Subjective Measures on Task Complexity Using Touchscreens in Flight Operations

Ben Wright[1], James Blundell[1], Wojciech Tomasz Korek[2], & Wen-Chin Li[1]

[1]Safety and Accident Investigation Centre, SATM, Cranfield University, [2]Dynamics, Simulation and Control Group, SATM, Cranfield University

SUMMARY

The following preliminary study uses subjective measures of situational awareness, workload, and system usability to assess the effect of touchscreen flight deck displays in simple and complex flying environments during a simulated flight task. Eighteen participants were evaluated whilst flying a simulated aircraft, conducting both simple and complex flight operations. Results showed that situational awareness improved, and perceived workload was maintained, when task complexity was increased during touchscreen interaction on the flight deck. This was likely driven by touchscreens providing increased attentional supply. This improves the flight deck human-machine interface (HMI) from a pilot-centred perspective by improving access to task-relevant information. There was no significant change in levels of touchscreen usability as flight task complexity increased, once again ratifying the use of touchscreens in assisting cognitive function in some task types. The application and limitations of these findings is discussed.

KEYWORDS

Aviation Safety, Flight Deck Design, Human-Machine Interface, Subjective Measures, Touchscreens

Introduction

The human factors ergonomic benefits associated with touchscreens in technology design have seen them become a prevalent method of technology operation (Orphanides & Nam, 2017). The interrelation between input and output afforded by touchscreens is an attractive way of improving system usability (Albinsson & Zhai, 2003). The user's experience of touchscreen operation has been identified as an important Human-Computer Interaction (HCI) design aspect when considering touchscreen implementation with respect to functional integration (Treeratanaporn et al., 2021), with interface design being an important factor in improving human performance and perceptions of touchscreen use (Tao et al., 2018).

Touchscreen use within the flight deck has been explored with mixed results, with evidence to suggest that the Human-Machine Interface (HMI) design of touchscreens and the type of tasks being performed affects the workload and performance of the pilot (Dodd et al., 2019). Mental workload has been found to increase in flight deck touchscreen displays with smaller touch targets (Dodd et al., 2014). However, Findings suggest that perceived workload assessed using the NASA-TLX subjective measure can be improved by optimising the layout of the touchscreen interface, although workload levels between the pilot and co-pilot vary due to differences in the positioning of the display (Wang et al., 2024). Furthermore, pilot distraction levels appear to be reduced when using touchscreen displays compared to other inputs for navigation and communication operations (Yadav et al., 2022). Despite this, it is important to consider which flight environments are suitable for touchscreen implementation, as findings suggest pilot performance and perceived usability are mixed when using touchscreen navigation displays in normal, turbulent, and startle and surprise flight scenarios (Xie et al., 2022). Moreover, the use of touchscreens for flight control operation have been found to reduce the handling quality of instrument approaches in comparison to other input modes (Li et al., 2022a). It is thus suggested that the touchscreen design framework is unsuited for some flight tasks, emphasising the importance of integrating a human-centred approach (Li et al., 2022b).

Subjective self-report measures of human performance metrics offer insight into HCI from a conscious viewpoint of the user whilst gaining insight into the user's cognitive function relative to the tasks administered. Interestingly, in Xie et al's (2022) research investigating touchscreens against analogue input types in varying flight task difficulty, objective measures, such as input accuracy analysis, found that pilot performance when using touchscreen inputs decreased in increased task difficulty compared to other input modes, whereas pilot self-report subjective measures of touchscreen use in the same tasks was contradictorily positive in some flight scenarios. This is perhaps due to the familiarity afforded by touchscreen inputs due to their prevalence in everyday life (Orphanides & Nam, 2017), which leads to bias in the user's perceptions of the accessibility and usability of the system. It is possible that the familiarity of touchscreen experience could be harnessed to ensure HCI is optimised when managing some flight task types. However, there are some contradictions between subjective assessments of touchscreens in some flight tasks. Li et al. (2023) found situation awareness and system usability was inferior with touchscreens compared to other input modes. Conversely, Rouwhorst et al. (2017) found touchscreen displays provided superior levels of situational awareness and workload for some flight tasks, whilst having little effect on other flight tasks - such as autopilot interaction. Variability in previous subjective flight deck touchscreen findings are likely to be partly dependent upon the ergonomic impact of touchscreen display location (Dodd et al., 2019). Despite this, it appears that user's perceptions of touchscreens in the flight deck appear to be positive despite the challenges associated with implementing such technology in an ergonomic way (Wang et al., 2018).

Considering the importance of keeping the end-user in the loop to design effective flight deck technology interfaces (Blundell & Harris, 2023), the present study focuses on the standardised measurement of perceived workload, system usability and situational awareness. In particular, these subjective constructs are evaluated within the context of a flight deck touchscreen interaction study that encompasses a wider range of flight deck interactions than has been previously reported: this includes manual flight path control, automation and aircraft system interaction. Furthermore, these interactions are examined and compared across two different task complexity conditions. The alternate hypotheses are as follows:

H1: There will be a significant difference in the mean NASA-TLX scores between the simple and complex flight tasks.

H2: There will be a significant difference in the mean SART scores between the simple and complex flight tasks.

H3: There will be a significant difference in the mean SUS scores between the simple and complex flight tasks.

Methodology

Participants

An opportunity sample of 18 participants was used in the study. Nine participants were female, eight were male, and one participant chose not to declare their gender. Participants had experience in the aviation industry (*M* = 7.71 years, *SD* = 8.48 years), with backgrounds in piloting aircraft, avionics engineering, air traffic control, and human factors. Participants were aged between 22 and 46 (M = 30.94, SD = 8.27). All participants consented to take part in the research.

Flight Simulator and Task Procedure

The primary apparatus used was the Future Systems Simulator (FSS) designed by Cranfield University in partnership with Rolls Royce (Figure 1). Touchscreen displays were: 1) Primary Flight Display (PFD); 2) Navigational Display, and; 3) Multi-function Display (operating autopilot altitude, heading and speed control, and fuel management). The simulator used Air Traffic Control (ATC) 'callouts' features. The FSS simulated the Gulfstream G500 aircraft during the experiment.

Figure 1. FSS flight deck consisting of PFD, Navigational Display and Multi-function Display

Following a briefing of the FSS's functionality, participants completed a simple and complex version of a task that required them to maintain a target altitude and heading displayed on the PFD. The simple task required responses to ATC callouts by recalling relevant information from the touchscreen display (e.g., current fuel quantity, airspeed etc). In the complex task, in addition to ATC information requests, participants responded to heading, speed and altitude change requests using the touchscreen displays and were required to fly in a short period of poor weather conditions. In both conditions, the sequence of ATC callouts was randomised. After each task participants completed workload, usability and situation awareness questionnaires. Participants operated the simulator on their own, with verbal help from a researcher if needed. Throttle control was automated, and rudder use was not required. The entire experiment lasted around 10 minutes.

Post-task Subjective Measures and Analysis

The NASA Task Load Index (NASA-TLX) was used as a measure of perceived mental workload using 6 subscales (performance, effort, frustration, mental demand, temporal demand and physical demand) rated by the participants from 1 to 100. The System Usability Scale (SUS) was used to measure the usability. The SUS consists of 10 statements with a 5-point Likert scale response. Situation awareness was measured with the Situation Awareness Rating Technique (SART-10D), which consists of 10 items that capture attentional demand and supply, and situational understanding. SART-10D composite scores were calculated using (Understanding – [Demand – Supply]). Due to missing data, 1 participant was excluded from the analysis of SART-10D scores.

NASA-TLX, SUS, and SART-10D data was analysed in IBM SPSS (version 29) using a repeated-measures experimental design. Task type (simple vs complex) was used as an independent variable, along with dimension (6 levels) for the analysis of the NASA-TLX data. Non-parametric tests were required for the NASA-TLX dimensional analysis due to data from the majority of dimensions being non-normally distributed (Shapiro-Wilk, $p < 0.05$). In contrast, SUS and SART-10D data met the assumptions for a paired-samples t-tests. An alpha of 0.05 was used. Bonferroni correction was applied in cases where an analysis involved multiple comparisons.

Results

Workload scores across NASA-TLX dimensions and task complexity are presented in Figure 2. scores by task type. A series of Wilcoxon signed-rank tests that comparing each task type on each NASA-TLX dimensions revealed no significant comparisons ($p > .05$). Thus, the null hypothesis for H1 could not be rejected. However, a Friedman test comparing workload across dimension scores, with task type data combined, showed that there was a significant difference ($\chi^2(5) = 49.68$, $p < .001$). Wilcoxon post-hoc test identified a significant difference between the physical demand dimension and all other dimensions ($p < .001$), and that there was a significant difference between the mental and frustration dimensions ($p = .002$). No difference was found between the other dimension comparisons.

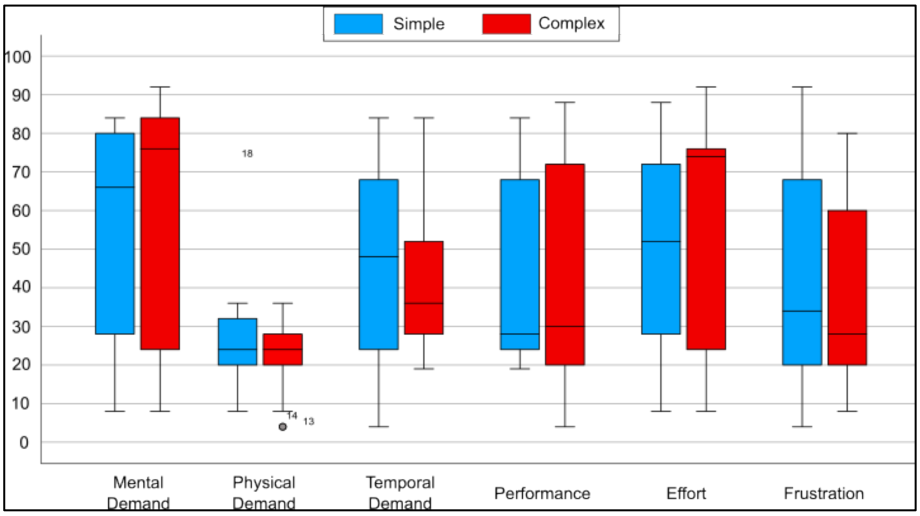

Figure 2. NASA-TLX Scores by Dimension

Figure 3 presents the mean SART-10D results between task types. Paired-samples *t*-tests showed a significant difference between the mean composite SART scores between the simple and complex tasks ($t(16) = 3.37$, $p = .004$, $d = .82$). This was driven by a significant difference in the mean supply dimension scores between the simple and complex tasks ($t(16) = 6.97$, $p < .001$, $d = 1.69$). There was no significant difference between the mean scores of the demand and understanding dimension ($p > .05$).

There was no significant difference between the mean SUS scores between the simple ($M = 63.61$, $SD = 15.82$) and complex tasks ($M = 60.83$, $SD = 17.76$) - $t(17) = 1.01$, $p = .326$.

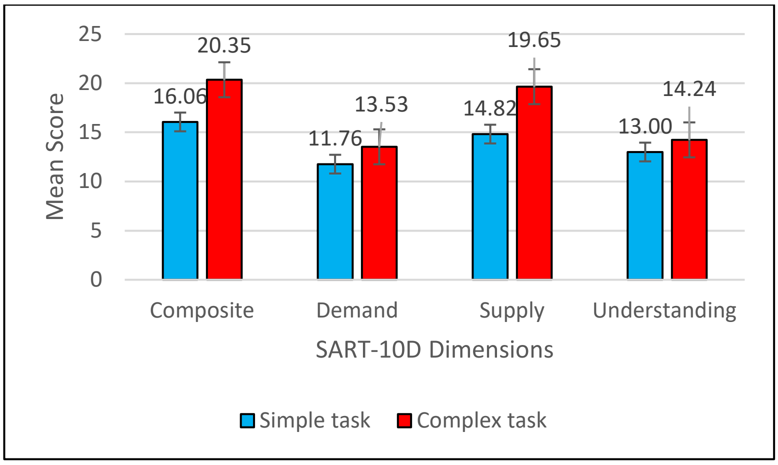

Figure 3. Mean SART-10D dimension scores by task

Discussion

The present study investigated how subjective measures of mental workload, situational awareness and usability of multiple flight deck touchscreen interfaces are impacted by the complexity of flying tasks. The key findings were that task complexity had no effect upon participant subjective workload when using touchscreen displays while aspects of situation awareness – measured via the SART-10D, were improved by task complexity. It is evident that touchscreens may assist in optimising flight task management during flight deck operations in increased task complexity.

The non-significance of NASA-TLX scores observed between complex and simple tasks suggest perceived workload levels were maintained as task complexity increased when using touchscreen displays. This finding contradicts previous research where subjective and objective workload increase as a function of flight task complexity (Blundell et al., 2020a, 2020b; Mansikka et al., 2019). One possible explanation for this finding is the ergonomic design of the current simulator. Considering Dodd et al.'s (2014) and Wang et al.'s (2024) research that stresses the importance of ergonomics within touchscreen interface design, the low scores in the physical demand dimension (irrespective of task complexity) suggests the ergonomic design touchscreens are effective in optimising ease-of-use and mitigating workload in complex task environments. This is perhaps due to their precision and interrelated input and output functionality, as discussed by Albinsson & Zhai (2003). The importance of ergonomic touchscreen design is also reinforced by the significantly lower combined frustration dimension scores in comparison to mental demand dimension scores, as participants may have found the touchscreen interface intuitive to use. It is worth noting that another explanation for these findings is due to practice effects, as all participants completed the simple and complex tasks in the same order, meaning they were more familiar with the nature of the flight tasks during the complex task condition, which may have mitigated the effect on any increase in workload.

Participant's situational awareness whilst interacting with the touchscreen improved during the complex flight task. The significant increase in the attentional supply during complex tasks corroborates Li et al. (2023), whereby touchscreens may assist in providing an improved attentional supply when needed. One reason for this could be due to the reduction in spatial displacement between the input and output of touchscreen operations. Another possible explanation is that the current study's mode of operation is similar to that of other touchscreen devices used in everyday life. Since the flight experience of the current sample was highly heterogeneous, familiarity of the touchscreen functionality may have offered a benefit to attentional supply with participants having access relevant touchscreen experience. This in turn improves situational awareness by giving the participant a greater comprehension of the required interaction mode (Endsley & Jones, 2003), which may also explain the maintenance of workload across task complexity levels.

The lack of significance between mean SUS scores for complex and simple flight tasks suggests perceptions of touchscreen usability were maintained as task complexity increased, ratifying the importance of user experience as a mitigating factor when managing flight deck tasks with the assistance of touchscreen inputs. Touchscreen use may improve usability by being familiar due to user's experiences with this interface in other contexts. However, research into simulated flight navigation tasks has found touchscreen navigation displays are less usable than standard displays due to a reduction in accuracy and throughput (van Zon et al., 2019). This may be due to the need for a greater level of tactile feedback required which enable the user to implicitly verify the operation has been conducted. As such, touchscreen implementation in the flight deck may not be appropriate for high-input tasks that require feedback, but rather as a tool for tailoring the provision of aircraft information required for a given flight task. System usability is at the heart of ensuring a human-centred approach to future flight deck design, and careful consideration should thus be taken when deciding how touchscreen implementation can optimise the useability of the flight deck. It is also worth noting that the SUS result identified in this paper may also be due to the small sample size used in the experimental design.

The present results are demonstrative of how touchscreens enable easy access to important information for the operator to feel situationally aware and maintain workload levels during flight operations. This research is beneficial as it demonstrates the possibilities of a flight deck that offers familiar HMIs as a means of reducing the negative effects of task complexity on human cognition. The findings of this research stress the importance of focusing on human-centred design frameworks for flight decks going forward, in particular by harnessing the notion of interface familiarity and ease-of-use to improve HCIs. Most notably, it is important to consider which flight operations should be designated touchscreen inputs in order to optimise pilot flight task management.

Limitations and Future Research

The primary limitation of this preliminary study is the lack of a comparison to an alternate, or traditional, mode of interaction (e.g. button / side stick). In addition, the small sample size used may affect the generalisability of the results obtained, and the use of participants with lower levels of experience in the aviation industry may not reflect the effect of pilot experience when using a new display input mode. Furthermore, the limitations placed on the number of aircraft operations being controlled by the participant may mean true workload management in flight operations is considered. The short duration of the study is also an issue, as it negates the effects of fatigue and continuous task demand on HCI in the flight deck, as is the lack of counterbalancing in the experimental design.

Future research should evaluate the efficacy of touchscreen flight deck displays with respect to a larger population. The use of a more representative sample of pilots in a pilot and co-pilot format may also improve the validity of results. Future research could also explore the usability of specific flight deck functionalities with respect to touchscreen displays in order to optimise HMIs as it is evident that not all flight deck operations can be optimally conducted via touchscreens. There are undoubtably challenges to overcome with respect to touchscreen technologies in the flight deck, namely the usability of touchscreens in turbulent conditions, and an analysis of how we can effectively present feedback for touchscreen operations in the flight deck.

Conclusion

Results of the preliminary experiment that investigated touchscreen displays in simulated flight tasks using subjective measures of human performance suggest that attentional supply, and thus situational awareness, are improved in complex flight tasks compared to simple flight tasks when using touchscreen displays, whilst perceived workload levels are able to be maintained. Touchscreens used to perform flight deck operations may improve access to important information via familiar means and ease-of-use when task complexity increases. This may be a contributing factor to maintaining perceived usability and perceived workload levels when using task complexity increases whilst using these interfaces. There are still many challenges to overcome when considering the implementation of touchscreen displays into future flight decks, namely how to optimise touchscreen design with respect to HCI to account for different flight tasks, operations and feedback methods. The following exploratory research has important practical application when looking at how to optimise pilot performance in order to improve safety outcomes within aviation.

References

Albinsson, P.-A., & Zhai, S. (2003). High precision touch screen interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/642611.642631

Avsar, H., Fischer, J. E., & Rodden, T. (2016). Designing touch screen user interfaces for future Flight Deck Operations. *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*. https://doi.org/10.1109/dasc.2016.7777976

Blundell, J., Scott, S., Harris, D., Huddlestone, J., & Richards, D. (2020a). With flying colours: Pilot performance with colour-coded head-up flight symbology. Displays, 61, 101932. https://doi.org/10.1016/j.displa.2019.101932

Blundell, J., Scott, S., Harris, D., Huddlestone, J., & Richards, D. (2020b). Workload benefits of colour coded head-up flight symbology during high workload flight. Displays, 65, 101973. https://doi.org/10.1016/j.displa.2020.101973

Blundell, J., & Harris, D. (2023). Designing augmented reality for future commercial aviation: A user-requirement analysis with commercial aviation pilots. *Virtual Reality*, *27*(3), 2167–2181. https://doi.org/10.1007/s10055-023-00798-9

Dodd, S., Lancaster, J., Miranda, A., Grothe, S., DeMers, B., & Rogers, B. (2014). Touch screens on the flight deck. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 6–10. https://doi.org/10.1177/1541931214581002

Dodd, S., Lancaster, J., DeMers, B., & Boswell, S. (2019). Multi-touch touch screens on the flight deck: The impact of display location, display inclination angle and gesture type on pilot performance. 2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC). https://doi.org/10.1109/dasc43569.2019.9081723

Endsley, M. R., & Jones, D. G. (2003). *Designing for situation awareness: An approach to user-centered design* (2nd ed.). CRC Press.

Kemble, N. K. (2020). A Mixed Methodology Study of the Effects of Age, Touchscreens, New Technology, Automation, and Interactions on Pilot Performance (thesis).

Li, W. C., Wang, Y., & Korek, W. T. (2022a). To be or not to be? assessment on using touchscreen as inceptor in flight operation. *Transportation Research Procedia*, *66*, 117–124. https://doi.org/10.1016/j.trpro.2022.12.013

Li, W. C., Liang, Y. H., Korek, W. T., & Lin, J. J. H. (2022b). Assessments on Human-Computer Interaction Using Touchscreen as Control Inputs in Flight Operations. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13307 LNAI, 326–338. https://doi.org/10.1007/978-3-031-06086-1_25

Li, W. C., Wang, Y., Korek, W., & Braithwaite, G. (2023). Future Flight Deck Design: Implementation of a Touchscreen as Flight Inceptor for Single Pilot Operation. https://doi.org/10.2139/ssrn.4357008

Mansikka, H., Virtanen, K., & Harris, D. (2019). Comparison of NASA-TLX scale, modified Cooper–Harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. *Ergonomics*, *62*(2), 246–254. https://doi.org/10.1080/00140139.2018.1471159

Orphanides, A. K., & Nam, C. S. (2017). Touchscreen interfaces in context: A systematic review of research into touchscreens across settings, populations, and implementations. In Applied Ergonomics (Vol. 61, pp. 116–143). Elsevier Ltd. https://doi.org/10.1016/j.apergo.2017.01.013

Rouwhorst, W., Verhoeven, R., Suijkerbuijk, M., Bos, T., Maij, A., Vermaat, M., & Arents, R. (2017). Use of touch screen display applications for Aircraft Flight Control. *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. https://doi.org/10.1109/dasc.2017.8102060

Tao, D., Yuan, J., Liu, S., & Qu, X. (2018). Effects of button design characteristics on performance and perceptions of touchscreen use. *International Journal of Industrial Ergonomics*, *64*, 59–68. https://doi.org/10.1016/j.ergon.2017.12.001

Treeratanaporn, T., Torat, T., & Kongtaworn, T. (2021). Usability evaluation of touch screen interface with various user groups. *2021 25th International Computer Science and Engineering Conference (ICSEC)*. https://doi.org/10.1109/icsec53205.2021.9684661

van Zon, N. C., Borst, C., Pool, D. M., & van Paassen, M. M. (2019). Touchscreens for aircraft navigation tasks: Comparing accuracy and throughput of three flight deck interfaces using Fitts' Law. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *62*(6), 897–908. https://doi.org/10.1177/0018720819862146

Virtanen, K., Mansikka, H., Kontio, H., & Harris, D. (2021). Weight Watchers: NASA-TLX Weights revisited. *Theoretical Issues in Ergonomics Science*, *23*(6), 725–748. https://doi.org/10.1080/1463922x.2021.2000667

Wang, L., Wang, Y., & Chen, Y. (2018). Survey on introducing touch-screen into civil aircraft cockpit: Opinions of aircraft designers and pilots. *CSAA/IET International Conference on Aircraft Utility Systems (AUS 2018)*. https://doi.org/10.1049/cp.2018.0250

Wang, X., Guo, W., Zhong, Z., Zeng, R., Zhang, J., & Wang, L. (2024). The research of touch screen usability in civil aircraft cockpit. *PLOS ONE*, *19*(2). https://doi.org/10.1371/journal.pone.0292849

Xie, Y., Zhou, R., & Qu, J. (2022). Fitts' Law on the flight deck: Evaluating touchscreens for aircraft tasks in actual flight scenarios. *Ergonomics*, *66*(4), 506–523. https://doi.org/10.1080/00140139.2022.2097318

Yadav, D. K., Zhao, Y., Gao, C., Wang, Y., Dong, T., & Bhatia, D. (2022). Contrastive Analysis of distractions to pilot caused by various flight instrument displays. *Journal of Transportation Technologies*, *12*(02), 209–227. https://doi.org/10.4236/jtts.2022.122013