

# Are eye tracking and skin conductance signals related to performance during cognitive tasks?

Julie Fournier<sup>1,2</sup> Agustina Fragueiro<sup>3</sup> René-Paul Debroize<sup>1</sup> Claire Cury<sup>1</sup>

<sup>1</sup>Univ Rennes, Inria, CNRS, Inserm, IRISA UMR 6074, EMPENN — ERL U 1228, Rennes, France

<sup>2</sup>Institut des Neurosciences Cliniques de Rennes (INCR), Rennes, France

<sup>3</sup>Dipartimento di Psicologia, Università G. d'Annunzio di Chieti-Pescara, Italy

## 1 Introduction

Cognitive load plays a significant role in performance, as it influences task engagement and arousal: both low and very high arousal lead to poor performance. Yet, performance is not easy to evaluate in real time, which makes monitoring cognitive engagement particularly useful in rehabilitation protocols such as neurofeedback, useful to improve learning processes during rehabilitation protocols such as neurofeedback [5]. Different physiological signals have been reported as good indicators of arousal and attention. Among them, eye tracking (measures of eye movements, directly linked with cognitive load) and electrodermal activity (linked with environmental stimulation) represent non-invasive, easy-to-monitor signals. In this study, we investigated the relationship between eye tracking and skin conductance features with performance during the execution of two different cognitive tasks (i.e., Addition and Stroop tasks). The goal is to improve a previously developed model that predicts the performance according to physiological signal [3].

## 2 Methods

**Materials** We implemented two cognitive tasks while recording eye and skin signals. More details on the experiment can be found in [3]. In the Stroop task [4], participants judged whether the ink colour of a top word matched the word written below, across 3 blocks (2 minutes task, 2 minutes rest each). In the Addition task [2], participants added two successive central digits and selected the correct answer among two lateral options, across 3 task/rest blocks of 1 minute each, with trial duration decreasing across blocks (3sec, 2sec, 1sec), increasing the difficulty with each block. In total, 18 volunteers (7 females, aged 20–60) participated after providing informed consent (COERLE/Inria, GDPR compliant, avis ID 2024-04).

**Data acquisition and pre-processing** While participants were performing the tasks, we recorded gaze behavior and electrodermal activity. The Matlab toolbox Ledalab [1] was used to extract the integrated phasic driver activity (ISCR). For gaze behavior, we extracted 4 different metrics: 1/ Pupil diameter (PD) 2/ Fixation duration (FD) 3/ Blink duration (BD) 4/ Gaze divergence (GD). After feature extraction, we removed outliers standing at 3 or more standard deviations from the mean, before normalizing all these features (PD, FD, BD, GD, and ISCR) using a z-score for each participant independently. Performance was computed using accuracy and response time, resulting in a score between 0 and 1 (the higher the better), defined for each trial as follows:

$$Perf = \frac{\text{stimulus duration} - \text{reaction time}}{\text{stimulus duration}} * \text{accuracy}.$$

**Analysis of correlation between performance and physiological signals** Due to sensor malfunction, we excluded ET data for one participant and EDA data for two. To compute correlations between the previously defined features and performance, we averaged data in 10-seconds non-overlapping time windows (TW) for each participant independently, to align with EDA time resolution. This results in 306 ET and 288 EDA observations for the Addition task (3 blocks lasting 1 minute) and 612 ET points and 576 EDA points for the Stroop task (3 blocks lasting 2 minutes). We then computed Spearman's correlations between performance and each one of the five features (PD, FD, BD, GD, and ISCR), independently for each task. To get a better understanding of features' evolution and association with performance, we computed the correlations over the full task, but also for each task block separately. Tests were corrected for multiple comparisons using the Bonferroni correction.

## 3 Results

For the Stroop task, PD is negatively correlated with performance ( $r_s(610) = -0.360, p < 1.0e^{-05}$ ) when considering the full task, and when considering independently the 3 blocks (not significant in block 2). FD

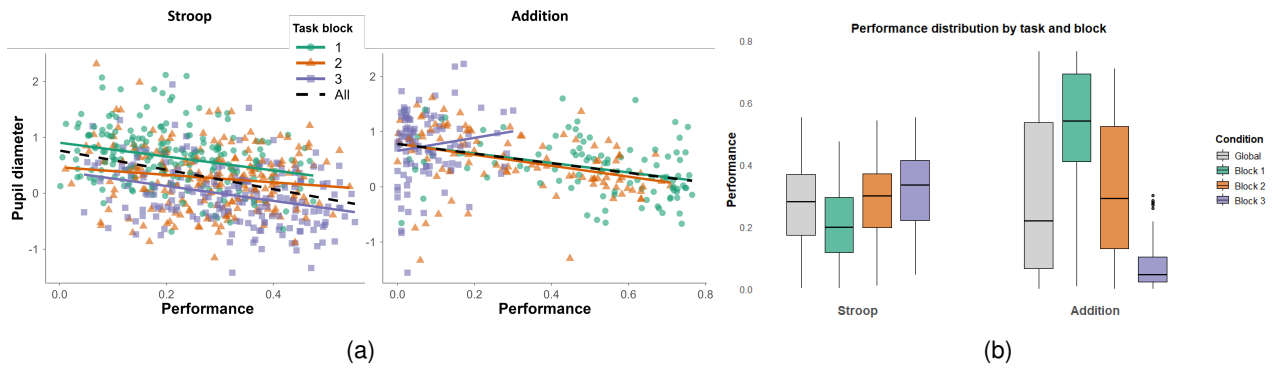


Figure 1: (a) PD versus performance for the 3 blocks of the Stroop (left) and Addition (right) tasks. Lines represent the linear relation between PD and performance for all task blocks and the full task. (b) Median and quartiles of performance on the 2 tasks during the full task duration and the 3 blocks independently.

is positively correlated with performance ( $r_s(610) = 0.151, p < 1.0e^{-03}$ ) for the full task (significant for block 1). GD is positively correlated with performance ( $r_s(610) = 0.153, p < 1.0e^{-03}$ ) only during the full task, as the ISCR, which negatively correlates with performance ( $r_s(574) = -0.202, p < 1.0e^{-05}$ ). Finally, there is no significant correlation between BD and performance for the Stroop task.

As for the Stroop task, for the Addition task, the PD is negatively correlated with performance ( $r_s(304) = -0.403, p < 1.0e^{-05}$ ) for the full task and the blocks 1 and 2. The correlation for block 3 is not significant and opposite. FD is positively correlated with performance ( $r_s(304) = 0.203, p < 1.0e^{-03}$ ) for the full task (significant in block 2). Finally, BD is negatively correlated with performance ( $r_s(100) = -0.168, p = 3.2e^{-03}$ ) for the full task (not per blocks). GD and ISCR are not significantly correlated with performance.

For both tasks, PD is the most correlated feature with performance. However, for addition task, the correlation coefficient is opposite in block 3 (see Figure 1a). The main difference between Addition and Stroop tasks is the difficulty level manipulation, which is constant across all the Stroop blocks but increases with each new Addition block. Figure 1b shows that performance in Addition block 3 is the lowest, reaching chance level. When participants cannot cope with the task efficiently, the correlation between PD and performance disappears, or even reverses.

## 4 Conclusions

The aim of the present study was to identify physiological features able to monitor cognitive performance. We investigated whether eye tracking and skin conductance metrics correlate with performance during the execution of cognitive tasks. Our results evidence that PD is the feature that correlates with performance the most: the better participants perform, the smaller their PD as a reflection of expertise and the consequent reduction in the cognitive load level required. However, when participants cannot deal efficiently with the proposed difficulty level, PD does not reduce as the cognitive load required does not decrease. Similarly, ISCR also represents the cognitive load required to cope efficiently with the task, but its changes are slower to detect, so we can observe its change only along the full task. The fixation duration is also related to performance, but we understand that it may be more dependent on the task instructions and modality than other eye tracking features. In future work, we aim to integrate the measure of eye tracking and skin conductance signals during neurofeedback sessions in order to monitor participants' engagement and performance. It could contribute to solving the inefficiency problem by adapting sessions in real time based on participants' engagement.

## References

- [1] M. Benedek and C. Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 2010.
- [2] L. A. Fos, K. W. Greve, M. B. South, C. Mathias, and H. Benefield. Paced visual serial addition test: an alternative measure of information processing speed. *Applied Neuropsychology*, (3), 2010.
- [3] A. Fragueiro, J. Fournier, R.-P. Debroize, and C. Cury. Eye-tracking and skin conductance to monitor cognitive task engagement. *Psychophysiology*, 2026.
- [4] D. Sato, Y. Yamazaki, A. Takahashi, Y. Uetake, S. Nakano, K. Iguchi, Y. Baba, R. Nara, and Y. Shimoyama. Water immersion decreases sympathetic skin response during color-word stroop test. *PLOS ONE*, 2017.
- [5] B. Sorger, F. Scharnowski, D. E. Linden, M. Hampson, and K. D. Young. Control freaks: Towards optimal selection of control conditions for fMRI neurofeedback studies. *NeuroImage*, 2019.