5-1-2021

The Effects of Increased Visual Information on Cognitive Workload in a Helicopter Simulator

Reilly Innes

Zachary Howard

Alexander Thorpe

Scott Brown

Ami Eidels

THE EFFECTS OF INCREASED VISUAL INFORMATION ON COGNITIVE WORKLOAD IN A HELICOPTER SIMULATOR

Reilly Innes[1], Zachary Howard[2], Alexander Thorpe[1], Scott Brown[1] & Ami Eidels[1]

[1]University of Newcastle, Newcastle, Australia
[2]University of Western Australia, Perth, Australia

Workload in highly demanding environments can be influenced by the amount of information given to an operator, and consequently, it is important to limit the potential overload. In the current study, we used the Detection Response Task (DRT) to assess the effects of enhanced heads-up display information ("symbology") on cognitive workload in a simulated helicopter environment. Participants (highly trained military pilots) completed simulated helicopter flights, which varied visual conditions and the amount of information given. During these flights participants completed a DRT. With increased heads-up display information, pilots landing accuracy improved across visual conditions. The DRT captured the increased workload resulting from the varying environmental conditions, and provided evidence for heads-up display information having negligible effects on workload. Our study shows that the DRT is a useful workload measure in simulated helicopter settings. We also show that the increased level of symbology appeared to assist pilots flight behaviour and landing ability, without compromising safety. This research highlights that a) the DRT is an easily implemented and effective measure of cognitive workload in a variety of settings and b) the potential for further cognitive workload evaluative methods in similar aviation and applied settings.

In modern society, we have seemingly unlimited sources of information available, however, processing, integrating, and using this information is difficult. For example, in helicopters, pilots have constant access to an array of flight metrics such as speed, altitude, roll and more, all intended to assist the pilot. The primary difficulty in using information is the limits of human attention and perception, which are commonly overlooked.

Driver distraction research has been used to highlight the limits of human attention and the potential consequences of such distraction on cognitive workload (Strayer & Johnston, 2001). Cognitive workload refers to the overall level of cognitive demand placed on an individual from a task/s (Lee et al., 2008; Innes et al., 2020). Cognitive workload includes demands related to the number of tasks at hand, the difficulty of those tasks, associated time pressures and the overall mental and physical effort exerted (Hart & Staveland, 1988).

The detection response task (DRT) is a behavioural measure of cognitive workload, which requires individuals to detect a salient stimulus (light or vibration) and respond as quickly as possible, whilst concurrently completing another task (such as driving; Engstrom et al., 2005; Strayer et al., 2013). There are various methods of measuring cognitive workload, such as subjective questionnaires, eye tracking, and heart rate variability monitoring, however, these fail to provide performance based results, which are important if main task performance is difficult to quantify (Innes et al., 2020). The DRT is easily applied to simple psychological task designs or more complex real world designs. For example, the DRT has been used extensively for simulated and real-world driver distraction studies to assess the level of workload induced from mobile phones, conversations with passengers and smart assistants (Strayer et al., 2013; Strayer et al., 2019). Response times from the DRT give an indication of cognitive workload, as fast responses indicate more available cognitive resources.

Previous cognitive workload studies have shown that errors and inferior task performance are more likely when cognitive workload is high. In a multitasking environment, cognitive demands can be difficult to assess, as adding items to process or increasing task difficulty can lead to a depletion of cognitive resources. When resources are low (i.e. workload is high), errors are more likely. Yet additional sources of information could also lead to redundacy gains, so performance may increase. In helicopter piloting and interface design, maximising informative assistance to aid performance whilst minimising distraction is vital. In this scenario, poor performance has critical consequences, and so needs to be optimised without overloading pilots.

Recently Airbus and Hensoldt have developed state-of-the-art sensor systems, which allow more information to be available to pilots. This information includes 3D mapping of the environment, clearly identified hazards and landing guides – all of which can be displayed in a pilots heads-up display (HUD). In the current study, we aimed to assess the cognitive workload induced under varying amounts of HUD information using the DRT. Highly trained helicopter pilots completed a simple flight path and landing in differing visual conditions, and with differing levels of HUD information. Both flight results (landing execution and flight metrics) and DRT results were assessed to evaluate the impact of HUD information on flight performance and workload. It was hypothesized that increased HUD information would lead to greater flight perfomance, however, we also hypothesized that cognitive workload would similarly increase.

Method

Participants

The participants' were limited to three pilots due to the highly specific requirements of the task. Thus, we ran a small-n study (Smith & Little, 2018) with high repetition to maximise availability of data. All three participants were highly trained helicopter pilots, each with more than 4,000 flying hours experience and extensive simulator experience. The three pilots completed the design seven, five and three times respectively. The uneven number of trials per participant was due to pilot availability, although no pilot varied greatly in DRT results or flight results between iterations.

Materials and Design

Data was collected in an Airbus MRH90 Taipan Multi Role helicopter simulator. The simulator incorporated three partially overlapping screens which made up 200◦ x 40◦ field of vision. The participant sat at a distance of approximately two metres from the screen. Controls in the simulator included a collective shaft, cyclic shaft and two foot pedals. The participants were shown an electronic map and a multi-function display, which indicated altitude, ground speed, collective power and helicopter roll. Participants were also fitted with a headpiece which was placed over the participants eyes. The headpiece acted as goggles, so that the participant could still see the simulator. In conditions where symbology was added, additional information was overlaid in their visual field. The location and angle of the headpiece was tracked at high rate so that information projected into the visual field mapped accurately and dynamically onto the visual environment.

Cogntive load was assessed via a DRT device, closely adhering to ISO 17488 (2016). The DRT device included a vibrating pad, which was taped to the participant's skin near their shoulder, and a response button, which was attached to the collective shaft nearest to where the pilots thumb sat. With an already crowded visual environment, we proposed the use of the tactile DRT to limit visual competition.

Each participant completed two simultaneous tasks – the flight simulation and DRT. For the DRT, a short stimulus was elicited via a vibration. The participant was required to respond via the response button to each iteration of the stimulus. The stimulus lasted for one second (or until the response button was pressed, whichever came first). The DRT stimulus was elicited at an interval of 3 - 5 seconds and occurred for the duration of each simulated flight. For the full DRT method, see ISO 17488 (2016).

	NO SYMBOLOGY (∅) Headpiece off	**2D MINIMAL (2D)** *LIDAR off	**3D MAX SYMBOLOGY (3D)** *LIDAR on, All symbology.
DAY VIS = 12000, TIME = 1600 DUST = OFF	DAY ∅ **A**	DAY 2D **D**	DAY 3D **G**
NIGHT VIS = 12000, TIME = 2000 DUST = OFF, FLIR = (ON) FLIRTIME =2000, FLIRVIS= 2400	NIGHT ∅ **B**	NIGHT 2D **E**	NIGHT 3D **H**
DUST VIS = 1200 TIME = 1600 DUST = ON	DUST ∅ **C**	DUST 2D **F**	DUST 3D **I**

Figure 1: Details of each tactical approach flown, with conditions of symbology and environment shown. There were two additional conditions where the DRT was not present – these were the 3D and no symbology Night conditions.

The flight simulation involved participants undertaking a short predetermined flight path and subsequent landing. There were three conditions of visual environment: High Visibility (Day), Low Visibility (Dust) and Night. There were three conditions of symbology: no symbology, 2D symbology and 3D. A full summary of conditions can be seen in Figure 1. The 2D symbology condition was made as similar as possible to the standard heads-up display used by military helicopter pilots in modern large-platform helicopters. The 3D symbology condition contained extra information, and the no symbology condition contained less. For an example of the three symbology conditions, see Innes et al., (2020). Two baseline conditions were added where participants completed the flight without the DRT in the Night condition *without* symbology and the Night condition *with* 3D symbology. The experiment was thus a 3x3+2 within subjects design (see Figure 1).

Procedure

All three participants were familiar with the simulator environment, and were given instructions about the DRT. The DRT commenced as soon as the pilot lifted the collective shaft for each condition. The flight path was identical for all 11 conditions. Participants were instructed to take off and fly for around one minute towards a mountain, where they would then complete a horseshoe turn at a deignted gate point. Following this the participants were instructed to begin descending to the Landing Zone, which was in the centre of a sports field.

Participants completed all 11 conditions. The order of the nine DRT conditions was randomized and the remaining two baseline conditions were completed following the corresponding DRT-active conditions. If, during a flight trial, the participant crashed or there were any technical issues, the run was restarted. Responses in these trials were recorded separately. Participants were given short breaks between flight trials and long breaks between blocks of flight trials. All flight data was recorded. DRT response times and misses were recorded.

Results

Several flight metrics were used to evaluate the quality of flight for each trial. These metrics included flight path variability measures and landing data. For brevity we report here only the latter. The main reason to evaluate landing/flight quality parameters was to ensure there was no task trade-off between the flight and the DRT. DRT response time and misses were analysed. Flights where the participant crashed were removed. For each metric we completed Bayesian repeated measures ANOVAs to measure effects of environmental conditions, symbology conditions and the interaction. All analysis was completed using the statistical program JASP (JASP Team, 2019).

Flight Metrics. We assessed the accuracy of landing data by borrowing appropriate precision measures from ballistic sciences. Participants were instructed to land at a specified and marked point in the virtual environment (centre of a football field). We measured the absolute distance from this landing zone (LZ) to the actual landing location ("landing error") and the "circular error probable" (CEP), which is the median error radius (Nelson, 1988, p.1). CEP results are shown in Figure 2.

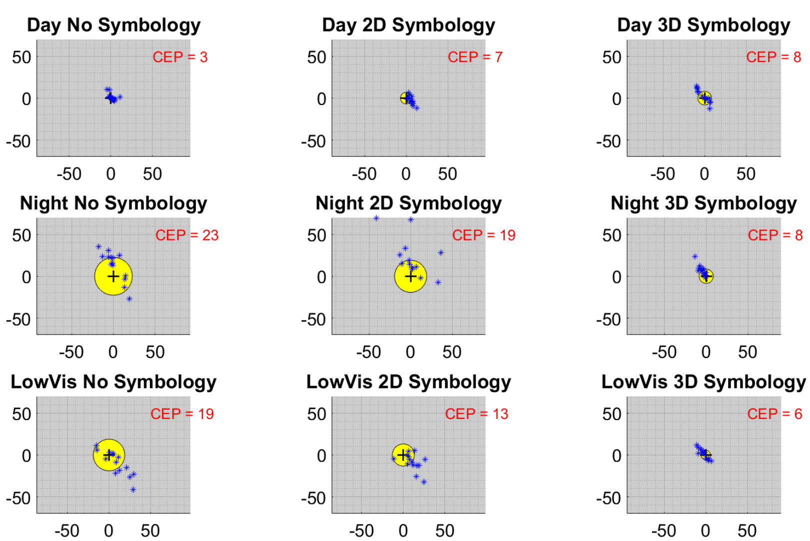

Figure 2: CEP plots for eah flight condition. The grid shows visual conditions as rows and symbology levels as columns. The yellow circle (which is quite small in some plots), shows the CEP for the landings. Each blue dot represents an individual flight landing.

A Bayesian repeated measures ANOVA on distance from the LZ (for each block of landings) revealed strong evidence for main effects of environment and symbology, and for their interaction effect (all BF_{10}>1000). Post-hoc analysis showed that landings in the night and dust conditions were significantly worse than the day condition (Night vs Day; BF_{10}= 137.78, Dust vs

Day; BF_{10}= 62.95). Landings completed in the 3D symbology condition were also much closer to the designated LZ than with no symbology (BF_{10}= 65.32) and 2D symbology (BF_{10} = 14.39). An interaction effect was also shown, such that the 3D symbology appears to be unaffected by the environmental condition, whereas no symbology and 2D symbology landed closer to the LZ in day conditions, but did much worse in the night and dust conditions. Figure 2 and the ANOVA results suggest that landings completed in the Day condition were most accurate. Furthermore, symbology was shown to have the greatest effect on landings, with 3D symbology landings more consistently closer to the designted landing spot than other symbology conditions.

DRT. Figure 3 shows mean DRT response times across flight and symbology conditions. A two-way repeated measures Bayesian ANOVA of log-transformed RT showed a preference for the model that included the effect of visual condition when observing differences across symbology and time of day (BF_{10}= 4.909). A comparison of visual conditions showed some evidence for a difference between the Day and Night conditions (BF_{10}= 3.335), and good evidence for a difference between Day and Dust conditions (BF_{10}= 6.842). A comparison of symbology conditions showed ambiguity for a difference between the conditions. A two-way repeated measures Bayesian ANOVA of misses showed a preference for the model which included symbology, environment and the interaction, with positive evidence in favour of the null (i.e. evidence for no difference between conditions of symbology and visual conditions; BF_{01}= 3.844). These results should be interpreted with caution, as the small sample size prevents reliable statistical testing.

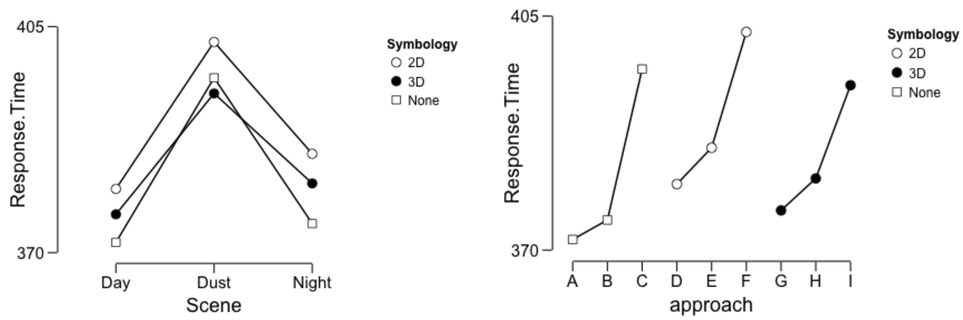

Figure 3: Left; Mean DRT response times for the visual conditions. Right; Mean DRT response times for the individual approaches. Data points in both panels are grouped across symbology conditions (2D, 3D, none) for ease of interpretation.

Discussion

Results indicated that landing performance declined in more difficult conditions (i.e. conditions with lower visibility and conditions with higher symbology), as expected. This was clear in the analysis of flight landings, but performance was difficult to distinguish through variability metrics. The variability metrics did however provide a descriptive analysis of the flight, where in low visibility and low symbology conditions, pilots tended to fly lower and slower. Importantly, it was shown that despite a difficulty increase from the degraded visual environment, workload was relatively unaffected by the symbology. There was no difference in workload, as measured by DRT responses, across the three symbology levels (none, 2D, 3D). However, this is not due to lack of senstivty as the DRT was sensitive enough to detect a change in workload between the Day and Dust conditions. This scenario appears ecologically valid, as pilots in brown-out experience degraded visual conditions and higher amounts of flight errors.

These results highlight the importance of workload evaluation, the sensitivity of the DRT to change in workload and the usability of 3D symbology for flight assitance in night and dust conditions. 3D symbology was especially useful in degraded visual conditions, but added little to performance in high visibility conditions. This is an important finding that highlights the need for adaptive user interface, where extra information may be useful in some conditions, but not others. Further, the additional information appeared to come at no extra cost to the pilots workload. There was a difference between 2D and 3D symbology, which could indicate that the 2D symbology was less intuitive than the 3D symbology, potentially using extra attentional resources or obscuring the field of view (in comparison to No symbology and 3D symbology). This finding is promising for implentation of 3D symbology to deliver increased information to assist pilots.

There were some practical limitations to the current experiment. First, each flight path was relatively short (around 3 minutes), which limited the DRT to around 40 trials. Further, the difficult part of the flight - the landing - only lasted around 30 seconds of the total time. Whilst this highlights the sensitivity of the DRT across such limited number of trials, more subjects and trials over a greater and more variable flight path (where data from specific sections was collated) could provide an avenue for future research.

Acknowledgements

This research was supported by a University of Newcastle Industry Linkage Pilot Grant to SB and AE. We thank Airbus and Hensoldt for their in-kind support.

References

Engström, J., Aberg, N., Johansson, E., & Hammärback, J. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems. University of Iowa

Hart, S. G., & Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139–183). Elsevier.

Innes, R. J., Evans, N. J., Howard, Z. L., Eidels, A., & Brown, S. D. (2020). A broader application of the detection response task to cognitive tasks and online environments. *Human factors.*

JASP Team. (2019). JASP (Version 0.11.0)[Computer software]. Retrieved from https:// jasp-stats.org/

Lee, J. D., Young, K. L., & Regan, M. A. (2008). Defining driver distraction. *Driver distraction: Theory, effects, and mitigation*, 13 (4), 31–40.

Strayer, D. L., Cooper, J. M., McCarty, M. M., Getty, D. J., Wheatley, C. L., Motzkus, C. J., . . . Horrey, W. J. (2019). Visual and cognitive demands of carplay, android auto, and five native infotainment systems. *Human Factors*, 61 (8), 1371–1386.

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). Measuring cognitive distraction in the automobile. AAA Foundation for Traffic Safety.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. Psychological science, 12 (6), 462–466.

Wickens, C. D., Hooey, B., L., Gore, B., F., Sebok, A., & Koenicke, C. S. (2009). Identifying black swans in NextGen: Predicting human performance in off-nominal conditions. *Human Factors, 51*, 638-651. doi: 10.1080/10508410802597382