 98%) in all categories. Brain activity is also discriminant, separating groups with minimum of 82.4% in all presented categories. The traditional metric used to quantify proficiency showed the worst accuracy among the metrics presented here, suggesting that the traditional performance measure might be imperfect or incomplete. Table I reveals that Spearman correlation between traditional metric and visual perception and traditional metric between EEG do not present statistically significant results and in some categories correlation is even negative. In order compare those results, one question from Category 3 and one question from Category 4 will be analyzed as examples. Figure 4 shows participants' performance in category C3.

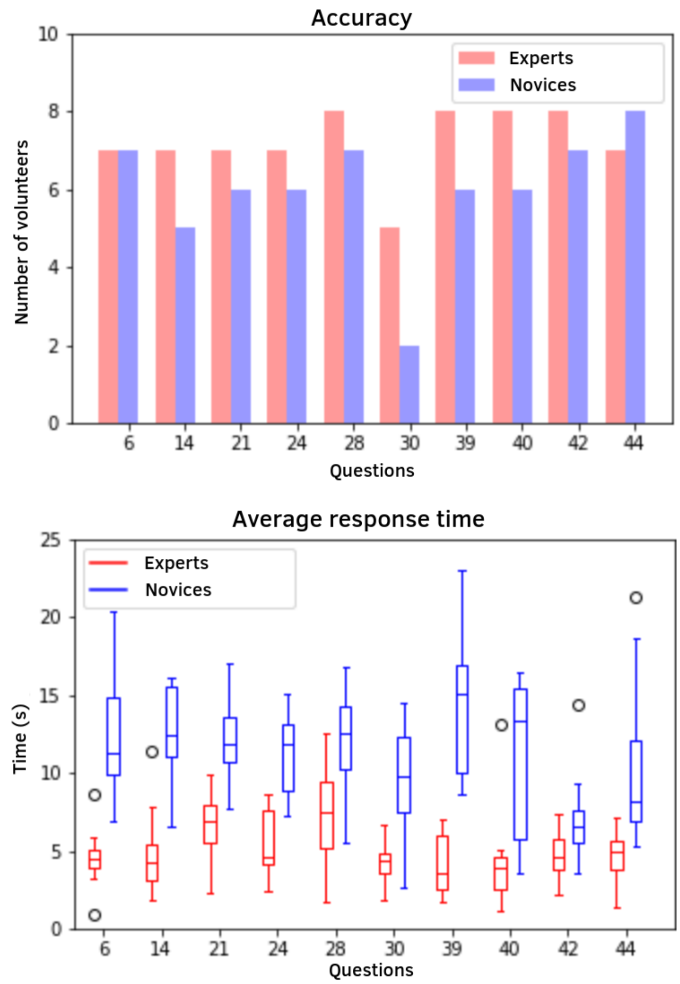

Fig. 4. (Participants' performance in Category 3.

Figure 4 shows that both groups performed well for this category. The expert group had more correct answers in almost all the questions, except the question 44, in which the novice group had one more hit. Comparing the time response between the two groups, it is noted that the novice group with took more time to analyze the questions presented and answer them.

Based on the information provided in Figure 4, EEG and eye-tracking hyperplanes are generated for question 39, in which the most proficient group had two more hits than the other group. Figure 5 shows the results for this question.

The question showed on Figure 5 asks to the participant if the black king is in checkmate. In this way, two conditions are necessary to answer it: the king is threatened by another piece and it cannot escape from this attack. So, one possibility to solve this question is to identify if the black king is threatened,

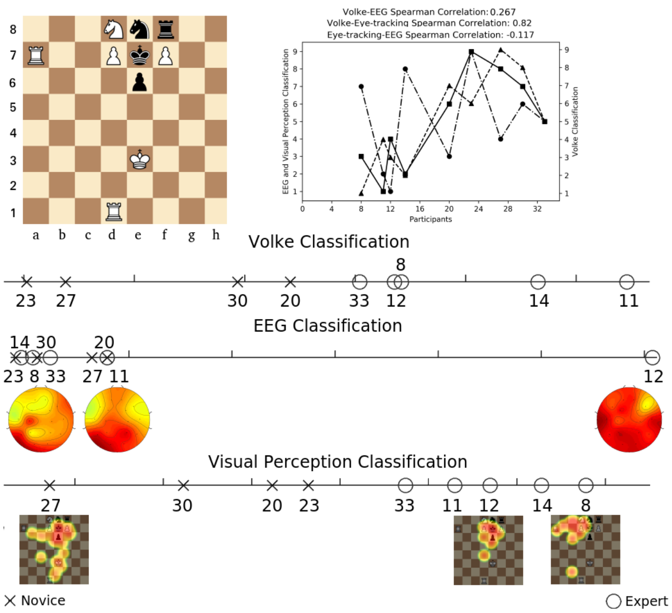

Fig. 5. (Top left) Chessboard diagram of the C3 (one of the easiest categories) on-screen question "Is the BLACK KING in CHECK MATE?". The correct answer is no; (Top right) Spearman's correlation of the degree of monotonicity between the series of paired participants ranked by the Volke, Visual Perception and EEG classifications shown as one-dimensional discriminant axes (in the middle); (Middle) From left to right, corresponding brain activation maps of the participants 8 (expert), 27 (novice) and 12 (expert); (Bottom) From left to right, corresponding spatial attention maps of the participants 27 (novice), 12 (expert), and 8 (expert).

if not, it is not necessary to check the second condition. Volunteers 8 and 12, proficients, analyzed where the king was and squares near to it, consequently they were able to answer this question correctly. On the other hand, volunteer 27, novice, presented fixation in squares that were not relevant to solve the question, such as the squares b7, b6, c6, e5, e4, f4, e2 and c4. These results are consistent with the prediction that the experts define fewer points of fixation to extract relevant information to solve the problem, and these fixations are grouped in the most relevant points of the scene. Eye-tracking hyperplane and traditional metric have separated all volunteers linearly, which explains its high correlation (0,82). EEG hyperplane has not presented a linear separation between groups, where all volunteers were projected near to each, except by volunteer 12. Brain maps show that volunteer 27 and 8 presented more brain activation in the area from electrode O1, related to the primary vision process [19], while volunteer 12 had activation in all areas of his brain, except for T5 electrode area.

Another example is Category 4. Figure 6 shows participants' performance in this category.

This category is considered the most difficult one among the categories that were selected to carry out the experiments. The graphics shown in Figure 6 depict its difficulty, in which only question 32 had 100% accuracy of the expert group while novice group had mostly 50% or less volunteers who answered the questions correctly, in addition to the average response time being higher than in all other categories, indicating that there was greater effort by volunteers to analyze the questions and resolve them.

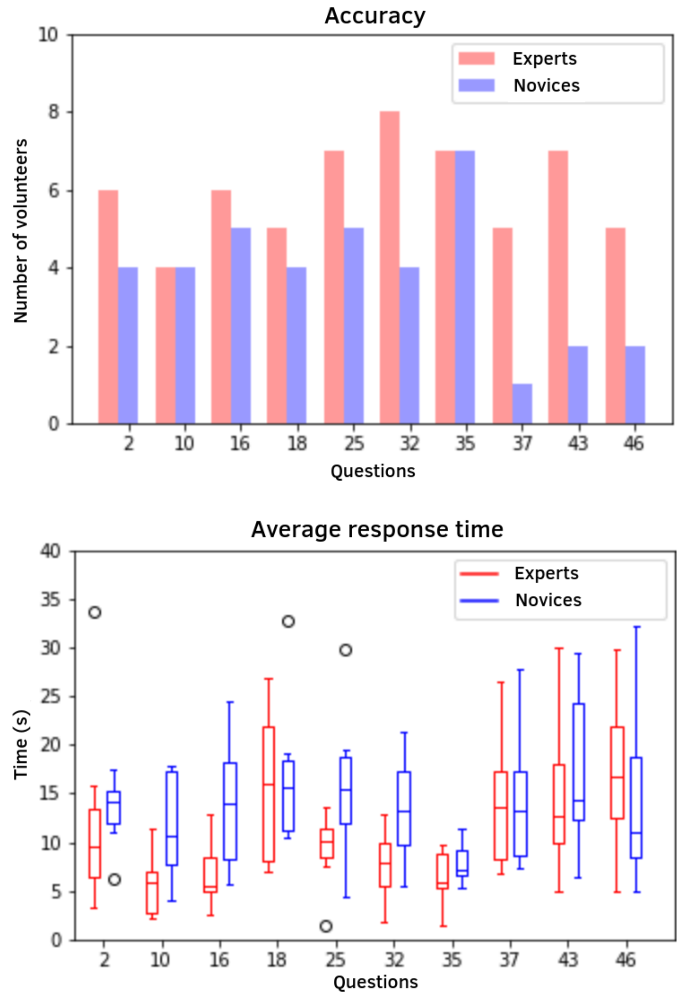

Fig. 6. (Participants' performance in Category 4.

Based on the information provided in Figure 6, EEG and eye-tracking hyperplanes are generated for question 37, in which the accuracy was the least from all the others questions in this experiment.

Figure 7 presents a question that volunteers must analyze how black queen can checkmate the white king. One solution to solve it is by checking squares in which black queen can put the white king in check, those possibilities are: Qxe5, Qxf3, Qc4, Qd4 and Qe4. Heatmap from volunteers 20, 11 and 8 are similar, in which only volunteer 20 has looked to the middle center top part of the board and all of them have looked to most important spaces of the board. Eye-tracking hyperplane separated all volunteers linearly in two groups, but Volke's classification considered volunteers 8, 11 and 12 (experts) as novices and volunteer 30 (novice) as expert. This difference between both hyperplanes explains the low correlation. EEG hyperplane also presented confusion in the classification of volunteers, in which volunteers 6 and

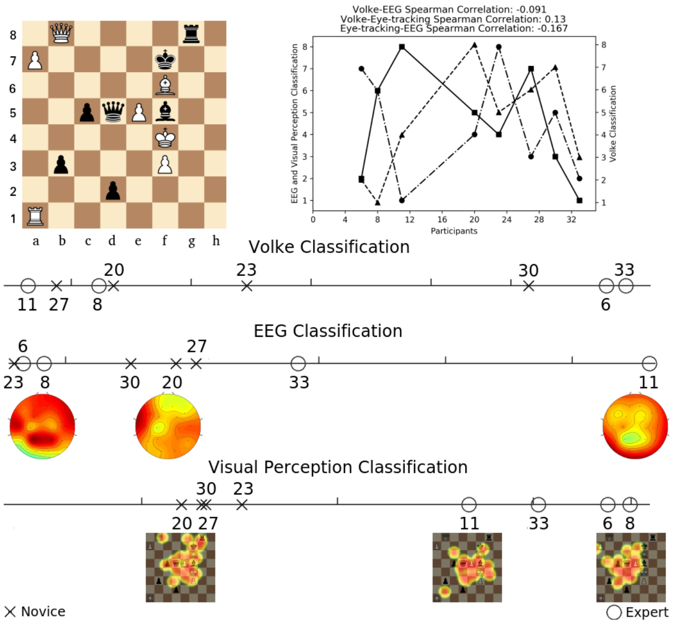

Fig. 7. (Top left) Chessboard diagram of the C4 on-screen question "May the BLACK QUEEN achieve checkmate in ONE move?". The correct answer is no; (Top right) Spearman's correlation of the degree of monotonicity between the series of paired participants ranked by the Volke and Visual Perception classifications shown as one-dimensional discriminant axes (in the middle); (Bottom) From left to right, corresponding spatial attention maps of the participants 29 (novice), 8 (expert), and 33 (expert).

8 (experts) were designed among novices volunteers and volunteer 11 had the best rating on the EEG hyperplane and the worst rating on Volke. This confusion between both metrics explains the low correlation between them. Brain maps reveal that volunteer 20 had his main activation in the area of T3 electrode, associated to Brodmann Area 21, which has the function to produce words and sentences [20], Volunteer 8 had activation predominantly in the occipital region and electrode T6 and volunteer 11 presented his main activation near to the electrodes P3, P4, T3 e T4, parieto-temporal area, related to the lolcalization storage and signals that call a person's attention [21].

VI. Conclusion

This work proposed and implemented a cognitive ranking measure based on eye movements and brain signals to assess the performance of chess players. Using several on-screen questions of distinguished complexities. Our findings revealed that the visual information and brain activity used by the participants on a question-by-question basis when making their decisions are discriminant and consistent regarding their overall performance when compared with the traditional measure of high accuracy and short reaction time. More than that, the traditional metric to quantify proficiency presented accuracy up to 73,3%, while the proposed cognitive one reached accuracy up to 87,5% and 98,9% for the electroencephalography and eye movements, respectively. We believe that the results reveal the potential of cognitive signals to translate and understand better human proficiency in chess, with an emphasis on the most discriminating patterns of eye movements.

As future work, we intend to extend this analysis to other cognitive signals, such as pupil dilation, which explains the cognitive effort and, perhaps, predict whether a person can become or not proficient, not only in chess, but any other domain that cognitive signals are important to this propose.

Acknowledgments

The authors would like to thank FAPESP (2018/13076-9 e 2012/22377-6), CNPq (444964/2014-2) and Royal Society (Newton Advanced Fellowship).

References

[1] A. P. Burgoyne, G. Sala, F. Gobet, B. N. Macnamara, G. Campitelli, and D. Z. Hambrick. "The relationship between cognitive ability and chess skill: A comprehensive meta-analysisIntelligence", *Intelligence* v.59, pp. 72-83, 2016.

[2] Gobet, Fernand, Jean Retschitzki, and Alex de Voogt. "Moves in mind: The psychology of board games", Psychology Press, 2004.

[3] H.-J. Volke, P. Dettmar, P. Richter, M. Rudolf and U. Buhss. "On-coupling and off-coupling of neocortical areas in chess experts and novices", *Journal of Psychophysiology*, v. 16, pp. 23-36, 2002.

[4] J. P. Fuentes-García, et al. "Chess players increase the theta power spectrum when the difficulty of the opponent increases: an EEG study." *International journal of environmental research and public health* 1. 17, n. 1, p.46, 2020.

[5] . P. Fuentes-García, et al. "Differences between high vs. low performance chess players in heart rate variability during chess problems.", *Frontiers in psychology*, v.10, p. 409, 2019.

[6] E.M Reingold and H. Sheridan. "Eye movements and visual expertise in chess and medicine", *Oxford Handbook on Eye Movements*, 2011, chapter 29, pp. 523–550.

[7] H. Sheridan and E.M Reingold. "Expert vs. novice differences in the detection of relevant information during a chess game: evidence from eye movements", *Frontiers in Psychology*, v.5, p. 941, 2014.

[8] H. Sheridan and E.M Reingold. "Chess players' eye movements reveal rapid recognition of complex visual patterns: Evidence from a chess-related visual search task", *Journal of Vision*, v.17, n. 3, p. 4-4, 2017.

[9] M. Teplan et al. "Fundamentals of EEG measurement", *Measurement science review*, v. 2, n. 2, p. 1-11, 2002.

[10] H. H. Jasper, "The ten-twenty electrode system of the International Federation." Electroencephalogr. Clin. Neurophysiol. v.10, p. 370-375, 1958.

[11] A.E. Elo, "The rating of chess players", *Arco Pub.*, 1978.

[12] L.R. Silva-Junior and C.E. Thomaz. "A multivariate correlation assessment of chess proficiency using brain signals", *Proceedings of the XV Workshop de Visao Computacional (WVC)*, Brazil, IEEE CS Press, 2019, pp. 10–15.

[13] M. S. Gazzaniga, "The cognitive neurosciences", *MIT press*, 2009.

[14] F. T. Rocha, et al. "EEG acquisition and processing for cognitive brain mapping during chess problem solving". *IEEE Latin America Transactions*, v. 14, n. 3, p. 1129-1134, 2016.

[15] J. Hoffman and C. Johnson, "Computational turbulent incompressible flow: Applied mathematics: Body and soul 4", *Springer Science & Business Media*, 2007.

[16] C. E.Shannon, "Communication theory of secrecy systems", *The Bell system technical journal*, v. 28, n. 4, p. 656-715, 1949.

[17] K. Fukugana, "Introduction to statistical pattern recognition", Elsevier, 2013.

[18] C. E. Thomaz, E. C. Kitani and D. F. Gillies. "A maximum uncertainty LDA-based approach for limited sample size problems—with application to face recognition", *Journal of the Brazilian Computer Society*, v.12, n. 2, p. 7-18, 2006.

[19] F. H. Martini, Michael J. Timmons, and Robert B. Tallitsch. "Anatomia Humana-: Coleção Martini". *Artmed Editora*, 2009.

[20] A. I. Holodny, "Right Brain Language Function and Academic Honesty", *Brain connectivity*, v.8, n. 6, p. 319-320, 2018.

[21] V. Santangelo and E. Macaluso. ”Visual salience improves spatial working memory via enhanced parieto-temporal functional connectivity.” *Journal of Neuroscience*, v. 33, n. 9, p. 4110-4117, 2013.
[22] R.A. Johnson and D.W. Wichern, *Applied Multivariate Statistical Analysis*, New Jersey: Prentice Hall, 1998.