Airbus-Hensoldt-UoN - Tech report 2018-2:

Cognitive Workload Measurement and Analysis

Reilly Innes, Zachary Howard, Ami Eidels, & Scott Brown

University of Newcastle, Australia

Contact: Ami.Eidels@newcastle.edu.au

In collaboration with Airbus & Hensoldt

Airbus-Hensoldt-UoN Tech report 2018-2 - Executive Summary

Background: Cognitive workload is of critical importance when operating sophisticated systems. In helicopters, pilots need to be able to process information from their avionics systems and from the external the environment, to fly certain flight paths and to ultimately complete the mission and return safely. When an individual is under high cognitive workload, they are more likely to make errors, or show a lower performance in the tasks they are performing (Kahneman, 1973). This is a critical issue, as flight conditions such as night, low altitude, or brownout may prove critical to the success of certain operations, yet may also impair processing by increasing cognitive workload. Hensoldt™ in co-operation with Airbus™ have developed a unique solution in a symbology system - which feeds sensor driven information directly to the pilot's visor – which may reduce cognitive workload in difficult flight conditions. Two reports outline a study that assessed the effectiveness of the symbology system.

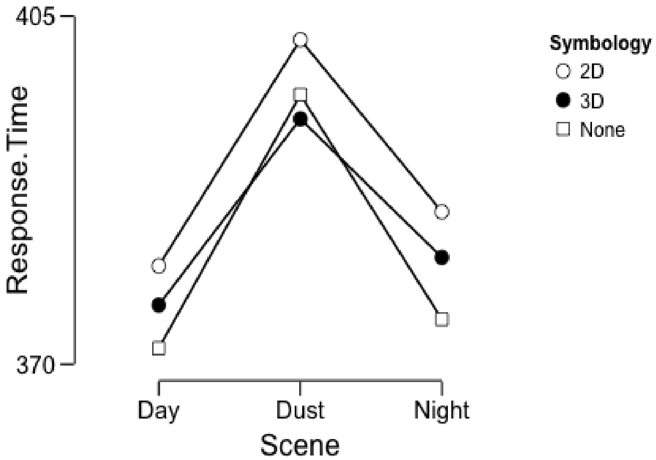

Figure 1. Mean DRT Response Time. 3D symbology does not increase workload.

Key findings: Response Time analysis from the Detection Response Task (DRT) showed that 3D symbology caused little-to-no addition to cognitive workload (Fig 1). Flight-path analysis from simulator data showed 3D symbology lead to improvements in flight performance, most particularly in landing accuracy (Fig 2). Taken together, the results suggest 3D symbology enables pilots to access more information from the environment and sensors, and demonstrably perform better on mission objectives, without increasing their level of workload. Like any field study there were limitations. Future endeavours will aim to more reliably address flight metrics other than landings by defining more concrete mission parameters (altitude/speed gates etc.).

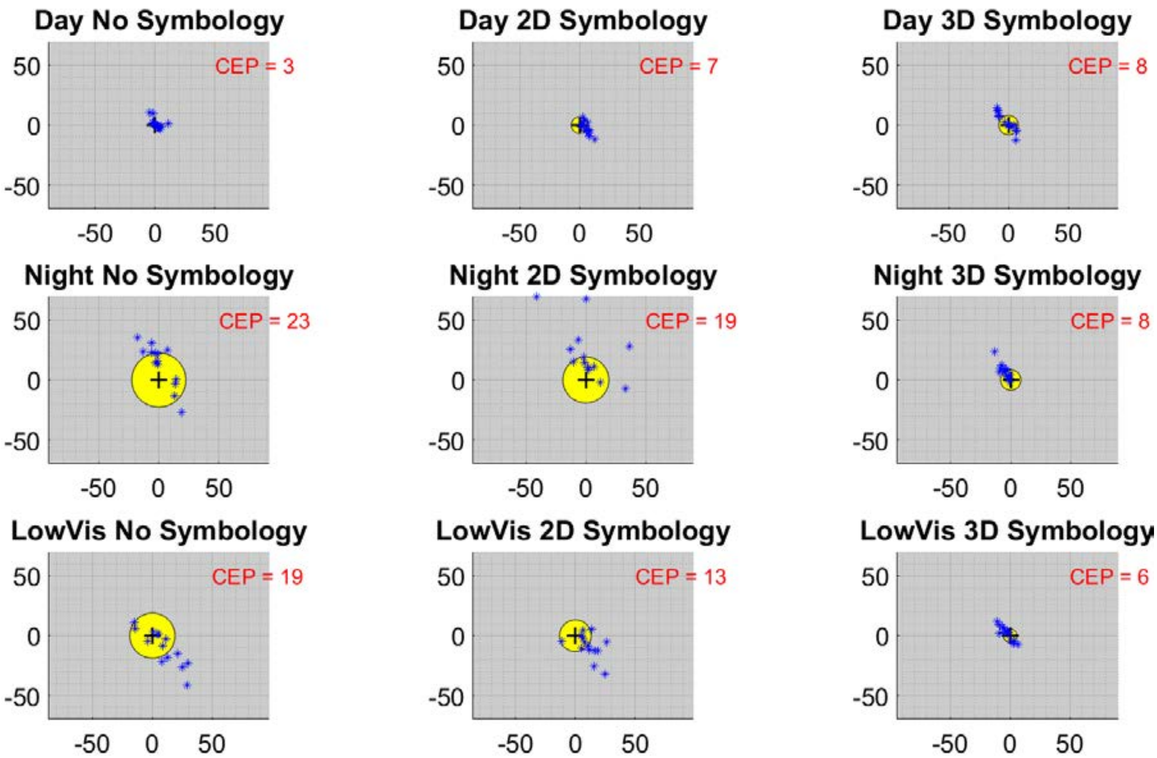

Figure 2. Landing performance in each flight condition, with (0,0) representing the designated target landing zone. Performance in 3D symbology (right column) was consistent regardless of visual environment.

As working environments of sophisticated systems grow in complexity, it is essential that designers consider the amount of cognitive demand placed on operators (Marshall, 2002). In helicopters, pilots need to be able to process information from their surrounds as well as information from the vessel in order to fly certain flight paths, routes and landings to ultimately safely operate and preserve the aircraft. Pilots may often be asked to fly certain flight profiles and routes, and this is especially the case with tactical troop transport vehicles that experience constant take offs and landings throughout a mission. Due to these increased demands the pilot may experience an increased cognitive workload to the point where their ability to operate the vehicle is impaired or they are prevented from managing the load elicited by any additional tasks.

Prior studies have shown that when an individual is under a higher cognitive workload, they are more likely to make errors, or show a lower performance in the tasks they are performing (Kahneman, 1973; Strayer & Johnston, 2001). This stems from the psychological concept of *capacity*, which is limited due to individuals only having a certain amount of attentional resources to allocate to their current tasks (Kahneman, 1973; Broadbent, 2013). The *Detection Response Task* (DRT) has been used extensively in driving settings as a measure of cognitive workload (Strayer & Johnston, 2001). Studies using the DRT have shown the difficulty associated with driving and texting, talking on a mobile phone, conversing and doing simple cognitive tasks (Strayer, Watson, & Drews, 2011; Strayer et al., 2013, 2015). The DRT involves detecting and responding to a short stimulus, such as a light or a vibration and runs for the duration of each flight. The DRT acts as a secondary task to measure the load of the main primary task. The DRT task is simple so adds little to no load to the main task and is used to assess the residual capacity from the main task.

Like driving behaviour, we expect helicopter pilots to possibly experience cognitive deficits when operating in information-heavy environments. Heightened levels of cognitive load may be the result of environmental conditions and/or information they perceive from aviation instruments and projections to their headpiece. Symbology projected onto their headset could provide information about the environment, landscape, obstacles and landing zones, helping in some way but also increasing the processing load demands. Additionally, research has shown that the perceptual load incurred by visual targets has an effect on the processing of information in the periphery of a display (Handy et al., 2001). This means that although this symbology is intended to aid the pilot in flying and landing the aircraft, it may add significant cognitive workload to the pilots' current working state, be ignored by pilots or even cause greater distraction from their main task. The project assess this effect.

Flight stimulators have become increasingly common as a method of training for aviation professionals, in both novice and expert pilots (Allerton, 2010). They have sound ecological and criterion validity, and at the same time are much cheaper and involve far less risk than flying in a live aircraft. Modern simulators are sophisticated enough to incorporate symbology into headsets and fly in simulated maps that are derived from real-world locations and may depict real-world combat situations. Using the DRT, we evaluated the effects of sensor technology and symbology on helicopter pilots' workload. Specifically, with Airbus and Hensoldt we tested the effects different levels of symbology have on cognitive load across environmental conditions in a helicopter simulator.

Experiment 1 - Method

Note: This section is the same as Tech Report 2018-1 Method section.

Tasks

Participants (pilots in the simulator) performed two tasks simultaneously: they had to fly the helicopter from a given starting point A, via a designated gate B, to a designated landing zone (LZ).

Flight: The flight task involved participants operating a helicopter simulator, which was modelled on an Airbus MRH90 Taipan Multi Role Helicopter. The simulators controls included a collective shaft, cyclic shaft and two foot pedals. The participants were shown a map and a multi-function display, which indicated altitude, ground speed, collective power and helicopter roll. Participants were also fitted with a headpiece placed over their eyes. The headpiece acted as goggles, so that the participant could still see the simulator, however, in conditions where symbology was added, additional information was overlaid in their visual field. The location and angle of the headpiece was tracked so that information projected into the visual field mapped accurately onto the visual environment. The simulator room included a 180° wall where projectors displayed the simulated environment. The participant sat at a radius of approximately 2 metres from the screen.

The flight task involved participants taking off from a designated starting point, flying up to a waypoint, turning and then descending into a landing point. The waypoint was approximately the mid point of the total flight. The average flight took 2-4 minutes.

Table 1 summarises the testing conditions.

	NO SYMBOLOGY **(∅)** Headpiece off	**2D MINIMAL** **(2D)** *LIDAR off GND SPD, RAD ALT, LZ DST, LINE	**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**

	NIGHT <u>NO DRT</u> ∅ **BX**		NIGHT <u>NO DRT</u> 3D **HX**

Table 1: Details of each tactical approach flown, with conditions of symbology and environment shown.

There were three conditions of *Degraded Visual Environment* (DVE): High Visibility (Day), Low Visibility (Dust) and Night. In all conditions, traffic was off, wind speed was set at 5km/h and weather was set to have no cloud or rain. The only parameters that were varied were visibility (distance in meters), time of day, dust (on or off) and FLIR (on or off). In the Day condition, visibility was set at 12,000m, time of day was set at 16:00, FLIR and dust were off. In the Night condition, FLIR was on and was set at 20:00 with FLIR visibility at 2,400m. General visibility in this condition was set at 12,000m, time of day was set at 20:00 and dust was off. In the Low Visibility, Dust condition, the dust appeared at 100m from the ground. Visibility in this condition was set at 1,200m, time of day was set at 16:00, dust was on and FLIR was off.

There were three conditions of symbology: no symbology, 2D symbology and 3D. In the no symbology, the headpiece remained fixed to the participants but displayed no information in the day or dust conditions. In the Night-no symbology condition, FLIR info was projected in the headpiece, with no additional symbology. In the 2D symbology condition, LIDAR was off, ground speed, radial altitude, location zone distance, and helicopter position were shown, as well as basic indicators for the waypoint and landing zone. In the 3D symbology condition, all symbology was shown, as well as the LIDAR. This included a grid over the environment, contours, landing zone information, horizon line and helicopter position.

The three symbology conditions were crossed with the three visibility conditions to yield a matrix with nine experimental conditions, as shown in Table 1. A further two control conditions were added where participants completed the Night condition with no symbology and the Night condition with 3D symbology, without the DRT task, making a total of 11 conditions.

DRT: The DRT implementation adhered to ISO 17488 (2016) guidelines. A finger response button was attached via a velcro strap to the collective shaft on the flight simulator. The collective was located to the left of the pilot and was easily operated via pressing the thumb down. It did not interfere with any other simulator controls. The button was used to respond to the DRT stimuli. A tactile vibration module was fixed to the skin of each participant's shoulder on their right side, using surgical tape. The tactile vibration module elicited a short vibration (lasting 1 second, or until the 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 flight.

Responses entered before the onset of the next vibration stimulus were deemed ``hits'', and a failure to respond within this interval was deemed a ``miss''. Second (and subsequent) responses entered before the onset of the next stimulus were deemed ``false alarms''. Response time was measured as the time between the onset of the vibration stimulus and the pressing of the switch.

Procedure

The participant was seated in the centre of the room and, given all had used the simulator before, no additional instructions on simulator operation were required. Two experimenters sat behind the participant so that they were out of the participant’s field of view. One operated the DRT, which commenced at the start of each flight and ended as soon as participants had landed and stopped movement. The second experimenter read out the assigned conditions parameters before the commencement of each flight.

Participants were introduced to the simulator and shown the flight path, waypoint and landing zone. Participants were then able to practice flying the designated flight path and landing without the presence of symbology. They practiced the flight up to five times, under different conditions of symbology at their discretion.

Following the practice flights, the DRT was attached to the participant and they were given three practice flights of responding to only the DRT without any flying. Once they were comfortable with the equipment and understood the task, participants were randomly assigned an order of conditions. Conditions with no DRT were always grouped with their corresponding condition of DRT presence, but were randomly ordered so that either DRT or no-DRT could occur first in the sequence.

The participants were instructed to commence the flight after the first elicitation and response of the DRT. Participants then flew the designated flight path. In the instance that the pilot overshot the landing zone or crashed, the data was still recorded and marked as “overshot” or “crash” in the flight log file. The flight was then restarted so that the pilot flew the condition again. This meant that in the conditional analysis, where crashes and overshoots are excluded, a full set of conditions was still recorded. A full block of 11 (or more in the case of crashes or overshoots) flights took between 30-50 minutes. Participants were given short breaks between flights and longer breaks between blocks to minimise fatigue.

Results

The average response times for conditions of successful and unsuccessful landings are shown in Figure 1. Mean response time was significantly higher in the unsuccessful landings than in the successful landings. Due to the significant impact on results from a limited number of unsuccessful (crash) flights, these flights are excluded from the following analysis (further analysis has been conducted on the full data and crash specific data).

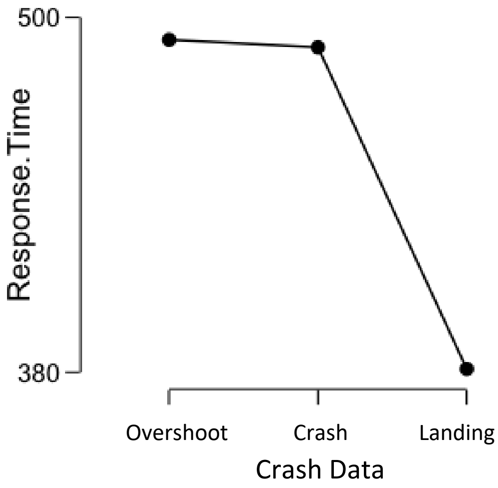

Figure 1: Average DRT response times for flights where overshoots – times when the pilot missed the landing zone, crashes, and successful landings were observed. Higher response times indicate higher cognitive workload.

Figure 2 shows the average DRT response times for all participants across each flight scenario. There is a clear trend which highlights that pilots in the DVE Dust condition showed the longest response times to the DRT signal, indicating the highest amount of cognitive workload. This was supported by a Bayesian ANOVA, with the Dust conditions significantly different to the Day and Night conditions. Pilots showed similar average response times across all levels of symbology, indicating that symbology had little to no negative effect on cognitive workload. However, a trend can be observed in 2D symbology conditions in which pilots always showed the highest workload relative to the environmental condition. A Bayesian ANOVA showed no significant difference in average response times across symbology. Furthermore, combining results from symbology and environment, it is evident that an interaction effect is present. Certain flight scenarios were shown to be more difficult than others, as seen in Figure 3, where scenario A and G show low cognitive workload, whereas scenario E shows a much higher cognitive workload.

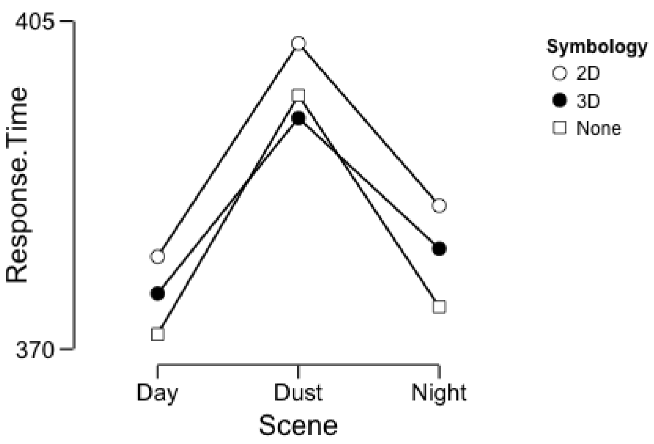

Figure 2: Mean DRT response time for each condition of environment and symbology with unsuccessful landing data excluded. Longer response times indicate higher cognitive workload. 3D symbology did not increase workload for pilots, as the time to detect the DRT signal was similar in the no- and 3D-symbology (and both were faster than 2D).

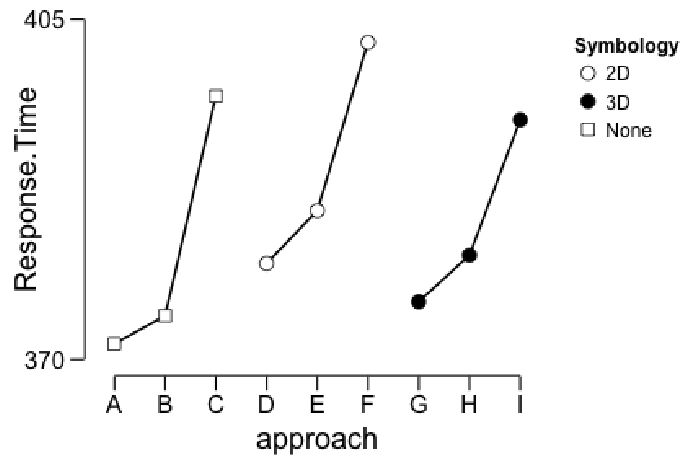

Figure 3: Mean DRT response time for each approach, split by condition of symbology. For further details about different approach types see Table 1. Higher response times indicate higher cognitive workload.

Whilst average response times answer some questions regarding the effectiveness of symbology and the impacts of DVE, it is also important to analyse the amount of lapses which occurred during each flight. Lapses were defined as responses that occurred after 500ms from the stimulus onset and included misses. Figure 4 and Figure 5 show lapses across conditions, with the data showing similar trends to response time analyses. In this analysis, it is more evident that the 2D symbology condition had the greatest effect on the amount of pilot lapses. This may be due to the symbology condition adding minimal assistance when searching for waypoints and landing zones or being more distracting, meaning less residual cognitive load was available. Furthermore, the Dust and Night DVE conditions had a greater amount of lapses than the Day condition, highlighting that clear day time flying was more manageable in the simulator than the other more difficult conditions. This was an expected result and highlights that the experimental manipulation of DVE was successful. These results are supported by Binomial tests indicating a significantly greater than expected frequency for the 2D symbology condition and a significantly lower than expected frequency for the Day DVE condition. Furthermore, scenario C (Dust, None), E (Night, 2D) and F (Dust, 2D) showed high frequencies of lapses. These results are important since lapses in cognition, induced by high cognitive workload in specific conditions, may explain several of the current findings. We consider lapse data to be of critical importance in our study, as lapses induced by heightened cognitive workload in the field could have significant consequences.

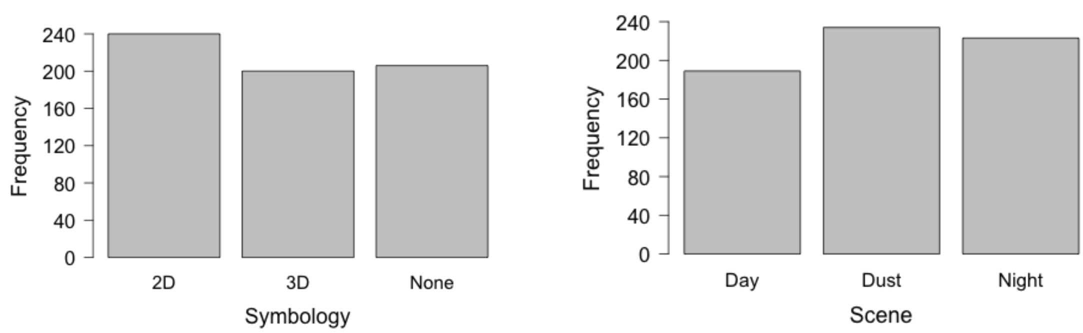

Figure 4: The left panel shows the number of lapses for each symbology condition. The right panel shows the number of lapses for DVE conditions. Longer bars indicate a greater amount of lapses (i.e. worse performance). There were fewer lapses of the DRT signal in the no- and 3D-symbology conditions than in 2D.

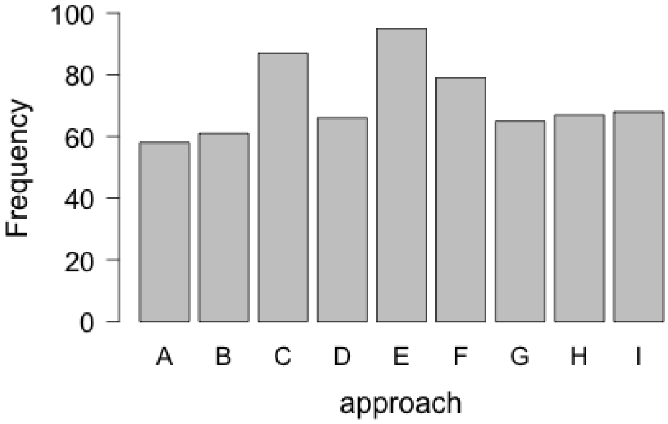

Figure 5: Frequency of lapses across the different flying scenarios. Longer bars indicate a greater amount of lapses (i.e. worse performance). See Table 1 for details concerning the approach conditions.

Finally, Figure 6 shows a time scale with weighted moving average response times for each condition. This method is used to get a smoother profile that gives an estimate of cognitive workload across the duration of the flight. These average cognitive load profiles are useful in determining particular points of high or low cognitive workload. Of particular interest are the peaks at end of each flight (the landing), where response times tend to increase sharply. This is an indication that our measure is sensitive to the pilots' increased workload during landing. Furthermore, the 50 second mark of each flight also tends to show an increase in average response time. This point is important as it was roughly this time in the flight where pilots had to complete a sharp turn and visually search for their landing zone. Other trends, previously shown in response time and lapses analyses are also displayed in the time series grid, with the 2D conditions showing a higher average response time, and Night DVE conditions showing significantly shorter flight times (we are not sure why, yet maintaining an increase in cognitive load at landing).

Figure 6: Weighted Moving Average response times plotted across flight time for each scenario. Higher average response times indicate a higher cognitive workload, mostly evident in the final section of approach and landing.

Experiment 1 Discussion

The results in this report show a preliminary investigation into the effects of symbology on pilots' cognitive load under varying environmental conditions. The primary finding is that Symbology has a minimal impact on cognitive load. The 2D symbology condition appeared to show the greatest impact on cognitive load. The 3D symbology and No symbology conditions were comparable, meaning that cognitive workload when using 3D symbology is similar to having no symbology present. This result means that the 3D symbology has a design which does not overload the pilot. Environmental conditions (DVE) did have an impact on cognitive load, with dust and night conditions showing a greater impact on cognitive workload than the day condition. This finding is important to the current study as it shows that the experimental manipulation was effective and consequently provides evidence that the DRT was measuring residual cognitive workload and not an extraneous factor, as no trade-off is evident. Both response time data and lapse data are taken into account in drawing this conclusion, with both analysis showing similar trends. It is important to consider the differences in these analyses, with average response times generally relating to cognitive workload and lapses relating to points of cognitive overload.

It was also evident that there was a significant difference between conditions where unsuccessful landings were included, with unsuccessful landings having a significantly higher average response times and frequency of lapses. Whether a cognitive overload was the cause or result of the unsuccessful landing is indeterminable from current data, and future experiments could focus on this.

Finally, when the weighted moving average was calculated over time, effects of DVE and symbology can be observed in the overall length of flight time, as well as increases to average response times at significant periods (i.e. at the turning point – roughly 50 seconds, and in the descent to land). This analysis may give a more accurate representation of the effects of DVE and Symbology, as more challenging periods of flight time, which may be short and infrequent, can be identified within certain scenarios. Furthermore, this type of analysis may be useful in developing future experiments which aim to better test the effects of Symbology and DVE on cognitive load using the DRT.

Experiment 2

Tasks

Flight: The flight simulator task was identical to experiment 1. Pilots flew two conditions of symbology; None or 3D. These were the same as these conditions in experiment 1. The environmental conditions were kept constant with visibility set at 400m. Time of day was set at 1600. FLIR and Dust were off. Traffic was off, wind speed was set at 5km/h and weather was set to have no cloud or rain.

For the flight path, the participant was asked to follow a route in the environment, going through five separate way points. The path followed a river in the virtual environment, so did not require any sharp turns or climbs. The participant was required to land at the final way point. Each flight took approximately 5 minutes.

DRT: The DRT was the same as in Experiment 1.

Participants

One participant completed four flights of the low level flight path, with two flights for each condition of symbology.

Results

Figure 7 shows little-to-no difference in response times between the no-symbology and the 3D symbology conditions. A Bayesian ANOVA supported this finding, showing a preference for the model which assumed no variance across symbology conditions.

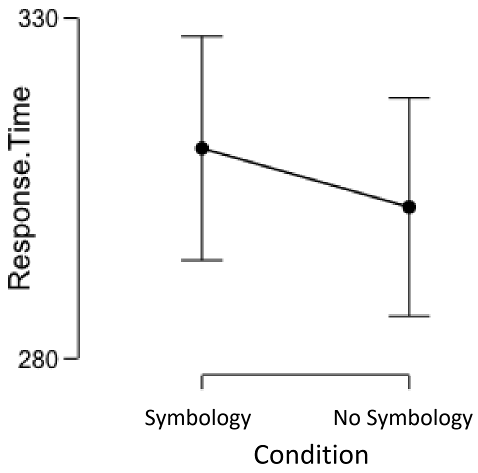

Figure 7: Mean DRT response times for each condition of symbology. Higher response times indicate higher cognitive workload. **Note:** The y-axis is on a small scale. Symbology (3D) appears to add no cognitive workload to pilots.

Experiment 2 Discussion

Experiment 2 was used as a pilot study for possible future designs, and showed evidence of effectiveness. There was no difference in DRT response times despite the addition of symbology to the pilots headpiece. These results provide evidence for 3D symbology adding no additional cognitive workload to the pilots current cognitive workload. To test whether the symbology increased flight performance, further analysis should be conducted on the flight path data to observe performance effects. The current study was limited in the amount of data collected, however, a reliable effect was shown with this small subset of data. More data, over longer distances, in conditions which show greater variance and have a more significant impact on cognitive load may need to be tested to ensure that the symbology has no significant effect on cognitive workload.

Conclusion

The current study used a DRT as the secondary task to measure the residual cognitive capacity of pilots during flight simulations. Pilots operated the simulator under three levels of symbology (no, 2D, and 3D) and three environmental conditions (day, night, dust). Results indicated no difference in average response times between levels of symbology, but 2D symbology showed a trend of more frequent lapses than the other two symbology conditions. Environmental conditions were shown to have a significant impact on cognitive load, with the Dust DVE condition showing the highest average response time, and the Day condition showing the lowest average response time and lowest average frequency of lapses. Pilots also showed a distinct trend of having significantly higher response times and lapses in scenarios where they crashed the vehicle or misjudged the landing zone.

In conclusion, the results of the current study show that Symbology has no great (negative) effect on helicopter pilot workload. That means that inclusion of symbology, at least in the scenario tested, does not overload the pilot. Symbology provides more information that can help in the accident reduction or mission completion, without taking a significant toll on pilot' cognitive load. Contrastingly, it was shown that environmental conditions played a large role in differences in cognitive workload. Future studies could focus on changes in workload over time and at significant waypoints as well as completing longer, less rigorous flights to evaluate the effects of Symbology and environments more thoroughly.

Key References

Allerton, D. J. (2010). The impact of flight simulation in aerospace. *The Aeronautical Journal*, *114*(1162), 747-756.

Broadbent, D. E. (2013). *Perception and communication*. Elsevier.

Handy, T. C., Soltani, M., & Mangun, G. R. (2001). Perceptual load and visuocortical processing: Event-related potentials reveal sensory-level selection. *Psychological Science*, *12*(3), 213-218.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.

Marshall, S. P. (2002). The index of cognitive activity: Measuring cognitive workload. In *Human factors and power plants, 2002. proceedings of the 2002 IEEE 7th conference on* (pp. 7-7). IEEE.

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.

Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., & Biondi, F. (2015). Assessing cognitive distraction in the automobile. *Human factors*, *57*(8), 1300-1324.

Strayer, D. L., Watson, J. M., & Drews, F. A. (2011). Cognitive distraction while multitasking in the automobile. In *Psychology of Learning and Motivation* (Vol. 54, pp. 29-58). Academic Press.

Authors Note:

The study was supported by the University of Newcastle Research Grant to S. Brown, A. Eidels, K. Nesbitt, and A. Brichta. Due to the constraints of industry-based testing, the experiment reported here does not have the same level of design control as is typical in laboratory experiments. For example, the population of interest (military-trained rotorcraft test pilots) is very difficult to sample from. These considerations limit the strength of inferences that can be drawn about causal relationships. Nevertheless, the results are valuable particularly because of their industry basis, for example in the way that they provide increased objectivity and more rigorous measurement of pilot workload compared with existing industry-based research practice.