Augmented, Virtual and Mixed Reality Simulation, Vol. 118, 2023, 69–79
https://doi.org/10.54941/ahfe1004442

Cockpit Task Management and Task Prioritization in a VR Flight Environment: A Pilot Study on the Stability-Flexibility Dilemma

Sophie-Marie Stasch and Wolfgang Mack

Universität der Bundeswehr München, Munich, Germany

ABSTRACT

Managing complex aircraft control and military tasks simultaneously in flight missions places substantial cognitive demands on pilots. To handle this challenge within the constraints of limited cognitive resources, pilots often employ cockpit task management strategies, such as task prioritization. Cognitive control plays a pivotal role in this process, as it entails directing attention towards relevant tasks while simultaneously filtering out distractions without missing safety-relevant information. The present paper relates these requirements to the stability-flexibility-dilemma of cognitive control. Different performance-related advantages and disadvantages are associated with the stability-flexibility dilemma in multitasking scenarios. On the one hand, cognitive stability is related to improved goal shielding, which in turn is associated with aggravated task switches. On the other hand, cognitive flexibility is linked to facilitated task switching but is also correlated with an increased likelihood of distraction by irrelevant cues. While the stability-flexibility-dilemma has already been investigated via task prioritization in a low-fidelity flight simulator, it remains to be explored in a more real-world flight environment. The presented study simulates a reconnaissance mission with eleven participants in a virtual-reality flight environment. Environmental factors such as weather conditions (non-windy or windy) and hostility levels (low or high) are systematically varied to manipulate task prioritization behavior. The effects of this manipulation on flight performance, workload, and eye-tracking metrics are statistically analyzed with a Bayesian repeated measures ANOVA. Results provide insight into how weather and hostility influence the cognitive control mode via task prioritization in near-realistic flight missions. Implications for the design of future studies are discussed.

Keywords: Virtual-reality, Multitasking, Aviation, Reconnaissance mission, Stability-flexibility-dilemma, Workload

MULTITASKING, TASK PRIOTIRIZATION AND COGNITIVE CONTROL IN FLIGHT MISSIONS

Flight missions are a cognitively challenging field in the domain of aviation. The main reason is that, in addition to the complex control of the aircraft, military tasks have to be concurrently performed in a dynamic task environment. Even minor errors and moments of inattention can potentially lead to fatal consequences (Lyu, Xiao & Zhou, 2018). However, cognitive resources are limited in time and modality and can quickly reach their capacity limits in multitasking scenarios. Therefore, pilots must adopt task management strategies to ensure adequate performance in the cockpit. Cockpit Task management (CTM) describes the process of initiating, monitoring, prioritizing, and terminating multiple concurrent tasks (Funk, 1991). Chou, Madhavan and Funk (1996) investigated the prevalence of CTM errors in flight accidents. Errors that played a significant role in aviation accidents were divided into three categories: The most frequent errors were attributed to task initiation (46%), followed by task prioritization (28%) and task termination (32%). The first category, task initiation, included errors like early descent, late configurations, and failures to tune navigation and communication radios. With these examples in mind, it is possible that the pilots might have been in a cognitive control state of high stability. The fact that new tasks were not properly started could be due to the fact that task-relevant cues as sensory input were not processed in sensory memory and/or working memory. Concerning task prioritization errors, the authors included distractions from weather and traffic watches. This type of error also suggests that the pilots might have been in a state of high cognitive flexibility, which might have impeded focused attention on the critical tasks. Distractions from irrelevant tasks might not have been filtered out efficiently. Finally, the task termination category included early autopilot disengagements, altitude overshoots, and improperly continued landing under unsafe conditions. This error type could be similar to errors of task initiation associated with increased cognitive stability. Cues indicating the termination of the current task might not have been processed in sensory memory and/or working memory.

In this context, the role of cognitive control is highly relevant because it involves the direction of attention to the most important tasks while irrelevant tasks must simultaneously be suppressed. Cognitive control, also known as executive control or cognitive regulation, refers to the flexible coordination of sensory, emotional, and motor processes to achieve higher-order goals (Botvinick & Cohen, 2014; Miller & Cohen, 2001). However, cognitive control is also subject to the stability-flexibility-dilemma (Braem & Egner, 2018; Goschke & Bolte, 2014). This dilemma describes the antagonistic demands of cognitive control and relates to the respective advantages and disadvantages of cognitive stability and cognitive flexibility. On the one hand, cognitive stability allows for the efficient shielding of interfering influences in the environment. On the other hand, cognitive stability is associated with a disproportionate attentional focus, leading to a diminished ability to switch to new tasks. Similarly, cognitive flexibility is also associated with relevant advantages and disadvantages. While this state facilitates task switches, it can also lead to irrelevant task distractions.

While the incorrect prioritization of tasks may have safety-critical consequences, correct prioritization of tasks is also considered an effective strategy for dealing with limited cognitive resources (Hoover, 2008). Kern (1998) strengthened the necessity of managing attention to have the necessary attention available to complete mandatory processes. The ANC (aviate-nagivate-communicate) axiom is one way to effectively prioritize the most critical tasks in the cockpit. While implications for the design of pilot training have already been elaborated (Hoover & Russ-Eft, 2005; Bishara & Funk, 2002), studies investigating the stability-flexibility-dilemma of cognitive control in flight environments are rather lacking. Stasch and Mack (2023a) successfully conducted a low-fidelity flight simulator study manipulating the cognitive control mode using task prioritization. The authors could demonstrate the effects of either control mode on performance, mental workload, and different eye-tracking metrics. However, the investigation of the stability-flexibility-dilemma in a more realistic flight environment remains unexplored.

CURRENT STUDY

To investigate how the stability-flexibility-dilemma can be manipulated via task prioritization in a more realistic flight environment, a pilot study with eleven participants was conducted. While the study conducted by Stasch and Mack (2023a) involved explicit task prioritization through instruction and a gamification method (Stasch & Mack, 2023b), the current study aims at manipulating environmental factors to guide the subject's attention in a more subtle and realistic way.

Method

A 2*2 design involving the factors wind speed (not windy = 0 mph or windy = 80 mph) and the number of unknown, hostile objects (low = 15% or high = 47%) was employed, resulting in four different mission scenarios. The order of conditions was balanced out between participants using a Latin Square design. Eye-tracking data was recorded at 200 Hz using the Head-Mounted Display (HMD) *Varjo Aero* (*see Figure. 1a*). Prior to the experimental start, participants gave their written consent to participating in the pilot study. Participants who had neurological diseases or a history of motion sickness were excluded from the study. Hypotheses about the influence of the manipulated factors on the intended task prioritization and the dependent variables are shown in Table 1.

Table 1. Hypotheses.

Independent variable	Intended Prioritization	Dependent variable
Weather	↑ windy= ↑ difficulty and focus on flight tasks (altitude, speed, navigation)	↓in performance ↑ number and duration of fixations on flight instruments (HUD & FMS)
Hostility	↑ hostility = ↑ difficulty and focus on identification task	↓ in performance ↑ number and duration of fixations (radar & out-of-cockpit)

Experimental Procedure

Participants filled out a demographic questionnaire at the start of the experiment. After that, they received instructions on how to fly the aircraft and with which schema the unknown objects could be identified. This was followed by the calibration procedure of the eye-tracker and a training mission to practice the experimental tasks. The aim of the experiment was to control the aircraft in a virtual reality flight environment by flying along a marked route and identifying unknown objects (*see Figure 1b*). There were three different unknown objects: hostile submarines and hostile aid units, as well as neutral ships, that acted as distractors. The identification process was achieved by observing the radar and by direct visual inspection by looking out of the cockpit. The participants had to report the number of unknown objects after reaching the end of the marked route. Additionally, they had to mark the position of the hostile tracks on a virtual map. Every mission was followed by a questionnaire measuring the subjective workload using the NASA-TLX (Hart & Staveland, 1988) and task difficulty of each subtask on a point-7-likert scale.

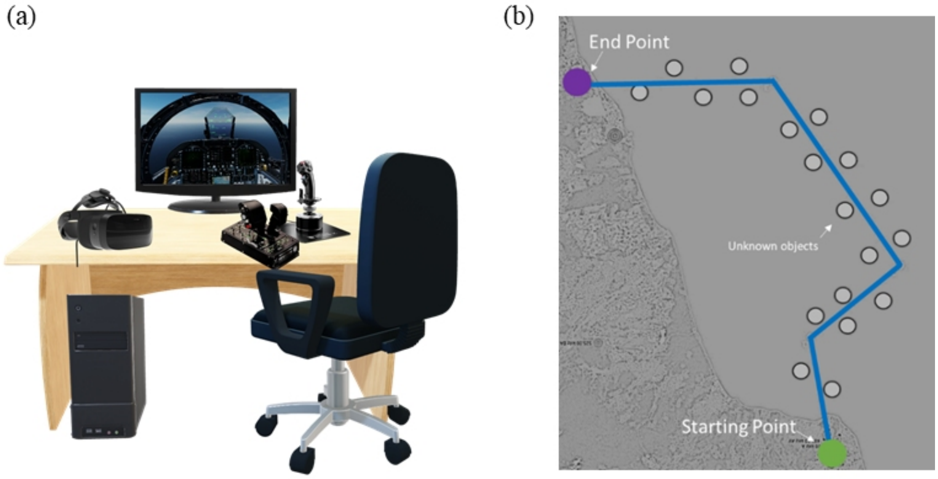

Figure 1: (a) Experimental set-up, including a HOTAS system and Head-Mounted Display (HMD), (b) Flight route with the position of unknown objects.

Participants

Eleven students ($M_{Age} = 24.90$, $SD_{Age} = 4.16$) from the University of the Bundeswehr Munich participated in the study. 36% of participants identified as female, 64% of participants identified as male. Moreover, 54% of participants indicated that they play video games at least once a month. Additionally, 54% of participants indicated to have flight experience with an average of 54.33 flight hours ($SD_{Flight\ Hours} = 105.48$). Participants rated their feeling of presence within the simulation on average with 4.22 on a 7-point Likert Scale ($SD_{Presence} = 1.86$).

DATA ANALYSIS AND RESULTS

A 2*2 Bayesian repeated-measures ANOVA (rmANOVA) was calculated for each dependent variable using JASP (JASP Team, 2023). All values were averaged across missions and all models were compared to the null model ($BF_{10} = 1.00$).

Flight Performance

Flight performance during the mission was operationalized by taking the deviations from the instructed altitude (2000 ft) and speed (250 kts) parameters as well as from the route deviation connecting the waypoints (in mi). Results indicate that neither the model including the hostility term ($BF_{10} = 0.37$), nor the model including the weather term ($BF_{10} = 0.54$), nor the model including both terms including the interaction effect ($BF_{10} = 0.13$) explain the altitude deviations better than the null model. Similarly, deviations from the flight route are neither better explained by the model including the weather term ($BF_{10} = 0.67$), nor by the hostility term ($BF_{10} = 0.45$) or the interaction term ($BF_{10} = 0.13$) compared to the null model. Regarding speed deviations, there is anecdotal evidence that the model including the weather condition ($BF_{10} = 1.63$) suits the data better than the null model.

Identification Performance

The number of correctly marked hostile objects was normalized by the total number of hostile tracks in each mission. The proportion of correctly identified objects enables a cross-comparison across the mission's despite of a different total number of hostile tracks. The best model explaining the correct positioning of unknown objects on the map (see Fig. 1b) is the model including the hostility term and weather term ($BF_{10} = 1923.76$), with $BF_{incl} = 1.97$ for hostility and $BF_{incl} = 828.39$ for weather. Adding the hostility term to the null model still makes a model improvement ($BF_{10} = 907.14$; see Fig. 2a). The proportion of correctly counted submarines is best explained

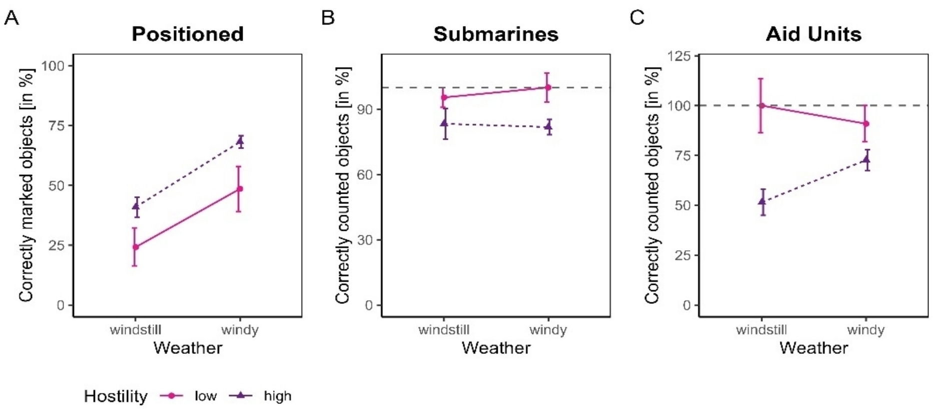

Figure 2: (a) Proportion of correctly marked objects on the map. (b) Proportion of correctly counted objects (submarines). (c) Proportion of correctly counted objects (aid units). All values were normed by the total number of available objects.

by the model including the hostility term ($BF_{10} = 4.88$) with substantial evidence in comparison to the null model (see Fig. 2b). The interaction model including weather, hostility and the interaction between both terms explains the correctly counted aid units the best ($BF_{10} = 4.27$), with $BF_{incl} = 16.02$ for the hostility term, $BF_{incl} = 1.47$ for the weather term and $BF_{incl} = 4.92$ for the interaction term between hostility and weather (see Fig. 2c). After the addition of the hostility term to the null model, results indicate anecdotal evidence against the inclusion of the weather term to the model ($BF_{10} = 0.44$). However, the addition of the interaction term between hostility and weather improves the model without the interaction term with substantial evidence ($BF_{10} = 4.43$).

Task Difficulty

The perceived difficulty of keeping the instructed speed was best explained by the interaction model including weather and hostility with anecdotal evidence ($BF_{10} = 1.81$). Adding the weather and the hostility term to the null model shows anecdotal evidence for improvement of the interaction model ($BF_{10} = 2.04$). Regarding the perceived difficulty of keeping the instructed altitude, there is anecdotal evidence that the model including the weather term explains the data best ($BF_{10} = 2.04$). Regarding the navigation, neither the weather model ($BF_{10} = 0.38$), nor the hostility model ($BF_{10} = 0.36$) or the model including the interaction term ($BF_{10} = 0.06$) explained the data better than the null model. For the task of counting the number of hostile objects, the model including the hostility term explained the data best ($BF_{10} = 6.60$).

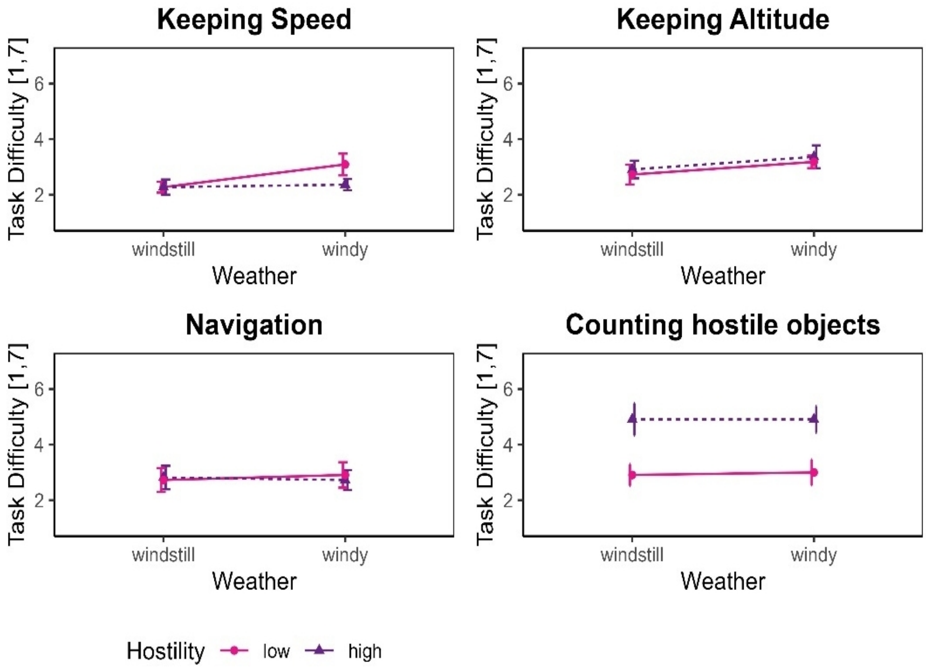

Figure 3: Perceived difficulty of the four experimental tasks on a 7-point Likert scale.

Workload

Results indicate that the model best explaining mental demand includes only the hostility term (BF_{10} =14.22) with strong evidence. Regarding physical demand, neither the model including hostility (BF_{10} = 0.57), nor the model including weather (BF_{10} = 0.41) or an additional interaction term (BF_{10} = 0.09) explains the data better than the null model. Temporal demand is best explained by the model including hostility (BF_{10} = 3.59) with substantial evidence. Concerning effort, the model including hostility explains the data best (BF_{10} = 2.66) with anecdotal evidence. Self-rated performance is best explained by the model including hostility (BF_{10} = 13.73) with strong evidence. The model including hostility explains frustration best with substantial evidence (BF_{10} = 4.23).

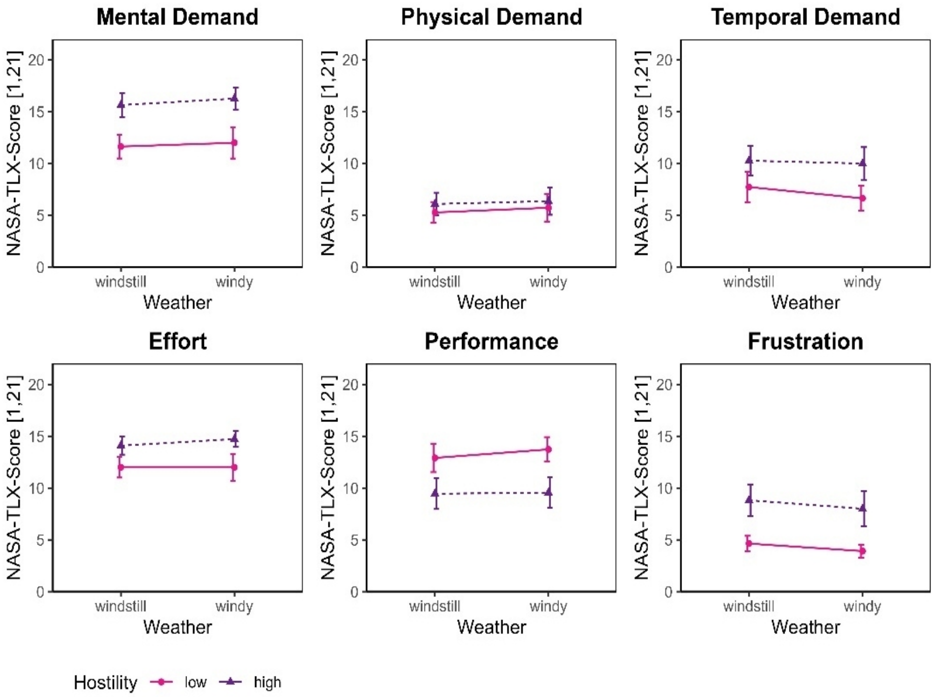

Figure 4: Mental workload measured with the NASA-TLX (Hart & Staveland, 1988).

Area of Interest (AOI) Specific Metrics

Regarding the fixation duration, the model best explaining the data only includes the AOI term (BF_{10} = 1150.78) with very strong evidence. Similarly, the number of fixations per minute is best explained by the AOI term with very strong evidence ($BF_{10} = 2.32*10^{+12}$).

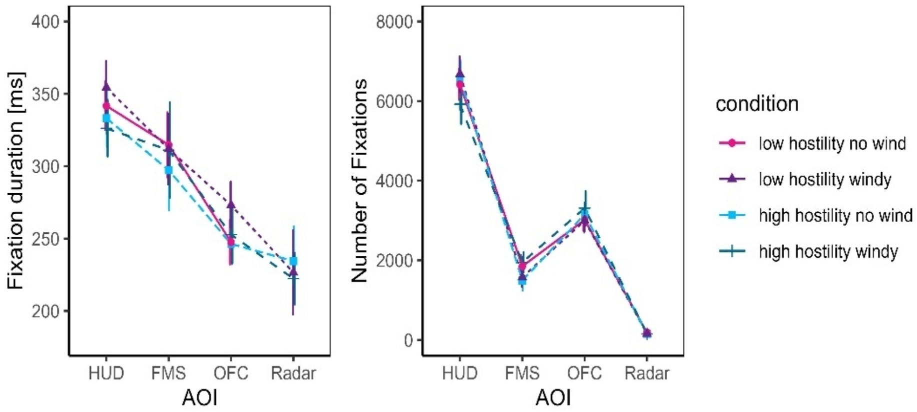

Figure 5: Fixation duration [in ms] and the number of fixations per condition and AOI. HUD = Head-Up-Display, FMS = Flight Management System, OFC = Out-Of-Cockpit.

DISCUSSION

Contrary to expectations (see Table 1), no evidence was found that the number of hostile tracks or different weather conditions influenced flight performance. Together with the fact that flight tasks (keeping speed, keeping altitude, and navigation) were perceived as relatively easy (see Figure 3) and were not influenced by the manipulated factors with substantial evidence, this finding might be attributed to a ceiling effect in performance. It could be that subjects assigned the highest priority to the flight tasks in each condition because most of their attention was on the HUD. Furthermore, it is likely that the flight tasks were designed as too simplistic, also because physical demand as one dimension of mental workload was unaffected by both factors (see Figure 4). Since the flight tasks were performed manually by adjusting throttle and stick, a higher cognitive demand should have been reflected in either performance or physical demand. Thus, it can be concluded that the influence of weather could not be proven in this experiment.

However, the influence of hostility can be discerned in the identification performance of the hostile objects. Participants proportionally counted more hostile tracks in the low hostile environment correctly compared to the high hostile environment (see Figure 2). Together with the fact that hostility influenced the perceived difficulty of the identification task, the result demonstrates that hostility influences multitasking behavior in flight missions and possibly leads to an increase in required cognitive resources. It is noticeable that participants marked the position of hostile tracks often more correctly in the high hostile environment compared to the low hostile environment. In fact, the probability of correctly marking the position by chance is higher when there are more hostile tracks in total, even if values were normalized. For this reason, this result should not be overinterpreted.

The question now arises to what extent hostility influenced task prioritization and the cognitive control mode in the experiment. Visual attention and eye-tracking metrics have a close relationship in the cockpit (Ghaderi et al., 2023; Ziv, 2016). However, no influence of hostility or weather was found in the eye-tracking results. One possible explanation for this lack of evidence lies in the analysis method of the eye-tracking metrics. When the unknown objects are in the field of view of the HUD, it is technically difficult to distinguish whether a flight parameter was monitored or if the participant looked at an object. Thus, it would be quite possible that the subjects looked at the objects and this behavior was not detected. However, this shift in behavior should have been manifested by longer and more frequent fixations on the radar, which indicate the position of the unknown objects. In general, though, the subjects looked least at the radar and most often at the HUD (see Figure 5). This result also demonstrates that the radar was not used as frequently as presumed. In summary, no influence of weather or hostility could be proven on attentional distribution. Therefore, no conclusive answer can be given as to what extent the cognitive control mode was altered by both factors.

The sample size of eleven participants in this pilot study restricts the generalizability of the results. Moreover, participants formed a mixed sample with varying flight experience, contributing to the restricted interpretation of the flight performance. Furthermore, subjects scored in the middle range of the presence score, which constrains the transfer of results to experiments in a high-fidelity flight simulator. The way the mission was set up occurs rarely in real flight missions. Nevertheless, an experimental investigation is necessary to statistically identify the effects of factors influencing multitasking behavior. The fact that the option to use an autopilot was not given only occurs in realistic flight missions when this system fails.

Future studies should therefore investigate which other factors are likely to lead to a modified prioritization of flight tasks. Further increasing the windspeed in the weather condition could provide the missing evidence of an influence of windspeed on flight performance. Positioning the unknown objects in a way that they do not fall in the visual field of the HUD would help to further explore how task prioritization influences eye-tracking metrics. Visual occlusions, such as clouds, could rather force participants to use the radar systems, making task switches technically easier to identify. The use of headphones can help to further increase the feeling of presence in the virtual environment.

A better understanding of how the distribution of attentional and cognitive resources is altered in flight missions will help to further reduce CTM errors. Possibly, adaptive assistance systems could take into account the pilot's current cognitive control mode and adjust the task distribution between pilot and automation to prevent performance declines.

CONCLUSION

The present pilot study investigated how task prioritization, as one way to manipulate the cognitive control mode, influences performance, workload, and eye-tracking metrics in a virtual flight environment. The task in the simulated flight environment was to identify hostile objects in an unknown environment (identification task) by flying off waypoints (aviate task). Task prioritization was intended to be manipulated by the weather and the number of hostile objects. Results indicate that the degree of hostility influences performance in the identification task and mental workload. An influence of weather or hostility on flight performance and eye-tracking metrics could not be proven. Future studies, including a larger sample size and a modified simulation, could help to make these results more profound. Ultimately, studies of this type can contribute to a better understanding of multitasking and cockpit task management in flight missions.

ACKNOWLEDGMENT

This research is funded by dtec.bw – Digitalization and Technology Research Center of the Bundeswehr (project MissionLab). Dtec.bw is funded by the European Union – NextGenerationEU.

REFERENCES

Bishara, S., & Funk, K. (2002). Training pilots to prioritize tasks. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 46, 96–100.

Botvinick, M. M. and Cohen, J. D. (2014). The computational and neural basis of cognitive control: charted territory and new frontiers, Cognitive science, Volume 38, No. 6, pp. 1249–1285.

Braem, S. and Egner, T. (2018). Getting a grip on cognitive flexibility, Current directions in psychological science, Volume 27 No. 6, pp. 470–476.

Chou, C.-C., Madhavan, D. and Funk, K. (1996). Studies of Cockpit Task Management Errors, The International Journal of Aviation Psychology, Volume 6, No. 4, pp. 307–320.

Funk, K. (1991). Cockpit Task Management: Preliminary Definitions, Normative Theory, Error Taxonomy, and Design Recommendations, The International Journal of Aviation Psychology, Volume 1, No. 4, pp. 271–285.

Ghaderi, M., Khalaj, A. B., Ben Abdessalem, H., Frasson, C. (2023). Attention Assessment of Aircraft Pilots Using Eye Tracking. In: Augmented Intelligence and Intelligent Tutoring Systems. ITS 2023. Lecture Notes in Computer Science, Frasson, C., Mylonas, P., Troussas, C. (eds). Volume 13891, Springer, Cham. https://doi.org/10.1007/978-3-031-32883-1_18

Goschke, T. and Bolte, A. (2014). Emotional modulation of control dilemmas: The role of positive affect, reward, and dopamine in cognitive stability and flexibility, Neuropsychologia, Volume 62, pp. 403–423.

Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, Advances in psychology, Elsevier, pp. 139–183.

Hoover, A. (2008). Prioritizing Tasks in the Cockpit: A Review of Cognitive Processing Models, Methods of Dealing with Cognitive Limitations, and Training Strategies, Collegiate Aviation Review International, Volume 26, No. 1.

Hoover, A. L. and Russ-Eft, D. F. (2005). Effect of Concurrent Task Management Training on Single Pilot Task Prioritization Performance, International Journal of Applied Aviation Studies, Volume 5, No. 2, pp. 233–251.

JASP Team (2023) JASP (Version 0.18.0) [Computer software] [Online]. Available at https://jasp-stats.org/.

Kern, T. (1998). Flight discipline. New York: McGraw-Hill.

Lyu, Y., Xiao, Y. and Zhou, Q. (2018). Classification and Cause Analysis of Human Errors in the Flight Accidents of International Modern Fighter Planes. In: Man–Machine–Environment System Engineering. MMESE 2017. Lecture Notes in Electrical Engineering, Long, S., Dhillon, B. (eds). Volume 456, Springer, Singapore. https://doi.org/10.1007/978-981-10-6232-2_31

Miller, E. K. and Cohen, J. D. (2001). An integrative theory of prefrontal cortex function, Annual review of neuroscience, Volume 24, No. 1, pp. 167–202.

Stasch, S., Mack, W. (2023a). Why the stability-flexibility-dilemma should be taken into consideration when studying pilots multitasking behaviour. In: Human Factors in Transportation. AHFE (2023) International Conference. Gesa Praetorius, Charlott Sellberg and Riccardo Patriarca (eds). AHFE Open Access, Volume 95. AHFE International, USA. https://doi.org/10.54941/ahfe1003846

Stasch, S., Mack, W. (2023b). A new experimental method to investigate multitasking strategies in flight environments via the use of gamification. In: Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2023 Annual Conference. D. de Waard, V. Hagemann, L. Onnasch, A. Toffetti, D. Coelho, A. Botzer, M. de Angelis, K. Brookhuis, and S. Fairclough (eds.). ISSN 2333–4959 (online). pp. 23–34. Available from http://hfes-europe.org

Ziv, G. (2016). Gaze Behavior and Visual Attention: A Review of Eye Tracking Studies in Aviation, The International Journal of Aviation Psychology, Volume 26, 3–4, pp. 75–104.