Creation of a Human-in-the-Loop Simulator Environment for Fifth Generation Stressor Research

Maykel van Miltenburg, Lodewijck Foorthuis, Rolf Zon
Royal Netherlands Aerospace Centre (Royal NLR)
Amsterdam, The Netherlands
Maykel.van.Miltenburg@NLR.nl, Lodewijck.Foorthuis@NLR.nl, Rolf.Zon@NLR.nl

ABSTRACT

The introduction of fifth generation platforms has led to a more information-driven air operation. It is characterized by the communication between different platform types within a mission and an increased operator autonomy due to improved information availability. Current research lacks understanding of the subtle differences between fourth and fifth generation stressors. Consequently, this requires additional knowledge on the interaction between stressors induced by fifth generation operations and human performance.

The fundamental challenge is to create a research environment in which (1) multiple fifth generation platforms operate together and (2) the equipment to measure impact of stressors on the operator is integrated. A second challenge is to generate scenarios that can be used in a research environment in which traditional (e.g. fourth generation) and potential new (e.g. fifth generation) stressors can be generated and manipulated. This paper describes the design of a research environment, which is suitable to measure the effects of these stressors.

The research environment consisted of an F-35 and an MQ-9 simulator with human operators. Mission scenarios were generated to simulate both traditional and potential new stressors. Workload and stress levels were measured using functional near-infrared spectroscopy (fNIRS), electrocardiography (ECG), and the NASA Task Load Index.

An experimental evaluation was carried out with former F-16 pilots using the developed missions and research environment. The evaluation results showed that the research environment can be used operationally and the equipment is able to measure impact of stressors on the operator. Furthermore, the environment allowed dynamic adjustment of scenario complexity during experiments, enabling adaptation to pilot performance. The research environment opens up possibilities for further generation specific stress-mitigation and training methodology development research. Ultimately, the study paves the way for future research on the impact of fifth generation platforms and operations on human performance.

ABOUT THE AUTHORS

Maykel van Miltenburg earned his Master's degree in Applied Cognitive Psychology (MSc) in 2014 and is an Human Performance expert within Royal NLR's department of Safety and Human Performance. He participated in many of the recent Royal NLR's human-in-the-loop studies in the military domain. His expertise lies on measurement tools such as rating scales and a combination of (psycho)physiological measurements such as electroencephalography, functional near-infrared spectroscopy, electrocardiography and eye-tracking.

Lodewijck Foorthuis is an R&D Engineer at the Training & Simulation department of Royal NLR. He has a degree in Mechanical Engineering and Aerospace Engineering from Delft University of Technology. He specializes in simulator federations: using multiple simulators in one integrated environment. Currently, he is working on the integration of fourth generation and fifth generation platforms in the same federation.

Rolf Zon is a well-recognized Human Effectiveness researcher. He holds a Master's degree in Cognitive Psychology from Leiden University in the Netherlands. Prior to his employment in 1997 at the Royal NLR, Rolf worked with his own consultancy company. At the Royal NLR he works as a senior consultant on Human Effectiveness in the Safety and Human Performance department. His current and past activities encompass a wide variety of human factors related research areas such as: Validation, (team) Situational Awareness, (mental) Workload, Usability, Fatigue. Further, designing and running human-in-the-loop experiments is a recurrent theme in Rolf's involvement in NLR projects.

Creation of a Human-in-the-Loop Simulator Environment for Fifth Generation Stressor Research

Maykel van Miltenburg, Lodewijck Foorthuis, Rolf Zon
Royal Netherlands Aerospace Centre (Royal NLR)
Amsterdam, The Netherlands
Maykel.van.Miltenburg@NLR.nl, Lodewijck.Foorthuis@NLR.nl, Rolf.Zon@NLR.nl

INTRODUCTION

The introduction of fifth generation platforms has resulted in an increased emphasis on information-driven air operations. These new platforms offer a multitude of opportunities for information acquisition, such as more advanced radar systems, improved camera systems, and enhanced communication capabilities with other units. These developments have led to a substantial increase in the amount of available information for pilots operating these platforms. The availability of up-to-date information and enhanced communication capabilities on fifth generation platforms has significantly increased the amount of information accessible to pilots, allowing them to act more flexibly and make continuous adjustments to their plans.

Fifth Generation Platforms Versus Fifth Generation Operations

It is important to distinguish between fifth generation platforms and fifth generation operations. The former refers to the new fifth generation hardware that provides more technological opportunities to individual platforms. The latter, however, refers to the fact that collaboration between different platforms and units offers new strategic as well as tactical approaches. The current paper focusses on development and evaluation of an initial environment for execution of fifth generation simulation experiments, whereas future work will focus on actually studying operator behavior in this fifth generation environment.

Looking ahead, it is expected that pilots will have increased flexibility and autonomy in executing missions, which may result in new demands on their cognitive capacity. This, in turn, could lead to specific high workload and stressful situations. However, current research lacks an understanding of the subtle differences between traditional (i.e. fourth generation) and possible new (i.e. fifth generation) stressors. Therefore, additional research is needed to investigate the interaction between stressors induced by fifth generation operations and human performance.

The Royal Netherlands Aerospace Centre (Royal NLR) carried out a study to prepare for new research by assessing the feasibility of a research environment that includes scenarios and (psycho)physiological measurement equipment to execute studies focused on fifth generation operations. The fundamental challenge was to design a research environment that can (1) incorporate multiple different fifth generation platforms operating together in one scenario, and (2) measure the impact of typical fifth generation events on operators' workload and stress levels simultaneously. Additionally, it was necessary to design scenarios that could effectively generate and manipulate stressors, allowing for differentiation between the effects of traditional and possible new stressors.

Advantage for the Aviation Community

This study holds significant value for aviation, specifically the aviation training community. It was found that the new developments resulting from fifth generation operations require pilots to possess distinct skills and competencies and adopt alternative approaches to missions in contrast to previous generations. In addition, in order to prepare pilots and optimize training accordingly, simulations and scenarios must be modified. This study therefore offers initial insights into what should be taken into account while preparing and redesigning training for the future military pilot.

Explanation of the Program

This study is a component of a larger program called “5th Generation Stressors”, which was established to investigate the phenomenon of fifth generation stressors through an extensive literature review, interviews, workshops, and small-scale experiments. The primary objective of this initiative was to enhance comprehension of fifth generation operations, their influence on pilots, in particular on workload and stress, and the potential repercussions for training.

The premise was that human performance will develop to a different level within the ‘fifth generation Air Force’, partly because of technological developments and a paradigm shift from platform-driven to information-driven operations. As an extension of its own perception and decision-making, the fifth generation operator will be surrounded and supported more strongly than before by new technology and a wide range of sensors. By merging the data from these sensors and other information sources in the network, the operator becomes an 'information manager', leading to a new division of tasks and cooperation between man and machine (Ministry of Defence, 2016a). Furthermore, the cooperation between people is also evolving as different platforms, unmanned systems, armed forces units, and countries collaborate, with communication and shared Situational Awareness (SA) playing crucial roles (Sweijs et al., 2018).

The new division of tasks and the more extensive possibilities of fifth generation platforms will have consequences for the way of operating, for example they can lead to new operational tactics, techniques and procedures (TTP). Within this way of acting, the human operator must also be able to perform the role as information manager under extreme and stressful conditions inherent to a military operation. It is important to consider the extent to which people can adequately process a large amount of information under complex, dynamic, mentally and physically demanding circumstances. Additionally, it is necessary to evaluate whether this new form of task load creates (cognitive) stress, which may not have been a concern in the past or to a lesser extent.

The new division of tasks between man and machine will also have consequences for education and training. The Royal Netherlands Airforce (RNLAF) anticipates that simulators will play an even more important role in the education and training of fifth generation pilots (RNLAF, 2015, 2017). Although 'zero flight time training' for the entire spectrum of military flight and tactical training may never be realized, RNLAF explicitly strives for a situation in which more training takes place in a simulator (synthetic training) than in the actual aircraft (in-flight training). While simulator use is usually associated with cost reduction, training effectiveness is the most important driver in the current discussion. For example, missions that are difficult to train in real life, can be trained in simulators. In addition, simulators are increasingly being used within Concept Development & Evaluation (CD&E) of new TTP, doctrines, future weapons, platforms and systems. The current arsenal of simulation tools has so far proved to be insufficiently capable of effectively and realistically reproducing the various physical, physiological and cognitive stressors to which the operator is actually exposed to, and this limits the applicability for its usage in R&T and CD&E purposes (Ministry of Defence, 2016b). This means that future acquisition and deployment of simulation equipment should consider the extreme (or “atypical”) and stressful operational conditions that require humans to perform their tasks. This requires more knowledge of the way in which these stressors influence the performance of human tasks, and thus on man-machine interaction.

Objective And Subjective Measurements

Besides simulating the situation that is needed, it is also important to be able to measure workload and stress related variables. Since it is complicated to 1) distinguish between traditional workload and stress, and new fifth generation induced workload and stress, and 2) execute non-intrusive reliable measures in simulated, or operational, environment, a substantial proportion of the work within this program was dedicated to workload and stress measurement techniques. The current study employs objective measures of workload and stress, including electrocardiogram (ECG) and functional near-infrared spectroscopy (fNIRS), in addition to subjective measures of pilot performance, workload and stress, and takes into account the relationship between these measures.

As task load and complexity increase, pilots must exert greater cognitive effort to maintain a good performance. This can result in increases in workload and stress, which affect the autonomic nervous system and the regulation of the body's neural and hormonal systems, ultimately impacting physiological processes such as heart or brain activity (Lean & Shan, 2012; Tao et al., 2019). Cardiac activity measured using ECG techniques is among the most widely used physiological measures for assessing workload (Charles & Nixon, 2019). Generally, heart rate (HR) increases and heart rate variability (HRV) decreases with increasing workload (Borghini et al., 2014; Hidalgo-Muñoz et al., 2018). These measures are suitable for workload research due to their non-invasive approach and relative robustness to movement artifacts.

Neural measurements of cognitive workload often focus on activity in the prefrontal cortex (PFC), the front portion of the brain. Research has shown that activity in this brain region increases with increasing workload (Geissler et al., 2021). fNIRS applied to the PFC is a sensitive method for measuring changes in cognitive workload (Sevcenko et al., 2022). fNIRS measures changes in terms of oxygenated (HbO) and deoxygenated (HbR) hemoglobin using red and near-infrared light. Increased activity in the PFC is associated with an increase in HbO and a decrease in HbR (Benerradi et al., 2019; Pinti et al., 2020). Because fNIRS is non-invasive and significantly less susceptible to electrical noise and head movements (compared to electroencephalography), it is a cost-effective, practical, and suitable method for measuring PFC activity (Fishburn et al., 2014). The combination of fNIRS with heart rate measurements has the important advantage of improving ecological validity (measurement robustness) since both are sensitive measures of workload when used separately (İşbilir et al., 2019).

Problem Statement

The introduction emphasizes the importance of gaining a deeper understanding of the workload and stress factors associated with fifth generation platforms and operations. Currently, there is a need for a research environment that can:

- Integrate various (fifth generation) platforms effectively;
- Measure operator workload and stress in the least intrusive manner possible;
- Simulate typical fifth generation scenarios;
- Provide distinct fifth generation stressors separate from traditional stressors within these scenarios.

Rationale Of The Approach

In this study, the objective was to create a research environment for small-scale experiments involving the integration of two fifth generation platforms. Due to the novelty and complexity of the simulation, a small-scale experiment was conducted with three experienced F-16 fighter pilots as subjects. While the limited sample size should be considered when interpreting the findings, this approach provided valuable insights into the feasibility of the constructed research environment and serves as a foundation for future research.

The focus was on preparing for larger-scale studies, and the study's contributions lie in successfully integrating the platforms and showcasing the potential of the measurement techniques used. Moreover, the experiment served as a demonstration of the simulation suite of Royal NLR's capability to replicate fifth generation scenarios involving multiple platforms.

Research Question and Sub-questions

The main research question is as follows:

- Is the constructed research environment in terms of technology sufficient to conduct research on the effect of fifth generation operations and platforms and its accompanying stressors?

The sub questions, underlying the main research question, are as follows:

- Is the fidelity and immersion of the constructed research environment sufficient for the pilots to experience fifth generation scenarios in an adequate manner?
- Is it possible to validly assess pilot performance in the constructed research environment?
- Is it possible to execute fNIRS and HR(V) measurements in the constructed research environment?
- Is it possible to observe differences in fNIRS and HR(V) data with varying levels of workload and stress within the composed scenarios?

METHOD

The current experiment aimed to investigate whether higher task loads, similar to those encountered in military operations, resulted in a measurable increase in workload and stress. Throughout and after the simulation the pilots' stress and workload levels are assessed using multiple assessment methods.

It further aimed to determine the feasibility of measuring subtle changes in task load induced by fifth generation scenarios can be measured within a research environment. Specifically, the experiment focused on the joint operation of an F-35 and MQ-9, with experienced F-16 fighter pilots being exposed to various stressors during the simulation.

To achieve these objectives, pilots were exposed to simulated scenarios that replicated the workload and stress experienced during military operations. Each scenario followed the same structure:

- Beginning with no direct threats;
- Progressing to multiple threats that required increased workload from the pilot;
- Ending with an event that needed to be resolved (a less threatening problem) which is potentially characteristic for fifth generation operations;
- Finally, one or more weapon deliveries needed to be executed.

Each pilot performed four scenarios, with variations in the total number of threats, the location of the weapon delivery, and the type of event to be resolved. This resulted in scenarios with low, medium and high task loads and complexities, allowing all pilots to operate at different levels of perceived workload. As the research was exploratory in nature, small adjustments to the scenarios were made during the experiments to ensure that they were not too difficult or too easy for the pilots, who had varying levels of experience. The first scenario served as a training exercise with the simulator and scenario.

After each scenario, the pilot completed a short questionnaire regarding subjective workload (based on the NASA Task Load Index; Hart, 2006) and stress. Finally, a debriefing emphasizing the feasibility, quality and relevance of the research environment took place after all scenarios have been completed. The findings from the questionnaires and debriefs are elaborated in the results and discussion of this paper.

Subjects And Facilities

In this experiment, three highly experienced former F-16 pilots participated. The pilots were recruited from the researchers' personal networks, ensuring that they are well-suited for the task at hand. The experiment involved the utilization of two interconnected facilities at Royal NLR: the Multi UAV (Unmanned Aerial Vehicle) Supervision Testbed (MUST) and the F-35 Concept Development and Experimentation Facility (CDEF).

CDEF Simulator Facility

The F-35 CDEF is a Human-in-the-Loop research simulator and has a wide range of applications for research and testing purposes. These include interoperability, investigating operational concepts and tactics and simulation based acquisition. The strength of the CDEF facility lies in its flexibility, as it can be easily customized to meet the requirements of any experiment.

MUST Simulator Facility

The facility known as MUST is a generic and reconfigurable UAV simulator designed for research purposes. In the fifth generation stressors experiment, the facility was used as an MQ-9B Predator UAV. MUST was utilized in counter-clockwise loiter mode (i.e. flying over a small area near the target in a counterclockwise direction) at an altitude of 10,000 feet, on the east side of an airfield. The facility was operated in a Human-in-the-Loop configuration, with a remote operator who was in direct communication with the CDEF and guided the F-35 pilot during weapon delivery. The operator communication between MUST and CDEF was established through a connection in the radio software ViperDIS (Battlespace Simulations Inc, 2017). Voice signals were transmitted and received via the ViperDIS software, allowing to configure radio frequencies and settings.

Cerebro

Royal NLR offers a battle lab capability named ‘Cerebro’ for military-oriented studies, demonstrations, and research purposes. This battle lab combines high-fidelity platform simulators with supplementary proof-of-concept demonstrators, which can be expanded to other battle labs (outside Royal NLR). Additionally, Cerebro has the potential to be linked with live systems when connected to, for instance, a Link-16. Cerebro has the capability to flexibly interoperate with other Royal NLR research simulation facilities, including mission support systems and Computer Generated Forces (CGF). This project used the previously described F-35 CDEF and MUST facilities, and connected these through Cerebro.

Sensors And Metrics

During the experiment, ECG and fNIRS data was measured. To record the data, sensors combined with the hybrid-8 hub of PLUX biosignals were applied to the pilots and calibrated.

For ECG measurements, the electric activity of the pilots' hearts was recorded using a three-electrode ECG sensor (3x30cm ECG sensor; PLUX biosignals). The ECG sensor has a bandwidth of 0.5-100 Hz, and the electrodes are placed in the Lead II Einthoven triangle configuration. The mean HR and HRV were extracted from the raw data using PLUX OpenSignals (r)evolution software (version 2.2.1). HRV extracted measurements included the standard deviation of the interval between normal sinus beats (SDNN) and the root mean square of successive differences between normal heartbeats (RMSSD). To obtain relative values, the results were compared to the baseline measurements conducted prior to each scenario.

For fNIRS measurements, the blood flow to the prefrontal cortex was recorded using a single fNIRS sensor (PLUX biosignals). The sensor has a resolution of 24 bits and a sampling rate of 500 Hz. It was placed on the right side of the forehead, against the dorsolateral prefrontal cortex. The raw data were obtained through the PLUX OpenSignals (r)evolution software (version 2.2.1) and converted into average HbO concentrations. To obtain relative values of HbO, the results were compared to the baseline measurements conducted prior to each scenario.

Additionally, after completing each scenario, the pilots completed a questionnaire about their perceived workload (based on the NASA Task Load Index; Hart, 2006) and stress levels. The NASA Task Load Index (NASA-TLX) consists of six bipolar subscales: mental workload, physical workload, temporal workload (time pressure), performance, effort, and frustration. The scale ranges from 1 to 100; higher scores on the NASA-TLX indicate a higher degree of subjective workload. The overall workload index is measured by averaging the results on these subscales (without weighting factors). Corwin et al. (1989) found in two studies that the NASA-TLX demonstrates high validity and reliability. A seventh item has been added to the NASA-TLX questionnaire to assess the subjective experience of stress. This seventh item is not included in the calculation of the overall workload index but will be analyzed separately.

For each scenario, every pilot was assessed on their skills in Flight Geometry, Weapon Management, Rules of Engagement and Communication. These four criteria were essential for the successful completion of the mission and are determined by an observer, a former F-16 pilot. The criteria were evaluated using a 5-point rating scale (1 = Poor; 2 = Fair; 3 = Adequate; 4 = Good; 5 = Excellent).

Explanation Of The Three Missions

The fifth generation stressors experiment scenarios comprise three types: a Low Task Load Scenario (LTS), a Medium Low Task Load Scenario (MTS), and a High Task Load Scenario (HTS). In each scenario, the pilot was assigned the task of disabling a specific target (i.e., weapon delivery). Ground and air threats were employed in each scenario, with varying levels of task load and complexity:

- The LTS included two Surface-to-Air Missile (SAM) defense systems positioned around the weapon delivery location, as well as two SAMs at a distance, and two Navy vessels dispersed across the sea. Additionally, four inactive enemy fixed-wing Fighters were located on the airfield.
- In the MTS and HTS, two additional SAMs were positioned around the weapon delivery location.
- The air threat in the LTS consisted of one enemy fixed-wing Fighter, while the MTS involved two of these enemy entities. The HTS featured two fixed-wing Fighters, and a pop-up group of an enemy entity in the

final segment of the scenario. Additionally, the HTS included a communication failure between the MQ-9 and the F-35. This communication failure represented a potentially new fifth generation event.

Moreover, the scenarios included a friendly Boeing E-3 Sentry Airborne Warning and Control System (AWACS) and a friendly General Atomics MQ-9B. An example of a MTS is shown in Figure 1 below.

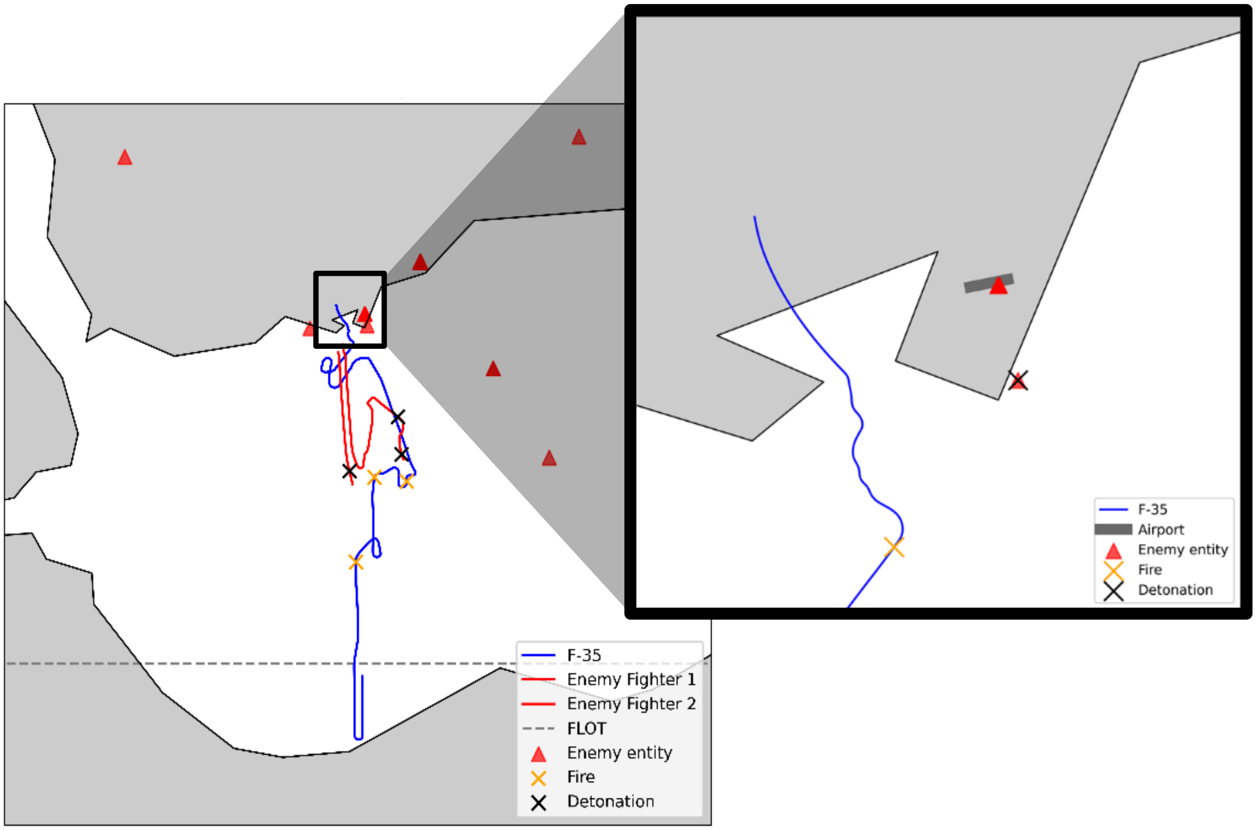


Figure 1: An example of a Medium Low Task Scenario showcasing the neutralization of two enemy fixed-wing Fighters and a successful weapon delivery. The hostile entities are represented by red triangles, with SAMs located in the north, another type of SAMs surrounding the weapon delivery location, and Navy vessels positioned in the sea.

Each scenario was divided into four segments, each segment lasting an average of five minutes, for a total duration of twenty minutes. This segmentation approach was applied to improve the reliability of the objective data analysis. The four segments were characterized by the events described below. From this point forward, any references toward the following segments are shown in *cursive*:

1. **Baseline:** The first segment comprised five minutes of flying time and did not involve any events.
2. **Mission Start:** The second segment started when the hostile entities became visible on the F-35 radar. The pilot communicated with the AWACS operator and adjusted their course toward the hostile group.
3. **Threats:** The third segment started when the hostile group came within range of the F-35. This segment involved a classic combat situation, during which the F-35 engaged in air-to-air combat with the enemy fixed-wing Fighters, if necessary using multiple maneuvers. After eliminating the immediate threat by disabling the airborne threat, the F-35 pilot adjusted their course toward the weapon delivery location as instructed by the AWACS operator. This marked the beginning of the final segment.
4. **Weapon Delivery:** The final segment was the most dynamic part of the scenario. Depending on the pilot's experience, multiple SAMs are used to adjust the stress level. This segment involved multiple evasive maneuvers by the F-35 pilot to avoid the active SAM threats. The segment ended at the moment of a successful weapon delivery.

Data Analysis

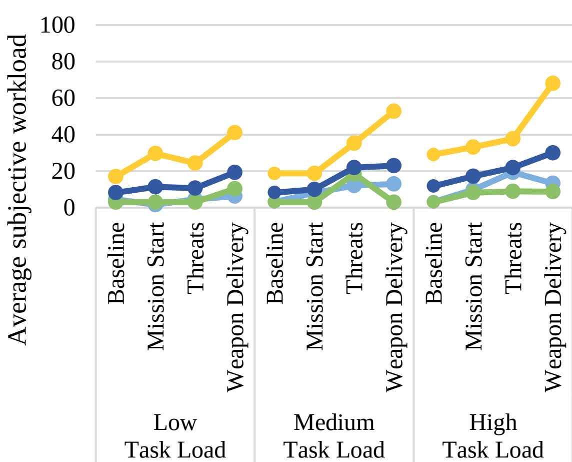
For HR(V) extraction via the HRV add-on (PLUX OpenSignals (r)evolution software; version 2.2.1), values of 0.30 seconds and 1.50 seconds were chosen for the minimum and maximum expected R-peak to R-peak interval length, respectively, along with a value of 20 for the number of R-peaks before and after, and 20% for the average acceptance of the window. For HbO extraction via the fNIRS add-on (PLUX OpenSignals (r)evolution software; version 2.2.1), the raw fNIRS data was converted to HbO.

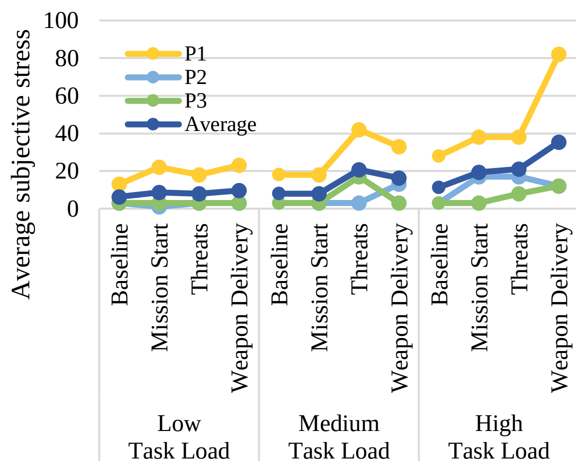
To analyze the different segments of each scenario, it was important that the data from the (psycho)physiological measurement instruments and the simulation data were synchronized in time. The simulation data contained the start and end markers of each segment in each scenario. The HbO, HR, and HRV parameters were averaged for each segment and plotted against the baseline measurement of the average in the *Baseline* of the same scenario.

RESULTS

Subjective Workload And Stress

In Figures 2a and 2b, subjective workload and stress of the three pilots (along with the average for all pilots) for each segment in each scenario are depicted on a scale ranging from 1 (very low) to 100 (very high). As can be observed, the subjective workload and stress levels increased both within and between scenarios. The subjective workload and stress during *Threats* and *Weapon Delivery* was in general higher than during *Mission Start*. Despite this, the pilots did not experience a significant degree of subjective workload and stress in each segment within each scenario. The results on stress were strongly similar to those of the subjective workload. A noticeable difference is that compared to workload, the level of perceived stress was lower at the beginning and increased more strongly during the experiment, peaking at the last *Weapon Delivery* of P1 with a score of over 80.





Figures 2a and 2b: The subjective workload and stress of three pilots (including the average for all pilots) for each segment in each scenario.

Performance

The performance of three pilots (including the average for all pilots) for each segment in each scenario is presented in Figure 3 on a scale from 1 (poor) to 5 (excellent). Overall, the pilots performed well to excellent on each segment within each scenario. One pilot (P1) scored lower than 4 (good) on one or more segments in each scenario: during *Weapon Delivery* in both the scenario with low task load and the scenario with high task load, and during *Threats* in both the scenario with medium task load and the scenario with high task load.

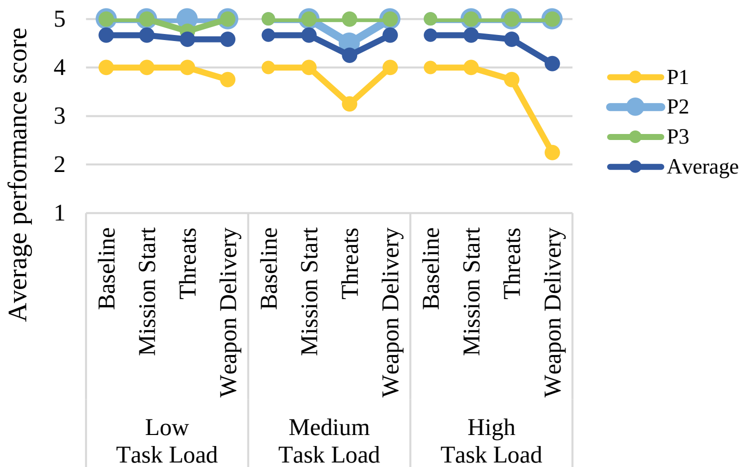


Figure 3: Performance of three pilots (including the average for all pilots) for each segment in each scenario.

ECG and fNIRS

Heart rate variability (HRV) decreased both within and between scenarios, indicating an increased level of workload and stress, except for pilot P3 in the low task load scenario. It is noteworthy that in the first scenario, Heart Rate (HR) for all pilots increased in each segment, also indicating an increased level of workload and stress. In the second and third scenarios, it appeared to stabilize more around the baseline measurement. Overall, both within and between scenarios, the HbO decreased, indicating a lower level of workload and stress, with the decline becoming increasingly more pronounced as the experiment progressed.

Debrief After The Experiment

An important aspect of the debrief was establishing whether the scenarios and simulation provided a realistic environment for the pilots, while experiencing the workload and stress levels as intended by the researchers. Regarding the workload and stress levels, the pilots unanimously reported experiencing varying levels of workload and stress, despite these levels being generally lower than in reality during operations. They also highlighted their ability to clearly differentiate between stressors within and between scenarios. Furthermore, they recognized subtle distinctions between traditional (i.e., fourth generation) and potential new (i.e., fifth generation) stressors.

Regarding the realistic environment, the pilots expressed their endorsement of the research environment, particularly its technological capabilities, as adequate for conducting studies on fifth generation operations and platforms. The pilots' positive reception of the simulation provided clear indications that the fidelity and immersion levels of the research environment were satisfactory for simulating fifth generation scenarios. The feedback from the pilots further reinforced the notion that the constructed research environment effectively captured the essential aspects of fifth generation operations.

DISCUSSION

For modern training in fifth generation operations, a flexible and adaptable environment is essential. The experiment successfully provided such an environment, which can be effectively utilized by an instructor. Dynamically adjusting the complexity of each scenario during the experiment provided the advantage of being able to adapt the task load and complexity level to the pilot's performance. In future research, the instruments used to assess performance, workload and stress could be utilized by an instructor to better determine when the task load and complexity of a scenario needs to be increased or decreased. The current research environment proved to be suitable for investigating the effect of (future) stressors and training with these stressors, as the necessary variables could be measured and adjusted appropriately.

In terms of assessing the impact of fifth generation stressors, it is important to consider the relevance of (psycho)physiological data. The utilization of fNIRS and ECG measurements in the research environment was successful and non-invasive. The recorded data demonstrated high quality, low noise, and sensitivity to variations in task load and complexity within and between scenarios. This confirms the effectiveness of fNIRS and ECG measurements in capturing (psycho)physiological responses and providing insights into the mental state of operators in fifth generation operations. However, contrary to expectations, the observed decrease in HbO and HRV may be associated with boredom or disengagement (Fishburn et al. 2014; Durantin et al. 2014) from the task, while the increase in HR suggests higher workload and engagement during weapon delivery. According to Izzetoglu et al. (2011) a decreasing HbO may be influenced by task-specific practice. Given the limited sample size, caution should be exercised when interpreting these measurements.

Limitations

While the simulated stressors were perceived as different, they did not include the broad range of characteristics of fifth generation operations or of a fifth generation platform. This was because compared to fourth generation operations similar stressors may still arise in fifth generation as well. Simulating stressors arising from the use of fifth generation platforms should therefore implement typical fifth generation specifications of the platform in the simulator. According to the pilots, this could have been done more extensively, but they could not discuss the exact desired details due to classification restrictions. The necessary details will be explored further and potentially implemented in future software updates of the simulator.

Although the fidelity and immersion of the simulation was deemed sufficient for the pilots to experience fifth generation scenarios in an adequate manner, the pilots perceived the fidelity as low compared to training simulators that they were familiar with. Furthermore, the pilots indicated that there were not enough Situational Awareness (SA) enhancing elements within the simulation. These elements refer to details that pilots can use to accurately assess the situation they are in. Ideally, for fifth generation operations, the pilot would be made more of a *battlefield manager*. To achieve this, a much larger group of entities, both friend and foe, would need to be added, either as simulators controlled by humans or as AI agents that can react to the pilot's behavior.

Another limitation of the study was the reliance on subjective evaluation methods for assessing pilot performance. Although an experienced pilot closely monitored the subjects, incorporating more objective performance assessments, such as analyzing simulator output or considering additional inputs made by the pilots, would enhance the comprehensiveness of the evaluation. By including these objective measures, a deeper understanding of pilot performance and behavior could be achieved.

CONCLUSION

The study successfully achieved its research objectives and provided valuable insights into the feasibility of the constructed research environment for studying fifth generation operations and platforms. The technology utilized in the research environment was considered adequate for investigating the impact of fifth generation operations and platforms, including their associated stressors. The fidelity and immersion provided an adequate experience for pilots in encountering fifth generation scenarios. Pilot performance assessment was feasible through subjective evaluation methods, and the utilization of fNIRS and HR(V) measurements effectively captured (psycho)physiological responses. These findings establish a foundation for future research in this field and indicate potential opportunities for enhancing assessment methods and measurement techniques in subsequent studies.

REFERENCES

Battlespace Simulations Inc. (2017). ViperDIS: Distributed Interactive Simulation (DIS) Software Suite. Retrieved from https://www.bssim.com/viper_dis_radio/

Benerradi, J., A. Maior, H., Marinescu, A., Clos, J., & L. Wilson, M. (2019). Exploring Machine Learning Approaches for Classifying Mental Workload using fNIRS Data from HCI Tasks. *Proceedings of the Halfway to the Future Symposium 2019*, 1–11. https://doi.org/10.1145/3363384.3363392

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, *44*, 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003

Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, *74*, 221–232. https://doi.org/10.1016/j.apergo.2018.08.028

Corwin, W. H., & Biferno, M. H. (1989). Assessment of Crew Workload Measurement Methods, Techniques and Procedures. *DOUGLAS AIRCRAFT CO LONG BEACH CA*, *1*, 237.

Durantin, G., Gagnon, J.-F., Tremblay, S., & Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural Brain Research*, *259*, 16–23. https://doi.org/10.1016/j.bbr.2013.10.042

Fishburn, F. A., Norr, M. E., Medvedev, A. V., & Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00076

Geissler, C. F., Schneider, J., & Frings, C. (2021). Shedding light on the prefrontal correlates of mental workload in simulated driving: A functional near-infrared spectroscopy study. *Scientific Reports*, *11*(1), 705. https://doi.org/10.1038/s41598-020-80477-w

Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(9), 904–908. https://doi.org/10.1177/154193120605000909

Hidalgo-Muñoz, A. R., Mouratille, D., Matton, N., Causse, M., Rouillard, Y., & El-Yagoubi, R. (2018). Cardiovascular correlates of emotional state, cognitive workload and time-on-task effect during a realistic flight simulation. *International Journal of Psychophysiology*, *128*, 62–69. https://doi.org/10.1016/j.ijpsycho.2018.04.002

İşbilir, E., Çakır, M. P., Acartürk, C., & Tekerek, A. Ş. (2019). Towards a Multimodal Model of Cognitive Workload Through Synchronous Optical Brain Imaging and Eye Tracking Measures. *Frontiers in Human Neuroscience*, *13*, 375. https://doi.org/10.3389/fnhum.2019.00375

Izzetoglu, K., Ayaz, H., Menda, J., Izzetoglu, M., Merzagora, A., Shewokis, P. A., Pourrezaei, K., & Onaral, B. (2011). Applications of Functional Near Infrared Imaging: Case Study on UAV Ground Controller. In D. D. Schmorrow & C. M. Fidopiastis (Eds.), *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems* (Vol. 6780, pp. 608–617). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21852-1_70

Lean, Y., & Shan, F. (2012). Brief review on physiological and biochemical evaluations of human mental workload: Brief Review on Physiological and Biochemical Evaluations of Human Mental Workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *22*(3), 177–187. https://doi.org/10.1002/hfm.20269

Ministry of Defence (2016a). Herijking van het Kennisportfolio Defensie 2016.

Ministry of Defence (2016b). Strategische Kennis en Innovatie Agenda (SKIA) 2016-2020.

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2020). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences*, *1464*(1), 5-29.

RNLAF (2015). Masterplan C-LSK 3.0.

RNLAF (2017). CLSK Research & Technology Roadmap (CRTR) 2019: Towards a 5th Generation Airforce.

Sevcenko, N., Shopp, B., Dresler, T., Ehlis, A.-C., Ninaus, M., Moeller, K., & Gerjets, P. (2022). Neural correlates of cognitive load while playing an emergency simulation game: A functional near-infrared spectroscopy (fNIRS) study. *IEEE Transactions on Games*, 1–1. https://doi.org/10.1109/TG.2022.3142954

Sweijs, T., Bekkers, F., & De Spiegeleire, S. (2018). *Playing to Your Strengths: A Different Perspective on Future Capabilities for the Royal Netherlands Army*. The Hague Centre for Strategic Studies.

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., & Zhang, T. (2019). A Systematic Review of Physiological Measures of Mental Workload. *International Journal of Environmental Research and Public Health*, *16*(15), 2716. https://doi.org/10.3390/ijerph16152716