Effect of a Real-Time Psychophysiological Feedback, Its Display Format and Reliability on Cognitive Workload and Performance

Sami Lini[1], Lise Hannotte[2] and Margot Beugniot[1]
[1]*Akiani, Talence, France*
[2]*ENSC, Talence, France*

Keywords: Cognitive Load, Pupil Dilation, Intervention Feedback, Multiple Object Tracking.

Abstract: For a long time, literature has identified some psychophysiological metrics that proved reliable to assess cognitive states in controlled conditions. Smaller, more reliable and more affordable sensors made the industrial community plan to design systems that would adapt themselves to the ability of their users to operate them. Thus an important human factors question must be asked: what is the impact of such a feedback on users' performance and cognitive workload? Does the display format of this feedback have an influence over subjects? What if the feedback provides erroneous data?

We designed a protocol to compare the influence of providing a cognitive load assessment gauge versus raw data versus no feedback in a Multiple Objects Tracking task. Reliability of this feedback was also evaluated. Performance in a dual task paradigm, pupil dilation and questionnaire were used to assess cognitive load. Trials duration and learning effect were used as control results. Raw feedback showed a negative effect while low reliability showed inconsistent results.

1 INTRODUCTION

Human monitoring issues (i.e. measuring the psychophysiological state of operators to assess their cognition) are getting more attention from the industrial community every day. Literature has, for some time now, identified several metrics that have proven reliable in controlled and operational conditions, such as heart rate variability (Egelund, 1982) or pupil dilation (Beatty and Lucero-Wagoner, 2000).

Physiological data are getting more and more accessible due to smaller, more reliable and more affordable sensors. Due to considerable scientific progress for real-time evaluation of cognitive workload (Afergan et al., 2014; George and Lécuyer, 2010; Kohlmorgen et al., 2007), the industrial field is planning to design systems that would adapt themselves to the ability of their users to operate them: for example, a cockpit display which would only show relevant information if the system has assessed pilots are suffering cognitive overload. In such a case, pilot would probably notice his display is decluttering and would be able to come to the conclusion that the system "thinks" he is not fully able to perform his duty.

Psychophysiological feedbacks have been used for decades in behavioral therapies, and some studies proved they have a significant effect on fear (Valins and Ray, 1967) or anxiety (Story and Craske, 2008). Feedback intervention is defined by Kluger and DeNisi (1996) as "*actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one's task performance*". Their meta meta-analysis showed that over 1/3 of these feedbacks have a negative effect on performances.

Cognitive workload feedback, i.e. a real-time feedback about one's cognitive load, can be considered as a feedback intervention. Thus, an important question must be addressed: what is the impact of such a feedback on a user's performance and cognitive state? Does the way this feedback is displayed have an influence on the users' performance?

In order to answer these questions, we designed an experimental protocol using a Multiple Objects Tracking (MOT) task (Pylyshyn and Storm, 1988). In this task subjects are asked to track a defined number of moving targets among identical distractors and identify them after a few seconds. We chose this task as it is a highly engaging visual attention task, widely used in the literature and

Lini, S., Hannotte, L. and Beugniot, M.
Effect of a Real-Time Psychophysiological Feedback, Its Display Format and Reliability on Cognitive Workload and Performance.
DOI: 10.5220/0005939500750079
In *Proceedings of the 3rd International Conference on Physiological Computing Systems (PhyCS 2016)*, pages 75-79
ISBN: 978-989-758-197-7

which can be compared to the monitoring tasks fighter pilots or radar operators are expected to perform (Allen, McGeorge, Pearson, and Milne, 2004).

In order to assess the impact of the display format of the cognitive workload feedback, three conditions were implemented: in one condition, _raw feedback_ was provided showing heart rate data, in a second condition a *gauge* gave an assessment of the cognitive load. The control condition displayed *no feedback* at all.

Data unreliability has been identified for a long time as a major cause of distrust in systems (Wickens, Gempler, and Morphew, 2000) and is known to influence decision making processes. In order to evaluate the impact of data quality over users' cognition, half of the trials showed obviously erroneous random data.

Results of a dual tasks paradigm, pupil dilation recorded from an eye tracking system and a questionnaire were used to assess cognitive load in each trial.

Hypotheses were as follow:

- (H1) A cognitive feedback has an effect on cognitive load and performance;
- (H2) The display format of feedback has an influence over performances and cognitive workload;
- (H3) An erroneous feedback leads to an increased cognitive workload and decreased performances.

2 METHOD

2.1 Participants

31 subjects (M = 21,6 years old, □ = 1,3 y.o., 20 males) took part to the study. They were recruited among the students of the National Cognitive Engineering School (Bordeaux, France).

Any text or material outside the aforementioned margins will not be printed.

2.2 Apparatus

Participants were tested individually. The procedure was explained, and ethical consent obtained. Participants then completed 96 trials of an MOT task, lasting approximately 45 min.

An Eye Tribe eye tracking system sampled subjects' data at 30 Hz (Lopez, Hansen, Sztuk, and Tall, 2014). Stimuli were displayed on a 24 inch 16:9 LCD screen. Luminosity was controlled during each session. Prior to the experiment subjects were familiarized with the task.

Ogama (Voßkühler, 2009) software was used to collect eye tracking data.

2.3 Apparatus

2.3.1 Multiple Object Tracking Task

Our experimental protocol is based upon a Multiple Object Tracking (MOT) task (Pylyshyn and Storm, 1988). In this well documented experimental task, subjects are asked to track a defined number (3 in our protocol) of moving targets among identical distractors (6 in our protocol).

At the beginning of each trial 9 identical balls appeared randomly placed within the display. Three of those balls were displayed in red for 1 second, which assigned them as targets to track.

Then all 9 balls started to move along independent trajectories with constant speed for a variable time (five or nine seconds). After 1 second, the targets turn white, making them less distinguishable from the distractors. When the balls' trajectories crossed, they collided in a predictable way (Drew, Horowitz, and Vogel, 2013).

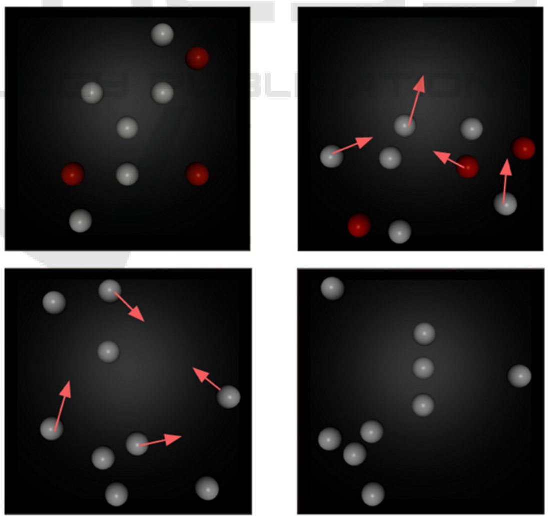

Figure 1: Multiple object tracking.

Immediately after each trial, subjects were instructed to identify, as quickly as possible, the three targets among all items in a "Mark all" manner (Hulleman, 2005) by clicking on them.

Each participant undertook one practice trial prior to completing the task.

2.3.2 Independent Variables

- Task Difficulty

We hypothesize that cognitive overload situations are not regulated with a cognitive feedback. Increasing the duration of the tracking task has been proven to decrease tracking performances (Oksama and Hyönä, 2004; Pylyshyn, 2004) thus making it more difficult. Subjects were asked to track the targets in two different conditions: for 5 seconds in the easy condition and for 9 seconds in the hard condition following Oksama and Hyönä (2004) results.

- Feedback Display Format

Given that no actual cognitive workload gauge exists yet, we decided to provide users with a fake feedback using the estimated difficulty of each trial and of the previous one. In order to assess the impact of the display format, two conditions were designed. Subjects were wearing a heart rate belt and facing an eye tracking system. In the first condition (raw), a raw display of the subject heart rate and pupil diameter is provided. In the second condition (interpreted), a colored gauge displayed a color gradient from green to red (the warmer the gauge, the higher the cognitive workload). Color was used in order to offer instantaneous visual information to the subject. In a control condition no feedback was given at all.

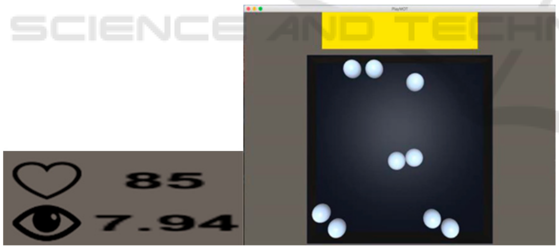

Figure 2: Raw and interpreted (coloured gauge)feedback.

- Reliability of the Feedback

As we wanted to evaluate the impact of an obviously erroneous cognitive feedback, we designed two different conditions. In the reliable control condition, a realistic feedback was provided to the user, taking into account the difficulty of both the current trial and the previous one. In the erroneous condition, feedback evolved in an obviously random way.

The task consisted of twelve (2x3x2) blocks of trials, with each block of trials consisting of eight trials (ninety-six trials total) in order to reach statistical thresholds.

Subjects were told that the feedback was displayed to give them information about a lack of focus and to help them achieve better performance.

2.3.3 Dependent Variables

- Dual Task

A dual task paradigm was implemented to increase cognitive load and collect more performance data. Subjects were asked to press on the “A” key of a keyboard when they heard a klaxon sound and the “E” key when it was a bell ring. Subjects were instructed to answer as fast as possible.

Both reaction time and answer were recorded as performance indexes.

- MOT Results

After each trial subjects were instructed to identify the three targets using a mouse. The number of correct answers is used as a performance index.

- Pupil Dilation

Pupil dilation was recorded using an Eye Tribe system with a sampling rate of 30Hz. Due to data loss, 9 subjects were excluded from the analysis. Pupil dilation has been well known to correlate to performances in memory span tasks (Kahneman and Beatty, 1966) and in attention tasks (Beatty, 1988)

- Questionnaire

Following the China Lake questionnaire approach (Gawron, 2008), we asked the subjects to self-assess their cognitive load on a scale from 0 (low workload) to 100 (high workload) at the end of each trial.

3 RESULTS

Non-parametrical paired statistical tests (Wilcoxon, Friedman) were performed using *R*, *Matlab* and *XLStat*.

In order to evaluate our setup, we first study effects that have been validated by the literature on Multiple Object Tracking: duration as a factor of difficulty (Oksama and Hyönä, 2004) and learning effect (Makovski, Vázquez, and Jiang, 2008).

3.1 Control Conditions

3.1.1 Effect of the Difficulty (Trial Duration)

One-tailed analysis of reaction time during the dual task showed significant differences ($p = 0.02$): lower values are recorded for the the easy condition.

Performance for the auditory task did not show any difference.

Tracking results at the MOT task showed a trend (p=0,08) in favor of the easy condition: more targets are retrieved in the easy condition.

One-tailed analysis of the questionnaire's answers showed a significant difference in favor of the easy condition (p=1.6e-05): lower results are shown for this condition.

Pupil dilation showed a significant difference in favor of the easy condition (p=0,008). Mean pupil dilation is lower for the easy condition.

3.1.2 Learning Effect

One-tailed analysis of reaction time during the dual task showed significant differences (p = 0.04): lower values are recorded for the first trials compared to the last ones.

Performance for the auditory task did not show any difference.

Tracking results at the MOT task showed a significant difference (p=0,01) in favor of the latest trials: the later the trial, the higher the number of items retrieved.

One-tailed analysis of the questionnaire's answers showed a significant difference in favor of the easy condition (p = 0,01).

Pupil dilation shows significant differences between the first and the last trials.

3.2 Display Format of the Feedback

Analysis of reaction time using Friedman's test during the dual task showed no significant differences between the 'no feedback,' 'raw value' and the 'colored gauge' conditions.

Performance at the auditory task did not show any difference.

Tracking results at the MOT task did not show any significant difference.

Questionnaire's answers showed no significant results.

Analysis using Friedman test of pupil dilation between conditions showed a trend (p=0,055). Trend differences (p=0,02) were found using a Wilcoxon test between the no feedback condition and the raw condition in favor of the no feedback condition as well as between the interpreted feedback and no feedback (p=0,046). A trend (p = 0,08) was found between the two kinds of feedbacks in favor of the interpreted feedback.

3.3 Reliability of the Feedback

Reaction time during the dual task showed no significant differences. Performance at the auditory task did not show any difference either.

Tracking results for the MOT task did show a trend (p=0,07) in favor of the unreliable condition.

The questionnaire responses did not show any statistical difference.

Pupil dilation showed a trend in favor of the reliable condition (p=0,07): pupil dilation is lower when data are consistent with the difficulty of the task.

4 DISCUSSION

Retrieval performances, reaction times, subjective evaluation and pupil dilation are consistent with the literature on the learning effect and the effect of trial duration as a difficulty factor. This validates our experimental setup.

Regarding our first hypothesis (H1), we found differences for the display format of the feedback. Pupil dilation results indicate that a physiological feedback seems to have a negative effect cognitive workload. We can explain this result by stating that a feedback needs resources from the subject. Subjects try to use this information while performing the task. The overall cognitive effort is then higher. This validates partially our H1 hypothesis: a psychophysiological feedback has a negative effect on cognitive workload not improving it in the MOT task.

Our second hypothesis (H2), that display format will show an influence over cognitive workload, was partially confirmed by higher mean pupil dilation with the raw feedback results. This result can be explained by the fact that raw data needs user's interpretation, which, therefore, increases the cognitive effort. This validates partially our H2 hypothesis.

Finally, our third hypothesis (H3) that erroneous feedback leads to increased cognitive workload and decreased performance was not validated, as the results only showed trends and are not consistent. When the feedback is not reliable, pupil dilation shows a trend toward a higher workload but retrieval performances are also better. We can explain this result by assuming that subjects noticed that the feedback was not usable and consequently invested more resources in focusing on the task while ignoring the feedback.

5 CONCLUSIONS

The present study investigated the effect of a psychophysiological feedback, its display format and its reliability on performance and cognitive workload during a Multiple Object Tracking task. This feedback was presented to the subjects as a means to improve their focus on the task in order to reach better levels of performance.

This task was chosen as it is a visual attention task which can be compared to the attention tasks fighter pilots or air traffic controllers are regularly expected to perform. Results on duration and difficulty are consistent with the literature (Makovski et al., 2008; Oksama and Hyönä, 2004).

In a highly engaging task such as the Multiple Object Tracking, displaying a psychophysiological feedback has a significant effect on subjects. More specifically, a psychophysiological feedback leads to higher cognitive workload compared to no feedback at all. Raw data increases cognitive workload compared to an interpreted colored gauge.

Reliability of the feedback showed inconsistent results: better performance with higher workload. We made the assumption that the feedback being ignored could explain this result.

As the MOT needs a lot of attention, eye tracking data should be investigated further in order to evaluate links between results and gaze patterns, particularly the attention provided to the feedback.

We believe the next logical step would be to evaluate effects of direct and psychophysiological measures as feedback intervention on users' cognition and performance.

REFERENCES

Afergan, D., Peck, E. M., Solovey, E. T., Jenkins, A., Hincks, S. W., Brown, E. T., … Jacob, R. J. K. (2014). Dynamic difficulty using brain metrics of workload. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3797–3806). ACM.

Allen, R., McGeorge, P., Pearson, D., and Milne, A. B. (2004). Attention and expertise in multiple target tracking. *Applied Cognitive Psychology*, *18*(3), 337–347. doi:10.1002/acp.975.

Beatty, J. (1988). Pupillometric signs of selective attention in man. In E. Donchin, G. Galbraith, and M. L. Kietzman (Eds.), *Neurophysiology and psychophysiology: Experimental and clinical applications* (pp. 138–143).

Beatty, J., and Lucero-Wagoner, B. (2000). The pupillary system. In *Handbook of psychophysiology* (2nd ed., pp. 142–162). New York, NY, US: Cambridge University Press.

Drew, T., Horowitz, T. S., and Vogel, E. K. (2013). Swapping or dropping? Electrophysiological measures of difficulty during multiple object tracking. *Cognition*, *126*(2), 213–223.

Egelund, N. (1982). Spectral analysis of heart rate variability as an indicator of driver fatigue. *Ergonomics*, *25*(7), 663–672. doi:10.1080/00140138208925026.

Gawron, V. J. (2008). *Human performance, workload, and situational awareness measures handbook*. CRC Press.

George, L., and Lécuyer, A. (2010). An overview of research on" passive" brain-computer interfaces for implicit human-computer interaction. In *International Conference on Applied Bionics and Biomechanics ICABB 2010-Workshop W1" Brain-Computer Interfacing and Virtual Reality."*

Hulleman, J. (2005). The mathematics of multiple object tracking: From proportions correct to number of objects tracked. *Vision Research*, *45*(17), 2298–2309.

Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585.

Kluger, A. N., and DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254.

Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K.-R., Curio, G., … Kincses, W. (2007). Improving human performance in a real operating environment through real-time mental workload detection. *Toward Brain-Computer Interfacing*, 409–422.

Lopez, J. S. A., Hansen, J. P., Sztuk, S., and Tall, M. H. (2014, February). Systems and methods of eye tracking calibration. Google Patents.

Makovski, T., Vázquez, G. A., and Jiang, Y. V. (2008). Visual learning in multiple-object tracking. *PLoS One*, *3*(5), e2228.

Oksama, L., and Hyönä, J. (2004). Is multiple object tracking carried out automatically by an early vision mechanism independent of higher□order cognition? An individual difference approach. *Visual Cognition*, *11*(5), 631–671.

Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, *11*(7), 801–822.

Pylyshyn, Z., and Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism*. *Spatial Vision*, *3*(3), 179–197.

Story, T. J., and Craske, M. G. (2008). Responses to false physiological feedback in individuals with panic attacks and elevated anxiety sensitivity. *Behaviour Research and Therapy*, *46*(9), 1001–1008.

Valins, S., and Ray, A. A. (1967). Effects of cognitive desensitization on avoidance behavior. *Journal of Personality and Social Psychology*, *7*(4p1), 345.

Voßkühler, A. (2009). OGAMA description (for Version 2.5). *Berlin, Germany: Freie Universität Berlin, Fachbereich Physik*.

Wickens, C. D., Gempler, K., and Morphew, M. E. (2000). Workload and Reliability of Predictor Displays in Aircraft Traffic Avoidance. *Transportation Human Factors*, *2*(2), 99–126. doi:10.1207/STHF0202_01.