Eye activity measures as indicators of drone operators’ workload and task completion strategies

Philippe Rauffet¹, Assaf Botzer², Alexandre Kostenko¹, Christine Chauvin¹, & Gilles Coppin³
¹University of South Brittany, Lorient, France
²Ariel University, Israel
³Telecom Bretagne, Brest, France

Abstract

We studied whether eye activity patterns in a simulated drone-operating task could be associated with workload levels and task completion strategies. Participants sent drones to suspected areas according to messages they received and according to self-initiated search. They were also required to validate whether suspected targets were indeed hostile prior to attacking them. We tested whether the number of suspected targets affected the number of eye transitions between task zones and whether it affected fixation duration in different task zones. We found that operators made fewer transitions between task zones as the number of targets increased. This was because they focused more on one zone and not on the others. Interestingly, the zone they attended to relatively more was the one they needed for attacking targets and not the ones where targets usually appeared. This was probably because attacking required extended cognitive operations. Findings demonstrated that eye activity patterns can be used to infer task completion strategies and to identify workload levels, once these strategies are described. Workload levels and task completion strategies should therefore be studied by a combination of hypothesis driven and exploratory driven methods. Eye activity patterns can then be used as triggers for assisting overloaded operators.

Introduction

Mental workload (MWL) and task completion strategies are interrelated. They both affect the degree of successful task completion as operators change their strategies and implement different mental workload’s regulation loops for improving task performance or decreasing task induced cognitive load (Hockey, 1997; Kostenko, Rauffet, Chauvin, & Coppin, 2016; Schulte, Donath, & Honecker, 2015). Designers of work environments would therefore like to gain insight into the levels of MWL that human operators experience and to learn about their task completion strategies. Such insights may assist in integrating adaptive automation to support operators in times of high MWL. As design validation criteria, these ocular MWL indicators can also point out the design of harder-to-operate interface components and may inform training programs to correct less productive task completion strategies (Byrne & Parasuraman, 1996; Hockey, 1997; Nickel & Nachreiner, 2003; Van Orden et al., 2001).

Performance measures, self-report measures (e.g., NASA-TLX) and physiological measures (e.g., heart rate variability; skin conductance) are all widely acceptable for estimating workload and to some extent for learning about task completion strategies. At the same time, however, they often fall short in detecting short periods of elevated workload that may trigger changes in task completion strategies, or conversely, indicate the recent occurrence of such changes (Hilburn & Jorna, 2001; Verwey & Veltman, 1996). Thus, these measures may sometimes not provide the information that practitioners would like to have. Eye activity measures, in contrast, are more adapt for learning about short periods of elevated workload and about shifts in task completion strategies. This is because they indicate what sources of visual input operators attend to and for how long, and they are therefore indicative of the cognitive processes that this input serves (Rayner, 1998; Salvucci, 2001).

For instance, Bijleveld, Custers & Aarts (2009) demonstrated that greater potential rewards in a digit-retention task led to increases in pupil dilation as would be expected when people invest more effort in tasks. Salvucci (2006) demonstrated how cognitive modelling of driver lane keeping, curve negotiation and lane changing corresponded with gaze distribution in driving. Botzer et al. (2015) demonstrated that aid from automation in a simulated quality control task led to changes in search patterns of faulty items and to changes in how much time decision makers inspected items. These changes corresponded with decision makers reported effort. Finally, Van Orden et al. (2001) demonstrated that blink frequency, fixation frequency and pupil diameter could be used to predict fluctuations in target density in a simulated anti-air-warfare task.

Still, in Van Orden et al. (2001), gaze distribution patterns of some participants were not related to task completion strategies, but rather to task disengagement. Next, while higher workload led to greater frequency of fixations but not to changes in fixation durations in a visual search task (Zelinsky et al., 1997), higher cognitive processing load did correspond with longer fixation durations in a flight task (Callan, 1998). Thus, MWL affects eye activity in different ways depending on the task, and one should therefore interpret eye activity measures according to task characteristics and according to prior knowledge and hypotheses. Exploratory inspection of the data together with alternative hypotheses about operators’ cognitive processes should be also employed to learn about task completion strategies and operators’ MWL.

Based on the methods and insights from a body of research on eye activity measures and human performance, we set to explore drone operators’ MWL and task completion strategies. We expected that greater density of hostile targets that operators need to handle in a simulated drone-operating task would lead to greater fixation durations in certain task zones and to fewer gaze transitions between task zones.
Method

Participants

Twenty-two participants, aged from 18 to 20 (mean: 19, standard deviation: 0.7) took part in the experiment. For reasons of homogeneity, all were men and had good experience with video gaming.

Simulation set-up and experimental task

The task was a simulation of securing an area with a swarm of drones that we ran on a demonstrator, named SUSIE (Coppin & Legras, 2012). SUSIE is supported by Java software, and it allows participants to interact with and to supervise a swarm of drones using a mouse-screen system.

Only one operator is required, but some tasks can be or are achieved by an artificial agent. The system provides different information to the operator from two sources: a dynamic map and a message banner (Figure 1). The dynamic map gives information about the areas that the drones control such as the vehicles in these areas and their state. The message banner indicates the coordinates and direction of a vehicle that the operators need to assign high priority to its neutralization.

Figure 1. Dynamic monitoring map

The main task is to detect and neutralize the threats (i.e., hostile vehicles) on the map. When a vehicle is generated by the software, it is hidden, i.e. it is present on the map but invisible (it has to be detected by drones sent by the participant). Before it is neutralized, the status of the vehicle changes several times (Fig. 2).

Figure 2. The different vehicle states
To advance from one status to the next, operators need to complete a number of sub-tasks, related to different areas of interest (AOI). Part of these AOIs, as the message zone, drone base and bombers base, are static and part, as the user-defined research zones and the blank zone, whose surface changes according to the creation or removal of user-defined research zones, are dynamic. Note that blank zones are used to identify vehicles, but also to define new user-defined research areas. In other words, the blank zone is all the space of the map where there are no specific AOIs. It is on this blank zone that participants create new user-defined research zones, and it is on this blank zone that participants achieve the identification sub-task (Table 1).

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Description</th>
<th>Associated Areas Of Interest (AOI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1: Read a message to locate a vehicle</td>
<td>Extract information on a suspected vehicle (coordinates and direction) from a message.</td>
<td>Message zone</td>
</tr>
<tr>
<td>T2: search for a new vehicle</td>
<td>Search vehicle by creating a user-defined research zone. The drone moves automatically in a zone, a vehicle is detected when a drone is flying over it.</td>
<td>Blank zone (participants have to create new research zones on the map)</td>
</tr>
<tr>
<td>T3: Lock a detected vehicle</td>
<td>Lock a vehicle (from state 2 to state 3) by flying over it a second time.</td>
<td>User-defined research zones</td>
</tr>
<tr>
<td>T4: Identify a vehicle as a target</td>
<td>Select and identify a vehicle as a threat or not (from state 3 to state 4). In the experimental version, this task is simulated by a basic cognitive task, which consists of sorting out nine numbers in an increasing order.</td>
<td>User-defined research zones (participants select a locked vehicle in a research zone) Blank zone (participants make the identification on a number square, appearing on the map, outside the research zones)</td>
</tr>
<tr>
<td>T5: Attack a target</td>
<td>Neutralize a vehicle (from state 4 to state 5) by drawing a corridor starting from the bombers base, which simulates the sending of a helicopter.</td>
<td>Bombers base</td>
</tr>
<tr>
<td>T6: Manage drones in user-defined research zones</td>
<td>Regulate the number of drones in a research zone. Drones go back to the drones’ base when fuel tank is empty, and participants have to refill research zones</td>
<td>Drones base</td>
</tr>
</tbody>
</table>
Performance of the sub-tasks in Table 1 was subject to temporal constraints:

- A detected vehicle had to be locked within 5 seconds following detection. Otherwise, it had to be detected once again.
- A locked vehicle had to be selected within 10 seconds following locking. Otherwise, it turned automatically from "locked" to "detected".
- An identified vehicle had to be neutralized within 100 seconds following its identification. Otherwise, it turned automatically from "identified" to "detected".

Scenario and objectives

The experimental session lasted 20 minutes and had two phases: Phase A of low difficulty (10 minutes) and phase B of high difficulty (10 minutes). We controlled the difficulty level by changing the rate of new vehicles and new messages. To limit the effects of the order in which the phases were presented (effect of learning, etc.) two similarly sized groups of participants were created. The first one performed phase A first, and the second group performed phase B first.

We communicated two objectives to all participants:

- **Reaching a minimum number of neutralized targets**: participants have to neutralize more than 25 vehicles during the mission.
- **Considering all priority targets**: For all messages (giving the location of priority targets), participants have to draw a zone according to the message in no longer than two minutes.

Data collection and processing

Measures and sensors

The system supporting SUSIE software is composed of a screen of 24” and a mouse connected to a PC. The devices used for data acquisition were an eye tracker FaceLAB5© for pupillary response, the log (text file) of scenario events (vehicle appearance, messages) and operator’s mouse actions recorded on SUSIE.
Using SUSIE’s log, one is able to compute indicators related to the density of stimuli and to operator performance:

- **Density level**: this variable characterizes the real-time task constraint, due to the density of informational stimuli relevant to dealing with the two objectives given to the participants. We defined density as the sum total of targets and messages that operators have not yet processed. This density varied from 0 to 35 stimuli across all participants and all experimental sessions, and we then categorized it into three density levels (low level: lower than 7 stimuli, medium level: between 8 and 15 stimuli, high level: higher than 16 stimuli).

- **Performance on message processing**: This is a binary indicator, computed every second. It decreases to 0 as soon as a message is not processed in time (i.e. if a new research zone is not created around the coordinates in the message during the 2 minutes following its appearance), and it stays or increases to 1 otherwise.

- **Performance on target neutralization**: This is a binary indicator, computed every second. It decreases to 0 as soon as a target is not neutralized fast enough (i.e. if time from first detection exceeds 2 minutes and target had still not been neutralized), and it increases to 1 otherwise.

- **Global Performance**: this indicator was computed as the mean of the two previous indicators.

In parallel to performance data, we also extracted eye movement variables that were related to task completion strategies using the Facelab 5.0 eye-tracking system:

- **Horizontal and vertical gaze concentration**: We defined this variable as the standard deviation in pixels (px) of gaze position on the X and Y axes (Wang, Reimer, Dobres & Mehler, 2014).

- **Proportion of Total Fixation Duration on each AOI**: It corresponds to the percentage of time spent to consult a specific AOI, based on fixation durations, as described in Masthoff, Mobasher, Desmarais & Nkambou (2012, p.132)

- **Transition rate between each pair of AOI**: We defined this variable as the number of gaze transitions per second from one AOI to another (Holmqvist, Nyström, Andersson, Dewhurst, Jarodzka & Van de Weijer, 2011, p. 424).
Density level was used to segment the experimental data over time. Each change in density level results in the ending of a temporal segment and the creation of a new one. The variables related to performance and eye activity were averaged on these segments.

**Experimental protocol: training and briefing**

The experiment was conducted in individual sessions of approximately 2 hours at the Lab-STICC laboratory and was divided into six phases:

1. *Greeting participants:* completing the profile questionnaire (age, video game experience)
2. *Explaining SUSIE principles:* a slide presentation was used to explain (verbally) the monitoring and drone management tasks and the interface. The experimenter provided the objectives of the mission that participants performed on the simulator.
3. *Practicing and training:* participants carried out the tasks to familiarize themselves with drone-swarm based monitoring activities.
4. *Sensors parameters and calibration setting:* Eye-Works Record software was launched to record the participants’ eye movements after faceLAB software had calibrated head, eyes, and test environment;
5. *Carrying out the monitoring tasks:* participants carried out the twenty-minute task scenario on SUSIE and completed a questionnaire at the end of the experiment;
6. *Debriefing and thanking:* We asked participants for their view of the set of proposed tasks, the interface, and the simulator in order to obtain their comments. To ensure inter-subject independence of the collected data, participants were asked not to share test contents with those around them.

**Experimental design and hypotheses**

We analysed the data from 17 participants after excluding 5 participants from the analysis due to intermittent failures in eye activity data acquisition. All were subject to the same scenario of twenty minutes, during which different constraint levels occurred. Our experimental design was thus a mixed factorial design of 17 Participants × 3 Constraint Levels. Table 2 presents the independent variables (related to informational constraint), and the dependent variables (related to performance and eye movement strategies). Statistical analyses of the data were conducted according to a General Linear Model.

**Table 2: Independent and dependent variables**

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Indicator of</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constraint Level</td>
<td>Informational density</td>
<td>3 ordinal classes: Low, Medium, High</td>
</tr>
<tr>
<td>Dependent variables</td>
<td>Indicator of</td>
<td>Range</td>
</tr>
</tbody>
</table>
### Performance on message processing

<table>
<thead>
<tr>
<th>Performance on message processing</th>
<th>Performance</th>
<th>Continuous variable, from 0 to 1 (mean=0.462; std=0.493)</th>
</tr>
</thead>
</table>

### Performance on target neutralization

<table>
<thead>
<tr>
<th>Performance on target neutralization</th>
<th>Performance</th>
<th>Continuous variable, from 0 to 1 (mean=0.523; std=0.498)</th>
</tr>
</thead>
</table>

### Global performance

<table>
<thead>
<tr>
<th>Global performance</th>
<th>Performance</th>
<th>Continuous variable, from 0 to 1 (mean=0.493; std=0.351)</th>
</tr>
</thead>
</table>

Where STCi is whether a subtask is completed or not when the sample is taken (yes=1 no=0) and N is the number of samples in the considered period.

### Gaze concentration

<table>
<thead>
<tr>
<th>Gaze concentration</th>
<th>Eye movement strategies</th>
<th>$\sum_{i=1}^{N} STC_i / N$</th>
</tr>
</thead>
</table>

**Horizontal concentration:**

- Continuous variable, from 0 to 852 (mean=323.7; std=171.1)

**Vertical concentration:**

- Continuous variable, from 0 to 537 (mean=210.9; std=104.2)

### Proportion of total fixation durations on each AOI

<table>
<thead>
<tr>
<th>Proportion of total fixation durations on each AOI</th>
<th>Eye movement strategies</th>
</tr>
</thead>
</table>

**Message zone:**

- Continuous variable, from 0 to 100 (mean=17.2; std=26.7)

**Drones base:**

- Continuous variable, from 0 to 100 (mean=10.0; std=19.9)

**Bombers base:**

- Continuous variable, from 0 to 100 (mean=5.5; std=16.8)

**User-defined research zone:**

- Continuous variable, from 0 to 100 (mean=58.0; std=36.7)

**Blank zone:**

- Continuous variable, from 0 to 100 (mean=9.3; std=23.1)

### Transition rate between AOI

<table>
<thead>
<tr>
<th>Transition rate between AOI</th>
<th>Eye movement strategies</th>
</tr>
</thead>
</table>

**Number of transitions/s between all AOIs:**

- Continuous variable, from 0 to 5.22 (mean=0.60; std=0.65)

**Number of transitions/s between each pair of AOI:**

Every 2-AOI transitions were also analysed in terms of frequency.

---

We posit the following five hypotheses with respect to the effects of constraint level on participants’ eye movement strategies and performance.
Effect of density level on eye movement strategies

Higher density of informational stimuli should result in attentional funnelling or perseveration on some task-related areas as a means to accommodate task demands. We therefore hypothesize that:

- H1: Gaze concentration will decrease when density of stimuli (messages and targets) increases.
- H2: Mean transition rate between all AOIs will decrease when the density of stimuli increases.
- H3: Proportion of total fixation durations on some AOIs will increase when the density of stimuli increases.
- H4: Transition rates between certain pairs of AOIs would increase when density of stimuli increases (this is because participants would give higher priority to some task areas and not to others).

Effect of density level on performance

- H5: We expect that a higher density of targets and messages will result in performance decrements.

Results

Effect of density level on eye movements

To evaluate the global gaze behaviour of participants we analysed their gaze variability and transition rate between AOIs. This context-independent analysis resulted in two main observations:

- H1 was partially supported, as we found a significant effect of the density of stimuli on horizontal gaze variability ($F(2,901)=20.657$, $p<.001$), but no significant effect on vertical gaze variability. As depicted in figure 4a, the gaze is more concentrated in the medium and high density levels ($M=294$px and $M=301$px, respectively) than in the low density level ($M=373$px). With respect to the lack of effect of density level on vertical gaze concentration, we believe that the reason for this may be that most activity remained in a large central, vertical strip on the display, whether density was lower or higher. This will be demonstrated in our analyses of fixation durations and gaze transitions.

- The effect of density on gaze concentration corresponded with the results from the analysis of transition rate between all AOIs (cf. figure 4b) and corresponded with H2. Overall, participants made fewer transitions per second between AOIs when the density of stimuli increased ($F(2,946)=5.783$, $p<.005$), dropping from $M=0.72$ transitions/sec in low density level to $M=0.51$ transitions/sec in the high density level.
To evaluate the eye movement patterns within task context we studied the proportion of total fixation durations on each AOI and the rate of transitions between each pair of AOIs of the five areas of interest we described in section 2.2 in the Method section.

Figure 5 shows a number of significant effects of the density level on fixation distribution and duration on AOIs. The statistical results highlighted three main observations:

- Participants focused relatively less on message zones when the density level increased ($F(2, 946)=4.84$, $p<.01$). Proportion of total fixation durations dropped from $M=21.8\%$ in low density level to $M=15.7\%$ and $M=15.9\%$ for medium and high density levels, respectively (cf. figure 5a).
eye activity as indicators of workload

CONSULTATION OF MESSAGES ZONE (5a)

CONSULTATION OF BLANK ZONE (5b)

CONSULTATION OF BOMBERS BASE (5c)
Consultation of blank zones - associated with identification of vehicles and consultation of bombers base associated with the neutralization of vehicles increased with density level (F (2,946) =5.74, p<0.005 and F (2,946) =7.05, p<.001, respectively). The proportion of total fixation durations increased from M=5.8% (low level) to M=12.1% (high density level) in blank zone, and from M=3.6% (low level) to M=8.4% (high density level) in bombers base (cf. figures 5b and 5c).

Finally, participants focused relatively less on user-defined research zones (cf. figure 5d) when density level was higher (F (2,907) =4.82, p<0.001) (M=54%, M=59.5% and M=63.7% for high, medium and low density levels, respectively).

Thus, in accordance with H3, higher density of stimuli led participants to focus more on certain task areas and less on other, with one exclusion - we found no significant effect of the density level on the distribution of fixations on the drones’ base. This is probably because neither when the density of stimuli was lower nor higher, was the monitoring of drone state a highly frequent sub-task.

Findings from the analyses above corresponded with a frequency analysis of AOI consultation and corresponded with H4:

- Transition rate between drones base and user-defined research zones decreased when density level increased (F (2,907) =4.936, p<.01), varying from M=0.123 transitions/s in low density to M=0.062 transitions/sec in high density.
- Participants executed less transitions between message zone and user-defined research zones when density level was high (F (2,907) =9.45, p>0.001). The number of transitions per second varied from M=0.121 (low level) to M=0.060 (high level).
No significant effect of the density level on transitions between the other pairs of AOIs was found.

![Transition Rate Graph](image)

**Figure 6**: Transition rate between drones base and user-defined research zones (6a), and between messages zone and user-defined research zones (6b). Vertical bars denote 0.95 confidence intervals.

### 3.2. Effects of density level on performance

Different indicators were analysed: participants’ performance on message processing (figure 7b), target neutralization (figure 7c) and on the sum total of the two former objectives (figure 7c). Two main observations were made:

- In accordance with H5, density level had a significant effect on global performance (F(2,946)=8.532, p<.001). Global performance decreased when density level increased (M=0.55, M=0.46 and M=0.44 in low, medium and high density levels, respectively).
Figure 7: Performance rate for both message and target processing (7a), for message processing (7b) and for target neutralization (7c). Vertical bars denote 0.95 confidence intervals.
- This decrease of global performance is mainly explained by a decrease of performance on message processing ($F(2,946)=21.13, p<.001$) ($M=0.59, M=0.35$ and $M=0.41$ for low, medium and high density levels, respectively). To the contrary, no significant effect was found on the performance on target neutralization ($F(2,946)=1.976, p=.139$).

**Discussion**

Our analyses of operators’ eye activity patterns and performance supported the main hypotheses. It was found that when the density of targets increased the overall variance in gaze positions along the X-axis was lower for medium and high compared to lower target density levels (Figure 4). This finding corresponded with a finer grained analysis showing that operators made fewer transitions between display areas and concentrated for relatively longer periods on certain areas and not on others. It was also found that performance was lower when the density of targets increased. This finding suggested that MWL was able to be manipulated and that changes in gaze behaviour were probably associated with different task completion strategies in response to changes in MWL. A deeper look into the eye activity patterns in terms of task dependent areas revealed how operators chose to deal with higher levels of MWL. When target density increased, operators reduced the proportion of time they spent on inspecting messages and by extension, the proportion of time they spent on inspecting self-defined research zones that one defines based on messages. In contrast, operators increased the proportion of time that they spent on inspecting blank zones and the bombers’ base that they needed for eliminating targets (Figure 5a and 5c). Performance measures corresponded with the strategy that eye activity measures reflected as greater number of targets led operators to losing points for not dealing with messages but not for failing to eliminate the targets that they had already detected (Figure 7).

The findings demonstrated the importance of context dependent interpretation of eye activity measures. It could be hypothesized that operators would invest different times in display areas when confronted with different levels of MWL, yet it was difficult to detail what areas they would inspect more or less carefully. The rate that operators could lose points for not dealing with messages and for not eliminating the detected targets was similar. Either way, score would drop down to zero every two minutes without action. Operators opted to prioritize the elimination of targets, probably because it made more sense to do so given the context of the task (securing a perimeter). In future studies, we will explore whether eye activity patterns would be sensitive enough to detect different strategies operators might employ in response to different rates of losing points (e.g., every 20 s for not dealing with messages and every two minutes for not eliminating the detected targets).

Eye activity measures corresponded with performance measures, showing their validity as a means to learning about task completion strategies in response to changes in MWL. At the same time, the correspondence between measures might raise the question why using eye activity measures in the first place and not just inferring about different levels of MWL and their consequences from operator performance? In this respect performance may sometimes remain unchanged even
when MWL increases, if people succeed to counter higher task demands by investing more effort. It is thus necessary to use a combination of measures when investigating MWL (Yeh & Wickens, 1988). Further, we computed average scores across 22 participants to detect a drop in performance between the three target density levels that we defined in the Method section. However, performance of individual operators may drop in response to lower/higher densities than the ones we defined but it would be impossible to anticipate this. Eye activity patterns, on the other hand, may be used as precursors of a later drop in performance. In future studies, with larger number of participants, we intend to test whether the changes that we described in the relative time operators spend in different display areas can indeed anticipate a later drop in performance. If this is indeed the case, then eye activity patterns may possibly serve as the basis of on-line algorithms to detect short periods of elevated MWL of single operators and trigger automatic or human assistance for those that experience too high task demands.

5. Conclusions

Different eye activity patterns can be detected in response to different levels of target density in a simulated drone-operating task. These patterns corresponded with operators’ scores in the task, suggesting that eye activity measures can be used to detect short periods of elevated MWL and changes in task completion strategies. The findings, therefore, constitute a platform for further investigation of the practical usage of eye activity measures in work environment. In the future, we intend to investigate the sensitivity of the indicators of MWL that we described for detecting short periods of elevated MWL of individual operators as a means to triggering automatic or human assistance in tasks.

References


