Eye-tracking and skin conductance to monitor task engagement during neurofeedback sessions

Agustina Fragueiro, René-Paul Debroize, Elise Bannier, Claire Cury

EYE-TRACKING AND SKIN CONDUCTANCE TO MONITOR TASK ENGAGEMENT DURING NEUROFEEDBACK SESSIONS

A. Fragueiro[1], R.-P. Debroize[1], E. Bannier[1,2], C. Cury[1]

[1]Univ Rennes, Inria, CNRS, Inserm - IRISA UMR 6074, Empenn ERL U 1228, Rennes, France
[2]Rennes University Hospital, Rennes, France

E-mail: agustina.fragueiro@inria.fr; claire.cury@inria.fr

ABSTRACT: The neurofeedback (NF) inefficacy problem refers to the variability in NF success and has been associated with attentional and motivational factors. Sustaining attention on any task over an extended period is demanding and leads to attentional drops. By using eye-tracking and skin conductance, we aimed at extracting physiological features linked to cognitive work, with the further purpose of monitoring changes in task engagement during NF sessions. Here, we present preliminary results on pupil diameter (PD) and phasic skin conductance responses (ISCR) linked to cognitive task execution. We observed that changes in both features are associated with performance and time-on-task. Thus, PD and ISCR decreased along the task while the performance increased. However, this trend is affected by manipulation of the task difficulty level. We also monitored, in the same participants, PD and ISCR during one NF session. Finally, we discussed preliminary ideas for target adaptation during NF sessions based on eye-tracking and skin conductance monitoring.

INTRODUCTION

Neurofeedback (NF) consists in feeding-back a patient with information about its neural activation to learn self-regulating its own brain activity [1]. It is therefore a powerful technique to trigger brain plasticity [2]. More importantly, NF has been postulated as a brain rehabilitation technique as it has the potential to reduce morbidity by correcting maladaptive patterns of brain function associated with a broad range of brain disorders [3]. NF is usually based on real-time electroencephalography (EEG) feature extraction, and it has been studied for several decades [4]. However, the NF inefficacy problem refers to the variability in NF success, as around 38% of participants undergoing NF training do not learn to regulate their own brain activity [5]. Among the different elements that may influence NF response, motivational [5] and attentional [6] factors have been identified as predictors of both performance and learning. Interestingly, motivation is likely to influence attention, as poorer performances can increase fear of incompetence and reduce mastering confidence which can lead to disengagement with the task and a potential label of "non-responder" [7]. To have the best success in NF training, it has been suggested to monitor participant's motivation [8], as well as, to adapt NF sessions to the participant [9]. A recent review on NF for post-stroke motor rehabilitation concluded that adaptation of NF target could lead to better meet patients' needs [10].

Sustaining attention on task-relevant information over an extended time is crucial for successful performance in any task, however, it is demanding and leads to attentional lapses (i.e., disengagement from the task). Performance relies on the "inverted-U shape" relationship between arousal and attentional states, linked to different on-task and off-task engagement states [11]. Only intermediate arousal activity is linked to task engagement and good performance [12]. Eye-tracking (ET) [13] and skin conductance (SC) [14] tools have been extensively used to measure physiological features related to both attention and arousal levels.

In the present study, by using ET and SC, we aim to extract features linked to cognitive task execution, with the further purpose of monitoring changes during NF sessions. To synchronize our multi-modal set-up, and to extract features of interest, we first collected data while participants were engaged in cognitive tasks. As proof of concept, the same participants performed one NF training session to observe physiological changes over time. Here, we present preliminary results on pupil diameter, as it has been strongly associated with cognitive load [15], performance [16], fatigue and task engagement [17], and on the phasic component of SC activity, as it has been linked with arousal [18] and it changes faster than the tonic component [14]. Finally, we monitored the same features during the NF session.

MATERIALS AND METHODS

Participants: Twenty right-handed volunteers (11 females, age range=$20-60$ yo) reporting normal vision to watch the screen participated in our study after signing informed consent. This study has been accepted by the COERLE, the Ethics Review Board at INRIA complying with the European General Data Protection Regulation.

Procedure: The protocol was divided into a 10-minute session of NF, followed by 30 min of randomised cognitive tasks, conducted while sitting in front of a screen with 1920×1080 display resolution in a dark room, while

simultaneously recording EEG, ET and SC (Fig. 1A).

Cognitive tasks: Participants completed two different tasks aiming to stimulate workload, (1) an adapted version of the Stroop task [19], and (2) an Addition task adapted from [20]. (1) During the Stroop task, participants were presented with two rows of words. In the row above the name of a colour (jaune, vert, rouge, bleu) is presented with the font in one of the following colours: yellow, green, red, and blue. In the row below, the name of a colour is written in black. The task of the participant was to judge if the meaning of the word below corresponds to the colour of the font of the row above or not, and press the corresponding key. Each trial was presented for 1250ms, followed by a 30ms inter-trial interval (ITI). Three task blocks (2min each), were interleaved with three rest blocks (2min each). (2) In each trial of the Addition task, a number from 1 to 9 was presented in the centre of the screen flanked by two other numbers. The task of the participant was to add the last two numbers presented in the centre of the screen and select the correct response among the flanker numbers by pressing the corresponding arrow on the keyboard. Three blocks of task (1min each) were alternated with three rest blocks (1min each). The time of appearance of the numbers decreases across the blocks to increase the difficulty level and prevent habituation (3sec in the 1st block, 2sec in the 2nd, and 1sec in the 3rd). An ITI of 30ms was set between trials. For both tasks, during rest blocks, a heart-coherence disengagement video was presented in which a blue dot increase (in 4sec) and decrease (in 4sec) repeatedly in the center of the gray screen.

Skin conductance: A BrainVision galvanic skin response set was used to acquire electrodermal activity from the index and middle fingers. After downsampling to 10Hz, SC responses were estimated through a Continuous Decomposition Analysis using MATLAB toolbox Ledalab [21]. Integrated phasic driver activity (ISCR), which corresponds to the area of the phasic driver within each temporal window, was extracted by setting 10sec consecutive temporal windows. Z-scores were computed along each task.

Eye-tracking: Eye activity was recorded by using a screen-based eye-tracker Tobii Pro X2-120 and the output was saved with Tobii Pro SDK. When pupil detection was judged as valid for both eyes according to the SDK validity codes, pupil diameter (PD) was averaged between both eyes. Z-scores were calculated for each task. To observe PD progression along time and for plotting purposes, PD was averaged within 10sec temporal windows.

Neurofeedback: Participants were instructed to perform a motor imagery task with both hands simultaneously. They were presented with a visual metaphor, a yellow ball moving inside a blue square rotated 90 degrees (Fig. 1B), depicting event-related desynchronization (ERD) activity of the C3 and C4 electrodes. The participant's goal was to keep the ball during all the runs in the upper corner of the metaphor, corresponding to a simultaneous motor imagery of both hands. They were informed that if their focus was more directed to the right or the left hand, the ball would move to the respective corner of the metaphor. The NF session consisted of one calibration run, eight training runs, and one post-run. Each run was composed of one rest (30sec) followed by one task block (30sec). To record EEG activity, we used a Brain Products actiCAP set of 32 active electrodes and an actiChamp amplifier. The Cz electrode was used as the reference. The OpenVibe software [22] was used for signal processing. Signals were epoched on the last 2 sec every 0.25 sec. A discrete Laplacian spatial filter with an 8 coefficient for the C3 and C4 channels and a -1 coefficient for their respective neighbours FC5, FC1, CP5, CP1 and FC6, FC2, CP6, CP2 was applied to the resulting samples, $Lap(C3)$ and $Lap(C4)$. Other channels were ignored. The power of $Lap(C3)$ and $Lap(C4)$ denoted $\mathrm{Bp}(C3)$, $\mathrm{Bp}(C4)$ is computed in the 8-30Hz frequency band. Powers $\mathrm{Bp}(C3)$ and $\mathrm{Bp}(C4)$ were continuously sent to a dedicated program via the LSL library to compute neurofeedback scores and display the ball in the right position on the metaphor. Neurofeedback scores for each electrode were computed with $Score(t) = \frac{\mathrm{Bp}_{ref} - \mathrm{Bp}(t)}{\mathrm{Bp}_{ref}}$ where Bp_{ref} is defined as $\mathrm{Bp}_{ref} = med(\mathrm{Bp}_{rest}|_{[10,20]})$, the median of the bandpower during the 10 sec central interval of the last rest block [23]. The goal target was set for each hand independently based on the 70th percentile of the scores achieved during calibration. If this score was lower than 0.15, the target was set to 0.15.

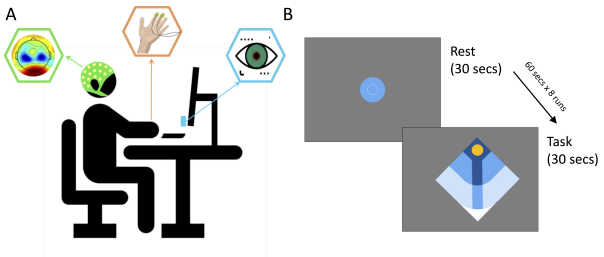

Figure 1: (A) Setup for data acquisition synchronously collecting EEG, eye-tracking and skin conductance signals, while the participant is engaged in a neurofeedback session and in resolving cognitive tasks presented on the screen. (B) A neurofeedback run composed of 30 sec rest, in which a heart-coherence video (increasing/decreasing blue dot), and 30 sec task, in which participants performed a motor imagery task of both hands simultaneously.

In addition, we preprocessed data offline with EEGLAB v2022.0 to test the difference between rest and task conditions during NF training. EEG was pass band filtered (1-40Hz) and re-referenced to average. EEG signal was corrected for ocular, muscular and noise artefacts using ICLabel 1.5 [24]. For power spectrum analyses, EEG was first selected with 1-29 sec time limits based on events corresponding to the beginning of the rest or the task conditions. For each condition, data was epoched in 2-sec no-overlapping temporal windows. Power in 8-30Hz fre-

quency band was compared between rest and task conditions. On the other hand, to observe ERS/ERD, data was epoched on 60 sec no-overlapping temporal windows, including rest (30 sec) and task (30 sec) conditions for each run (8 x subject).

Statistical analysis: To observe differences in performance along the task blocks, a one-way ANOVA was conducted for each task including accuracy as a dependent variable and block order (i.e. 1,2,3) as independent factor. For statistical analyses on both ISCR and PD, each task was analyzed independently and z-scores were averaged within each block. First, a one-way ANOVA was conducted including ISCR z-scores as the dependent variable and condition as the independent factor (i.e. baseline, task, rest). A two-way ANOVA was conducted including ISCR as the dependent variable and condition (i.e. task vs rest) and order (i.e. 1,2,3) as independent factors. To investigate the relationship between ISCR and accuracy, Pearson's correlations were conducted between ISCR and accuracy for each task block independently, for the total task (average among all three blocks), and for the whole task but including each block score separately. In addition, a delta score (block 3 - block 1) was calculated for both ISCR and accuracy scores. The correlation between these delta scores was also tested. The same analyses were conducted using mean PD z-scores within each block as a dependent variable on 13 subjects as the eye-tracking data for the first 7 subjects was excluded due to a change in the screen luminosity. To test learning during the NF session, paired sample t-test were conducted between calibration and post NF scores for C3, and for C4 on 18 subjects due to recording problems in the post-block for the first 2 subjects. The difference between NF scores for C3 and C4 was tested for both the calibration and the post runs. Finally, to observe changes in the NF score along the session, a one-way ANOVA was conducted one for C3 and one for C4 including NF scores as dependent variable and run as independent factor. To test changes in the synchronization between both hands, a one-way ANOVA was conducted including the difference in the NF scores between C3 and C4 as dependent variable. To compare power spectrum between the rest and task conditions during the training, paired t-tests were conducted for all channels independently (31 channels) and p-values were Bonferroni corrected. Analyses to investigate changes in ISCR along the NF session were conducted on 16 subjects due to recording problems on the first 4 subjects. A two-way ANOVA was conducted including ISCR as dependent variable and condition (i.e. rest vs task) and run (i.e. from 1 to 8) as independent factors. The same analysis was repeated for PD as dependent variable in 13 subjects. All p-values corresponding to post hoc tests included in this study were Bonferroni corrected. Statistical analyses were conducted on JASP 0.17.2.1.

RESULTS

Behavioral performance: For the Stroop task, mean accuracy across participants in the task was 80% ($sd = 0.13$) for the 1st block, 89% ($sd = 0.09$) for the 2nd block, and 91% ($sd = 0.07$) for the 3rd block. The difference in accuracy between blocks was significant ($F(2,57) = 7.263$, $p = 0.002$). Specifically, accuracy in the 1st block was lower than in the 2nd ($t(38) = -2.793$, $p = 0.021$) and 3rd ($t(38) = -3.642$, $p = 0.002$) blocks. Thus, performance increased along the task. For the Addition task, the mean performance across participants in the task was 89% accuracy ($sd = 0.15$) for the 1st block, 80% ($sd = 0.17$) for the 2nd block, and 44% for the 3rd block ($sd = 0.18$). As expected, the difficulty level between blocks was different ($F(2,57) = 40.755$, $p < 0.001$), with the 3rd block being significantly more difficult than the others (both $p < 0.001$).

Skin Conductance: For the Stroop task (Fig 2A), the one-way ANOVA evidenced a significant difference in ISCR between conditions (i.e. baseline, task, rest) ($F(2,137) = 26.071$, $p < 0.001$). Post hoc comparisons evidenced that the ISCR for the rest condition was significantly lower compared to the baseline ($t(78) = -5.022$, $p < 0.001$) and the task condition ($t(118) = -6.629$, $p < 0.001$). The two-way ANOVA, including ISCR as dependent variable and condition (i.e. task vs rest) and block order (i.e. 1,2,3) as independent factors evidenced a significant main effect of both condition ($F(1,114) = 73.537$, $p < 0.001$) and order ($F(2,114) = 17.583$, $p < 0.001$), and a significant interaction ($F(2,114) = 10.104$, $p < 0.001$). Post hoc comparisons showed that ISCR during rest was lower than the task condition ($t(118) = -8.575$, $p < 0.001$) and that ISCR in the 3rd block was lower comparing the 1st ($t(78) = -5.898$, $p < 0.001$) and 2nd blocks ($t(78) = -3.480$, $p = 0.002$). Specifically, the task 1st block was significantly different from both the 2nd ($t(38) = 4.659$, $p = 0.002$) and 3rd ($t(38) = 6.673$, $p < 0.001$) task blocks, and from all the rest blocks (all $p < 0.001$). ISCR in stroop task blocks 2 and 3 was not different, but the 2nd task block was different from the 1st and 3rd rest blocks (both $p < 0.002$). The rest blocks were no different. Mean ISCR during all three

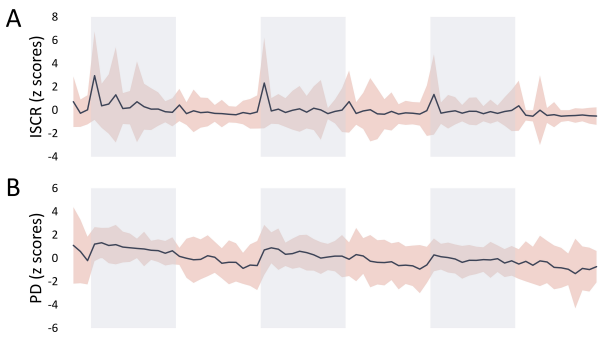

Figure 2: Stroop task. (A) Skin conductance responses (ISCR) and (B) Pupil's diameter (PD). Black line: the mean across subjects. Red area: +/-2 standard deviations. Blue shadow: task blocks. White areas: rest blocks (2 min each block). Total duration: 12 min plus 30 sec baseline at the beginning.

tasks blocks negatively correlated with mean total accuracy in the Stroop task ($r = -0.48$, $p = 0.034$) and, when observing the relationship for each block independently, ISCR negatively correlated with accuracy in the 2nd block ($r = -0.44$, $p = 0.05$) and the same tendency was observed in the 3rd block ($r = -0.39$, $p = 0.09$). Delta ISCR did not correlate with delta or mean accuracy. The correlation between accuracy scores and ISCR for the complete task, including all blocks independently, was significant ($r = -0.47$, $p < 0.001$), reflecting the decrement in ISCR and the improvement in accuracy along the task (Fig. 3A).

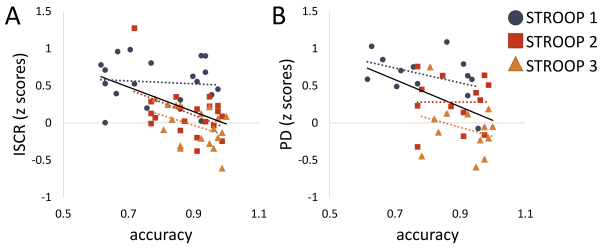

Figure 3: Stroop task. Pearson's correlations between accuracy scores and (A) Skin conductance responses (ISCR), and (B) pupil's diameter (PD). Dotted lines: linear correlations for each block. Black line: correlation over all blocks.

For the Addition task (Fig. 4A), the one-way ANOVA evidenced a significant difference in ISCR between conditions (i.e. baseline, task, rest) ($F(2,137) = 42.773$, $p < 0.001$). Post hoc comparisons evidenced that the ISCR during the addition task was significantly different from both the baseline ($t(78) = 6.210$, $p < 0.001$) and the rest ($t(118) = 8.607$, $p < 0.001$). The two-way ANOVA, including ISCR as dependent variable and condition (i.e. task vs rest) and block order (i.e. 1,2,3) as independent factors evidenced, as expected, a significant effect of condition ($F(1,114) = 84.986$, $p < 0.001$), a tendency for an order effect ($F(2,114) = 2.832$, $p = 0.063$) but no significant interaction. Post hoc comparisons showed that ISCR was higher during the task compared to the rest ($t(118) = 9.219$, $p < 0.001$). Specifically, ISCR during all the task blocks (i.e. 1, 2, 3) was significantly higher compared to ISCR in all the rest blocks (all p<0.001). ISCR was not different among the task nor the rest blocks. The correlation was not significant between ISCR and accuracy scores for the Additon task. However, we can observe a negative tendency between ISCR and accuracy in the 1st task block ($r = -0.40$, $p = 0.081$), and this relationship seems to be inverted in the case of an increment in the difficulty level linked with an abrupt decrement in accuracy in the 3rd task block (Fig. 5A). We observed a negative correlation between ISCR and accuracy scores along all the tasks when including each block score independently ($r = -0.28$, $p = 0.031$), as accuracy decreased along the task while ISCR increased with the increment on the difficulty level.

Eye Tracking: For the Stroop task (Fig. 2B), the one-way ANOVA including PD as dependent variable evi-

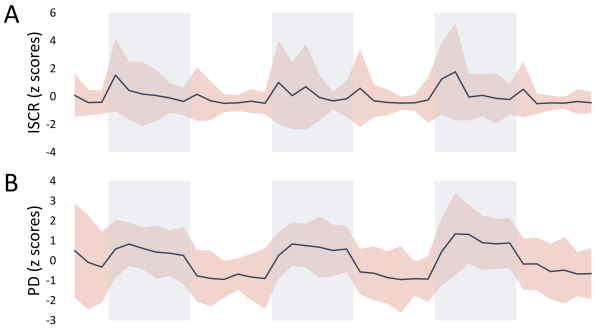

Figure 4: Addition task. (A) Skin conductance responses (ISCR) and (B) Pupil's diameter (PD). Lines and coloured areas are the same as for Fig. 2. Total duration: 6 min plus 30 sec baseline.

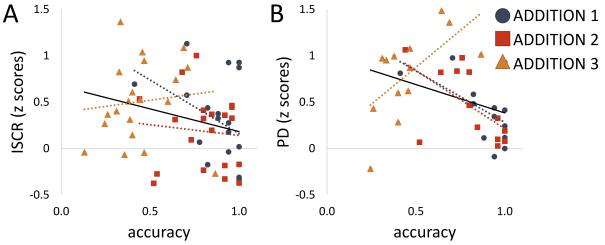

Figure 5: Addition task. Pearson's correlations between accuracy scores and (A) Skin conductance responses (ISCR), and (B) pupil diameter (PD). Dotted lines: linear correlations for each block. Black line: correlation over all blocks.

denced a significant difference between conditions (i.e. baseline, task, rest) ($F(2,88) = 20.407$, $p < 0.001$). Post hoc comparisons showed that PD during the rest was significantly smaller compared to the baseline ($t(50) = -4.734$, $p < 0.001$) and the task ($t(76) = -5.687$, $p < 0.001$). The two-way ANOVA, including condition (i.e. task vs rest) and order (i.e. 1,2,3) as independent factors, evidenced a significant main effect of both condition ($F(1,72) = 79.651$, $p < 0.001$) and order ($F(2,72) = 16.636$, $p < 0.001$), but no significant interaction ($F(2,72) = 2.396$, $p = 0.098$). Post hoc comparisons showed that PD was bigger during the task ($t(76) = 8.925$, $p < 0.001$) and that PD progressively decreased along the blocks, as PD was significantly different across all three blocks (all $p < 0.05$). Specifically, post hoc comparisons showed that PD in Stroop 1st block was significantly different from PD in all rest blocks (all $p < 0.001$) and the 3rd task block ($t(24) = 5.556$, $p < 0.001$). PD in Stroop 2nd block were also different compared to all the rest blocks (all $p < 0.01$). Finally, PD in the 3rd task block was different from the rest 3rd block ($t(24) = 3.893$, $p = 0.003$), but not from rest blocks 1 and 2. When testing the correlation between accuracy and PD, no significant correlation was found. However, the correlation between PD and accuracy scores along all the tasks, including each block scores independently, was significant ($r = -0.46$, $p = 0.004$), as PD decreased while accuracy increased along the task (Fig. 3B).

For the Addition task (Fig. 4B), the one-way ANOVA ev-

idenced a significant difference in PD between conditions (i.e. baseline, task, rest) ($F(2,88) = 67.208, p < 0.001$). Post hoc comparisons showed that PD during the addition task was significantly bigger compared to the baseline ($t(50) = 3.711$, $p = 0.001$) and rest ($t(76) = 11.586$, $p < 0.001$) and that PD during the rest was smaller also compared to the baseline ($t(50) = -4.482$, $p < 0.001$). The two-way ANOVA, including PD as dependent variable and condition (i.e. task vs rest) and order (i.e. 1,2,3) as independent factors, evidenced a significant main effect of both condition ($F(1,72) = 237.530$, $p < 0.001$) and order ($F(2,72) = 7.008$, $p = 0.002$), but no significant interaction. Post hoc comparisons showed that PD was bigger during the task ($t(76) = 15.412$, $p < 0.001$) compared to the rest and that PD increased along the task duration, as PD in the 3rd blocks were bigger compared to the 1st and 2nd blocks (both $p < 0.01$). Specifically, PD did not significantly change among the addition task blocks, nor the rest blocks, and PD in all the task blocks was significantly different from all the rest blocks. Thus, in the Addition task, we did not observe a decrement in PD along the tasks. When comparing PD and accuracy for the Addition task (Fig. 5B), a negative correlation was observed in the 2nd task block ($r = -0.59$, $p = 0.033$) and, although not significant, the same tendency was observed for the 1st task block ($r = -0.54$, $p = 0.055$). An opposite correlation was observed for PD and accuracy in the 3rd block ($r = 0.63$, $p = 0.022$). Averaged accuracy did not correlate with averaged PD, however, the PD delta positively correlated with accuracy in blocks 2 ($r = 0.85$, $p < 0.001$) and 3 ($r = 0.65$, $p = 0.017$), and with total accuracy ($r = 0.82$, $p < 0.001$). When bigger the increment in PD is along the task, higher the performance is. There is a negative correlation between PD and accuracy scores along all the tasks when including each block score independently ($r = -0.37$, $p = 0.022$), as accuracy decreased along the task while PD increased.

Neurofeedback: During the calibration, the NF score was significantly different between C3 and C4 ($t(17) = -2.189$, $p = 0.021$). When comparing calibration and post, although not significant, there was an increment in the NF score only for C3 ($t(17) = -1.580$, $p = 0.066$) (Fig. 6A). The one-way ANOVA conducted along the 8 runs, was not significant for NF scores in C3 nor C4, nor for the absolute difference between them (Fig. 6B). When comparing band power (8-30 Hz) between rest and task along all the NF training, the difference was significant for both C3 ($t(19) = 4.183$, $p = 0.016$) and C4 ($t(19) = 4.3703$, $p = 0.010$), but also for CP1 ($t(19) = 4.2515$, $p = 0.013$) and CP2 ($t(19) = 4.0952$, $p = 0.019$) (Fig. 6C). In the ERS/ERD conducted for training runs (i.e. 30 sec rest, 30 sec task), we observed 8-30 Hz ERD starting at the second 30, corresponding to the beginning of the task (Fig. 6D).

The two-way ANOVA including condition (rest vs task) and order (i.e. from 1 to 8) as independent factors and ISCR as dependent variable evidenced only a significant effect of order ($F(7,240) = 23.098$, $p < 0.001$), and a

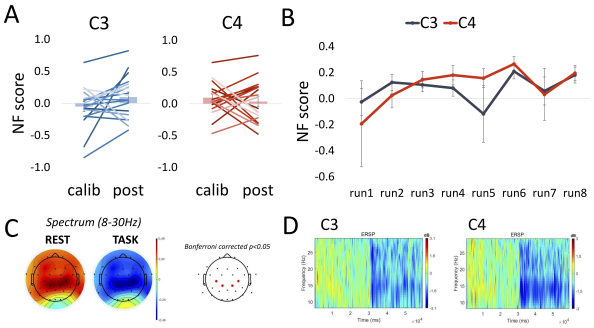

Figure 6: Neurofeedback. (A) NF scores calibration vs post blocks for C3 (blue) and C4 (red). One line per participant. (B) mean and SE for NF scores in the 8 task blocks. (C) Topoplot of task vs rest, and channels with a significant difference (in red) after Bonferroni correction. (D) ERS/ERD in 8-30Hz frequency band along NF runs (30 sec rest then 30 sec task).

significant interaction ($F(7,240) = 3.283$, $p = 0.002$). Post hoc comparisons confirmed that ISCR for the 1st and 2nd runs was significantly higher compared to all the other runs (all $p < 0.01$). The same analysis with PD as a dependent variable evidenced a significant effect of condition ($F(1,192) = 18.263$, $p < 0.001$) and a tendency (not significant) for an order effect ($F(7,192) = 1.500$, $p = 0.169$). Specifically, PD in the rest conditions was significantly smaller compared to the task ($t(206) = -4.160$, $p < 0.001$). Although not significant, PD in runs 1 and 2 tended to be bigger compared to run 7.

DISCUSSION

As expected we observed differences in both ISCR and PD between the task and the rest conditions for both the Stroop and the Addition tasks. However, we observed a different trend for both physiological features (i.e. ISCR and PD) in the two different tasks. While in the Stroop task we observed a decrement in both physiological signals along the task, on the contrary, we didn't observe the same trend in the Addition task, in which the difficulty level was increasing along blocks. Thus, in the Stroop task performance increased along the task blocks while physiological signals decreased, probably linked to an expertise and/or habituation effect. In the Addition task, both ISCR and PD remained stable along the task, while the relationship between the physiological features and performance changed based on the task difficulty level. In the case of manipulation of the difficulty level, participants who showed less decrement in physiological features were able to keep a higher performance. For NF, we observed a decrement in physiological signals during the session. This may be an indicator of habituation linked to expertise or, on contrary, linked to fatigue or even to a lack of cognitive effort allocated on the task. In a future study, we aim to develop a method that progressively adapts the NF target based on individual performance and physiological features monitoring. In NF protocols, with

the repetition of sessions, participants may loose engagement on the task and this may be linked to: (1) the task has become too easy, or (2) they give up due to the difficulty in achieving the goal. For instance, if NF scores are high and physiological features are dropping during the session, we may increase the target difficulty level to make the session more challenging. However, if NF scores are high, but also physiological features are high, the difficulty level might be challenging enough. If NF scores are low and physiological signals drop, the target might be set easier to re-engage the participant's interest. Finally, we would like to briefly comment on the difference in the NF score for calibration between c3 and c4. We understand that, as all participants were right-handed, the motor imagination of a movement with the left hand might be less automatic compared to the right. Interestingly, we observed differences only for the calibration, during the NF runs, the difference tended to decrease.
This study aimed to present preliminary results on the use of skin conductance and eye-tracking features to monitor task engagement during NF sessions. These first results encourage to individually adapt NF targets to keep participants engaged in the task. In future analyses, we will consider also other physiological features that might be of interest to track cognitive load and attention, such as blinks, saccades and fixations, as well as changes in the tonic component of skin conductance.

ACKNOWLEDGEMENTS

This work is funded by the Inria Exploratory Action EyeSkin-NF.

REFERENCES

[1] Sitaram R *et al.* Closed-loop brain training: The science of neurofeedback. Nature Reviews Neuroscience. 2017;18.
[2] Loriette C, Ziane C, Ben Hamed S. Neurofeedback for cognitive enhancement and intervention and brain plasticity. Revue Neurologique. 2021;177.
[3] Stoeckel L *et al.* Optimizing real time fMRI neurofeedback for therapeutic discovery and development. NeuroImage: Clinical. 2014;5.
[4] Marzbani H, Marateb HR, Mansourian M. Neurofeedback: A Comprehensive Review on System Design, Methodology and Clinical Applications. Basic and Clinical Neuroscience. 2016;7.
[5] Haugg A *et al.* Predictors of real-time fMRI neurofeedback performance and improvement – A machine learning mega-analysis. NeuroImage. 2021;237.
[6] Hammer EM *et al.* Psychological predictors of SMR-BCI performance. Biological Psychology. 2012;89.
[7] Kadosh KC, Staunton G. A systematic review of the psychological factors that influence neurofeedback learning outcomes. NeuroImage. 2019;185.
[8] Sorger B, Scharnowski F, Linden DE, Hampson M, Young KD. Control freaks: Towards optimal selection of control conditions for fMRI neurofeedback studies. NeuroImage. 2019;186.
[9] Alkoby O, Abu-Rmileh A, Shriki O, Todder D. Can We Predict Who Will Respond to Neurofeedback? A Review of the Inefficacy Problem and Existing Predictors for Successful EEG Neurofeedback Learning. Neuroscience. 2018;378.
[10] Le Franc S *et al.* Toward an Adapted Neurofeedback for Post-stroke Motor Rehabilitation: State of the Art and Perspectives. Frontiers in Human Neuroscience. 2022;16.
[11] Unsworth N, Robison MK. Tracking arousal state and mind wandering with pupillometry. Cognitive, Affective, & Behavioral Neuroscience. 2018;18.
[12] Aston-Jones G, Cohen JD. An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. Annual Review of Neuroscience. 2005;28.
[13] Skaramagkas V *et al.* Review of eye tracking metrics involved in emotional and cognitive processes. IEEE reviews in biomedical engineering. 2021;PP.
[14] Critchley HD. Review: Electrodermal Responses: What Happens in the Brain. The Neuroscientist. 2002;8.
[15] Piquado T, Isaacowitz D, Wingfield A. Pupillometry as a measure of cognitive effort in younger and older adults. Psychophysiology. 2010;47.
[16] Brink RL van den, Murphy PR, Nieuwenhuis S. Pupil Diameter Tracks Lapses of Attention. PloS One. 2016;11.
[17] Hopstaken JF, Linden D van der, Bakker AB, Kompier MAJ, Leung YK. Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data. Journal of Experimental Psychology: Human Perception and Performance. 2016;42.
[18] Lane RD, Chua PML, Dolan RJ. Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. Neuropsychologia. 1999;37.
[19] Sato D *et al.* Water immersion decreases sympathetic skin response during color–word Stroop test. PLOS ONE. 2017;12.
[20] Fos LA, Greve KW, South MB, Mathias C, Benefield H. Paced Visual Serial Addition Test: An alternative measure of information processing speed. Applied Neuropsychology. 2000;7.
[21] Benedek M, Kaernbach C. A continuous measure of phasic electrodermal activity. Journal of Neuroscience Methods. 2010;190.
[22] Renard Y *et al.* OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain–Computer Interfaces in Real and Virtual Environments. Presence: Teleoperators and Virtual Environments. 2010;19.
[23] Lioi G *et al.* Simultaneous EEG-fMRI during a neurofeedback task, a brain imaging dataset for multimodal data integration. Scientific Data. 2020;7.
[24] Pion-Tonachini L, Kreutz-Delgado K, Makeig S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. NeuroImage. 2019;198.