A Study of Human-Machine Teaming For Single Pilot Operation with Augmented Reality

Narek Minaskan*
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany

Alain Pagani†
German Research Center for Artificial Intelligence (DFKI)
Kaiserslautern, Germany

Charles-Alban Dormoy‡
Université de Bordeaux,
CATIE, Talence, France

Jean-Marc Andre§
Bordeaux INP-ENSC
IMS UMR CNRS 5218
Talence, France

Didier Stricker¶
German Research Center for Artificial Intelligence(DFKI)
Kaiserslautern, Germany
TU Kaiserslautern
Germany

ABSTRACT

With the increasing number of flights in the recent years, airlines and aircraft manufacturers are facing a daunting problem: shortage of pilots. One solution to this is to reduce the number of pilots in the aircraft and move towards single pilot operations (SPO). However, with this approach, the safety and quality of the flights must be guaranteed. Due to the complex nature of piloting task, a form of human-machine teaming is required to provide extra help and insight to the pilot. To this end, it is natural to look for proper artificial intelligence (AI) solutions as the field has evolved rapidly through the past decades with rise of machine learning and deep learning. The ideal AI for this task should aim to improve the human decision-making and focus on interaction with human rather than simply automating processes without human intervention. This particular field of AI is designed to communicate with the human and is known as cognitive computing (CC). To this end, several technologies can be employed to cover different aspects of interaction. One such technology is augmented reality (AR) which as of today, has matured enough to be used in commercial products. As such, an experiment was conducted to study the interaction between the pilot and CC teammate, and understand whether assistance is required to enable safe transition towards SPO.

Index Terms: Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods

1 INTRODUCTION

One solution to the growing shortage of pilots expected with the increase in the number of flights in the coming years is to reduce the number of pilots required for flights. There are several strategies including Single Pilot in Cruise (SPIC), Reduced Crew Operation (RCO) and Single Pilot Operation (SPO). SPIC and RCO refer to a reduction in the number of pilots required for a long-haul flight during cruise. As instance, this allows to have only two pilots instead of three. One pilot remains at the controls during the cruising phase while the second can rest and then take over. There are two pilots for the preparation, taxiing, take-off, descent and landing phases therefore, only the cruise phase is concerned by SPIC. However, in SPO, the pilot is alone at the controls for the entire duration of the flight.

For both the SPIC and the SPO it is essential that the level of flight safety is maintained at its highest level in all circumstances.The major fear in SPO is if the pilot is totally incapacitated (e.g. heart attack), leaving the plane without a pilot on board. Total pilot incapacitation is rare and has been assessed by DeJohn et al. [6, 7] in American airline pilots at 0.045 and impairment rate of 0.013 per 100,000 flying hours. Evans and Radcliffe [8] demonstrated an increased risk of pilot incapacitation with age. In the event of overloading or partial or total incapacitation of the pilot, several solutions are envisaged such as assistance provided by automated systems, assistance on board the aircraft or assistance from an operator on the ground.

The multiplication of new systems to assist the pilot and increase their capabilities, such as the Synthetic Vision System (SVS) or the Enhanced Vision System (EVS) will play a key role in the shift to SPO. Cummings et al. [5] propose functional requirements for assistance provided by automated systems. One such is oral and bidirectional communication with automation. The goal being to replace the co-pilot and his role as pilot-monitoring by automation. But increasing automation in a cockpit would only reinforce the automation paradox and increase the burden of system monitoring by the pilot [3, 16]. In order to increase trust in the systems and thus reduce monitoring, the human autonomy team (HAT) is the focus of several researches [14, 19]. A study of Bailey et al. [2] showed that pilots in SPO in a legacy cockpit,were able to handle abnormal situations safely and with acceptable performance conditions. However, flight performance decreased, and safety margins and workload were assessed by the pilots as unacceptable, particularly in an emergency situation. It is imaginable that for SPO, a good HAT with specific tools could manage abnormal situations as good as two pilots crew in the future.

The HAT is one solution towards SPO. This paper goes further by proposing a Human Intelligent Machine Team (HiMT), with cognitive computing (The machine) as a teammate for the pilot (The human) instead of more automation to avoid the paradox of automation [3, 15]. The CCTeammate aims to be implemented in legacy cockpits and will be able to evolve if there is a total overhaul of cockpits for single-pilot aircraft and will be usable during the total renewal of fleets. The recommendations of Cummings et al. and Shively et al. [5, 19] for HAT comply with the CCTeammate. It keeps verbal and non-verbal communication, with a multimodal HMI², to create a HiMT. The CCT does not have the possibility to take control of the aircraft or to make tangible actions with the

---

*e-mail: nareg.minaskan_karabid@dfki.de
†e-mail: alain.pagani@dfki.de
‡e-mail: ca.dormoy@catie.fr
§e-mail: jean-marc.andre@ensc.fr
¶e-mail: didier.stricker@dfki.de

cockpit to let the pilot-in-command in control of the aircraft. The case of total pilot incapacitation is out of the scope of this article.

This paper is structured as follows: in section 2 an overview of the experiment with usecase scenario and communication modalities are provided. Next, section 3 explains the technical implementation. Then, the methodology and results are presented in section 4. Discussion of the results and conclusion are discussed in 5 and 6 respectively.

2 EXPERIMENT DESIGN

Before setting up a meaningful experiment, it is necessary to carry out an analysis of the tasks (and the activity) in order to understand the stakes. A distinction is made between task analysis (what is to be done) and activity analysis (what is done), whether at the behavioral or cognitive level. This first activity (analysis of the tasks and the activity) was carried out by a preliminary interview of the pilots and asking them to share their knowledge and experiences. In a user-centered design approach, pilots interviews were conducted to learn more about the tasks in all phases of flight, from the preparation of the flight to the shutdown checklist (or turnaround times). The purpose was to define the tasks they want to delegate or agree to delegate to the AI, share, and the ones they do not wish to delegate, share, for each flight phase and each condition (normal and abnormal).

Based on the results of the interviews, it was obvious that the experiment scenario must contain a complex situation which will require the pilots to understand a great amount of data presented to them and will have them to make quick decisions due to time constraint. It is worth mentioning that the interviews did not solely focus on defining a usecase but also studied the human aspects such as acceptability and trust.

2.1 Usecase Scenario

The initial questions that were imposed are as follows:

- How cognitive computing could bring more efficient support to the pilot in the light of existing limitations e.g. misunderstanding of the information delivered, and the risk of wrong decisions being made?
- Which flight phases deserve specific effort?
- How cognitive computing can help pilots differently than automation?
- What are the available technologies?

The usecase is an important part of the experiment which will leverage the potential of cognitive computing, to demonstrate and challenge the HiMT. The main assumption was the following: situation awareness (SA) and mental workload of the pilots can be enhanced with the help of a CC and an AR system in different situations and contexts. SA is a key element for the decision-making model of the pilots. If the SA is enhanced, their decision-making will be enhanced, through trusted information provided by the CC. A Look at accident statistics [1] showed that 49% of accidents happen during approach and landing. By taking into account the descent and the initial approach phases, this reaches 60% of the accidents between 2008-2017. It was also noticeable that runway safety (RS) represents the major risk of accidents, compared to controlled flight into terrain (CFIT) and loss of control-In flight(LOC-I) [11]. The objective was therefore to take these statistics into account and to design a usecase based on them. Therefore, the most relevant scenario will be a complex scenario including the approach and landing phases with a risk of runway excursion, LOC, CFIT, a system failure to manage and a time constraint such as lack of fuel. To this end, the Bremen landing scenario was chosen as it was representative of major accidents and closest to our interests (Fig. 1). This well documented scenario was already used in the Future Sky Safety project and is consisted of nine phases: descent, approach, go-around, system failure, weather change, time constraint, runway length constraints and specificity (need of LAPA calculations, documentation search (e.g. quick reference handbook and operational manual), and landing. SA and mental workload are the measures used for evaluation. The results of the Future Sky Safety project with a two pilot crew, are taken as a baseline.

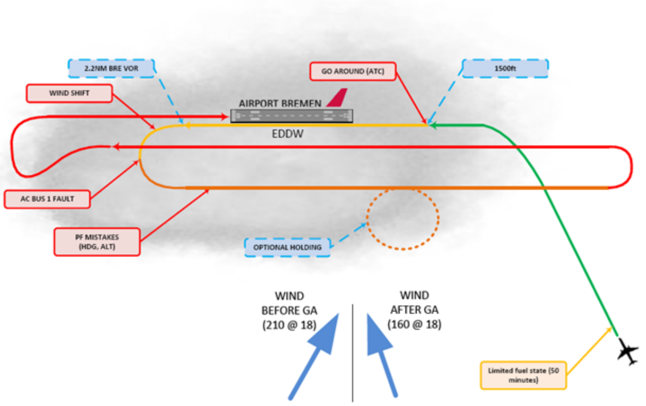


Figure 1: **Green Phase:** The flight begins at top of descent with a low amount of fuel. The final approach to runway (RWY) 27 is canceled by Air Traffic Controller (ATC) . **Yellow:** Crew performs a Go Around and experiences an AC BUS 1 FAULT in the turn to downwind and the wind shifts by 50°. **Orange:**RWY 27 is no more possible (wind shift and A/C limitations), RWY 09 is the only option. **Red:** The pilot must fly back to RWY 09, prepare a CAT2 approach with manual rollout. The colors are related to the time constraint induced by the quantity of fuel.

2.2 Communication Level and hypothesis

Two levels of communication were defined for the CC teammate: CC Assistant on request, and CC pro-active teammate. For the statistical analysis and experiment, each of the these levels are referred to as a "step" such that, step 1 has no assistance, step 2 with CC on request, and step 3 corresponds to proactive teammate.

2.2.1 CC Asistant on request

In this case, the CCAssistant will be active only if the pilot asks a question or requests help. If this is the case, the variant will be as described in "proactive CC teammate". In this modality the pilot does not wear AR glasses.

2.2.2 proactive CC teammate

- **Level of communication, quality of explanation:** Deliver information to the pilot. It is informative message by voice or by AR. It delivers information without explaining where it comes from and why, or the repercussion it may have on the rest of the flight.
- **AR messages:**Visual information appearing in a virtual side panel. Caution and warning can also be displayed in 2D in the direct field of view of the pilot.
- **Communication modality:** Limited speech commands from the pilot and informative voice messages about current state of the plane.

2.2.3 Hypothesis and measurements

Within the framework of the usecase scenario and defined communication levels, the main hypothesis are the following:

- A safe SPO is possible without assistance or with an assistant on request (An SPO without assistance or with an assistant on demand is not sufficiently safe).
- A safe SPO requires a collaborative (pro-active) intelligent teammate

NASA TLX and SART [10, 17] questionnaires were used to measure the cognitive workload and SA in order to understand which CC teammate is the most useful and appropriate for safe SPO.

3 IMPLEMENTATION

The AR application was developed for Microsoft HoloLens 2 with Unity engine. Due to complex nature of bidirectional communication between pilot and CCT, pilot and Ait Traffic Controller (ATC), it was decided to adopt a Wizard-of-Oz approach. As such that the wizard will provide the necessary verbal communication and will display the desired virtual information on HoloLens (Fig. 2). It must be noted that the wizard is not the same as ATC and in fact, the wizard does not communicate with ATC, rather is a silent listener and interpreter when it comes to communication with ATC, which only displays the information from ATC.

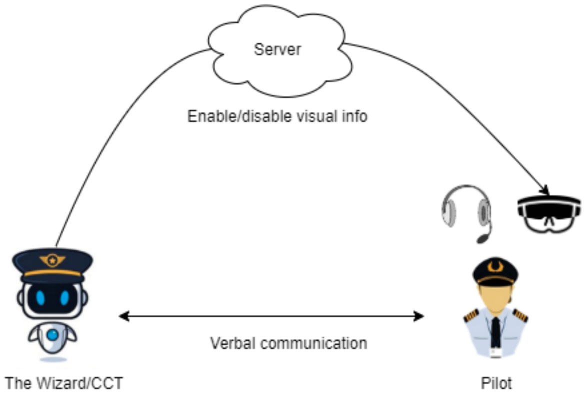


Figure 2: Communication between the pilot and the CCT. The wizard plays the role of the CCT and engages in direct verbal communication with the pilot. At the same time, the wizards issues commands via a desktop application to enable/display visual information or data to the pilot.

To this end, two separate applications were developed: A desktop application and an AR application for HoloLens. The desktop application was developed with C# windows forms together with MQTT library [4] for networking. This application contains checkboxes and value fields where the wizard can enable/disable information or display data and act as a server. The AR application developed with Unity engine contains Mosquitto MQTT broker [13] library for receiving messages from the server and Microsoft Mixed Reality Toolkit [12] for enabling interactions. Two types of visual information were developed: 2D and 3D. The 2D information was displayed in upper half of field of view whereas the 3D information were displayed on a 3D window (Fig. 3) or, in case of data, above their respective panels where the pilot will enter their values. The 3D window and values were pinned to the desired location in cockpit by enabling world anchors. This made the position tuning simpler as at each run, there is no need to redo the positioning. The grabbing were then disabled throughout the experiment so the pilots won't accidentally grab the window or values while they move their hands.

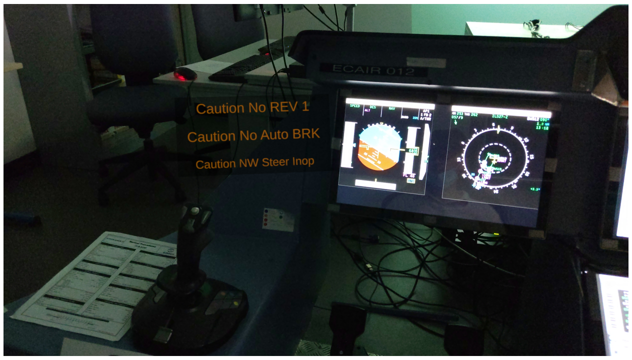


Figure 3: 3D information displayed to the pilot in a virtual side panel/window.

4 METHODOLOGY AND RESULTS

Eighteen healthy pilots ($M = 34,05$ y.o. ; $SD = 9,46$; flights experience = 2836 Flight hours, range = 2185, 20% female, 50% captain and the other 50% copilots), were recruited and were paid for their voluntary participation in this study. Confidentiality was guaranteed. All participants gave written informed consent. Crews were not familiar with each other. 10 pilots had participated in the baseline (two crew pilot operation), 5 pilots in step 1, 5 pilots in step 2, and 8 pilots in step 3 (As discussed in section 2.2).

4.1 Flight simulator

For the usecase, the cockpit demonstrator located in the Bordeaux INP premises has been used. The cockpit interface consists of hardware elements and touchscreens for the overhead and a part of the pedestal (Fig. 4). The simulator is equipped with physical sidestick, throttle levers, flaps, speed brake. A flight control unit (FCU) similar to the one available in an A320 is present to manage the autopilot flight parameters (altitude, speed, heading, etc. . .). The pilot has a Primary Flight Display (PFD) and the Navigation Display (ND). In the middle there are the upper and lower ECAM (Electronic Centralized Aircraft Monitoring) as shown in Fig. 4. The simulator has a fixed structure. The flight model and the scenery are taken from Prepared 3D simulator (developed by Lockheed Martin). The A320 functionalities and behavior are simulated with high fidelity. Such as the Flight Management and Guidance System (FMGS) with S8 logic and proper SID/STAR tracking; lateral and vertical flight management which follows the ARINC 424-19 specification in full detail; the entire range of aircraft systems; and the complete custom Fly-By-Wire implementation featured in the actual aircraft.

Figure 4: A320 flight simulator

TeamSpeak has been used for the ATC-Pilot audio communication as well as Pilot-CC teammate (WoOZ) communication. Three different cameras were used to record the whole cockpit. A camera was placed in the co-pilot's seat facing the pilot to have a complete view of the pilot and be able to observe their behavior. A second camera was placed in the middle of the cockpit to have a complete view of the cockpit and to be able to see all the pilot's interactions and screens. And finally, a third camera placed to visualize the overhead panel.

4.2 Procedure

Before each flight simulator session, the pilot was briefed about the flight simulator specificities (e.g. touch screens) and the flight scenario. For steps 2 and 3, a specific briefing was done to introduce "Jack", the CC teammate. It was explained to them how to communicate with Jack, what it could and could not do (e.g. take control of the aircraft, push buttons) and a paper summarizing all these points was systematically given to them. The training scenario consisted to take-off from Bordeaux (LFBD) to fly a Standard Instrument Departure (SID) and the Standard Terminal Arrival Route (STAR) and to land. The simulator was systematically stopped at 100ft above ground. The purpose of this training was for the pilot to take the simulator and the touch screen and become familiar with the interactions with the CCT. For the experiment scenario the pilot is briefed about the origin airport (LFBD) and the destination. The weather at destination is given on a paper. The pilots have all the time they want to know the Bremen (EDDW) destination airport on the paper charts or on the electronic flight bag (EFB), and to make the performance calculations. And the same for the alternate airport Hannover (EEDV). They are briefed on the flight plan, that they are at 5min from the Top of Descent, the distance to the airport, the flight level (FL330), the STAR entered in the FMS (PIXUR3P) and the configuration of the aircraft. The screens of the simulator were switched off during this phase of briefing and are switched on again only at the launching of the experimentation and thus of the flight scenario.

4.3 Results

Prior to analyses, data were cleaned, and the assumptions of normality were tested to ensure that they hold. Violations for assumptions of normality were identified using Shapiro–Wilk test for all variables in order to guide selection of statistical tests. Univariate analyses were conducted using Kruskal–Wallis tests for non-normally distributed variables to determine whether there were significant differences between the 3 steps compared to the Baseline. All analyses were conducted using Jamovi 1.6.23 statistical software.

4.3.1 Cognitive Load

With regards to cognitive load, mean scores show that the overall workload did not significantly change between the different steps as shown in Fig. 5, $\chi^2(3) = 3.20$, $p = 0.361$. Step 3 ($M = 73.1; SD = 11.3$) still slightly higher than the baseline ($M = 63.4; SD = 15.3$) but the difference is not significant.

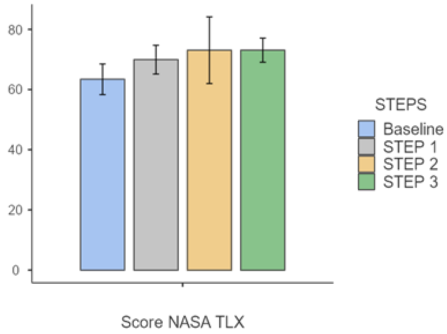


Figure 5: NASA TLX score

4.3.2 Overall Level of Situation Awareness

The SART questionnaire requires participants to rate demand on attentional resources, supply of attentional resources and understanding of the situation on a 1-7 scale. Responses to the SART result in a subscale for each of the aforementioned dimensions as well as a combined score based on the difference between attentional demand and the sum of supply and understanding ratings [20].

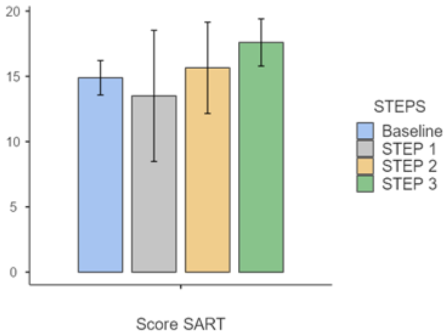
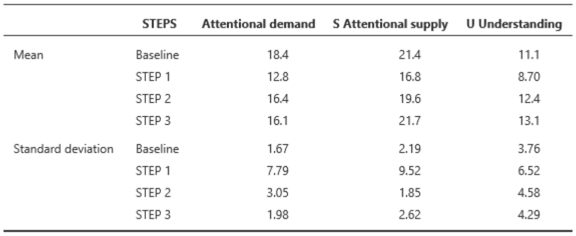


Figure 6: SART score

As seen in Fig. 6, even if the difference is not significant, $\chi^2(3) = 1.53, p = 0.675$, the highest level of SA was in step 3 followed by the baseline.

The separate analyses of the three dimensions of SART revealed no significant results(Fig.7).



	STEPS	Attentional demand	S Attentional supply	U Understanding
Mean	Baseline	18.4	21.4	11.1
	STEP 1	12.8	16.8	8.70
	STEP 2	16.4	19.6	12.4
	STEP 3	16.1	21.7	13.1
Standard deviation	Baseline	1.67	2.19	3.76
	STEP 1	7.79	9.52	6.52
	STEP 2	3.05	1.85	4.58
	STEP 3	1.98	2.62	4.29

Figure 7: Descriptive analysis of the three dimensions of SART

The attentional demand was slightly lower in step 1 compared to the baseline and step 2, and 3 but the results were not significantly different $\chi^2(2) = 2.54, p = 0.100$. The attentional supply is slightly higher in step 3 ($M = 21.7, SD = 2.62$) compared to the baseline, but the results were not significantly different $\chi^2(3) = 3.76, p = 0.288$. The results regarding the understanding of the SA are slightly higher in step 3 compared to the baseline, but the difference is not significant, $\chi^2(2) = 2.54, p = 0.468$.

4.3.3 Correlation Between NASA TLX and SART

The overall NASA-TLX score and SART overall correlated negatively with $r = -0.465; p = 0.015$, demonstrating that when the SA is high, the workload is lower with some overlap (Fig. 8). SART is a subjective measure, concerns have been expressed that it is overly related to workload [18].

4.3.4 Correlation Between SART and its Dimensions

The overall SART score and attentional demand correlated negatively with $r = -0.408, p = 0.035$, demonstrating that when the SA is high the attentional demands seem to be lower. The overall SART score and attentional supply correlated positively with $r = 0.558, p < 0.002$, demonstrating that when the SA is high the

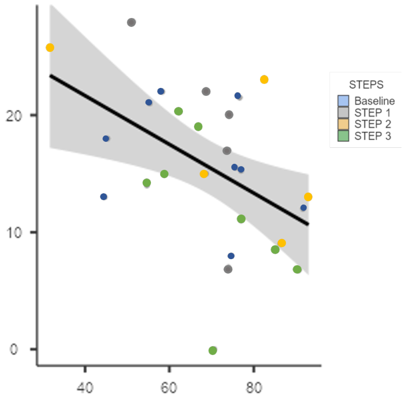


Figure 8: Correlation between NASA TLX and SART.

attentional supply increase. Finally, the overall score SART and understanding correlated positively with $r = 0.630, p < 0.001$, demonstrating that when the SA is high there is a higher score of understanding.

5 DISCUSSION

The objective of this study was to show that implementing a proactive CC teammate in SPO could improve pilot's SA and decrease or maintain an acceptable workload in abnormal situations. Even if the difference is not significant, our results seem to support this hypothesis. Indeed, the SA was slightly higher in step 3 (pilot with a proactive CC teammate) compared to the baseline (2 pilots crew). This could be explained by the fact that the proactive CC teammate gives important information responsible for the improvement of the pilots' SA. Also, in step 3 the pilot and the proactive CC teammate had a better collaboration with more exchanges, which induced a better understanding, less attentional demands and better attentional supply with a higher SA. On the other hand, the proactive CC teammate, in some cases, increased the pilot's workload. This result does not preclude the fact, that a proactive CC teammate increase workload because of the SPO context, with which pilots are not familiar. Also, the use of virtual assistance is something new for them and in some cases the proactive CC teammate didn't take into account the current task, which led to task switching or dual task and contributed in some cases to overload the pilots. In our further research, a new step will be included. In the step 4 the CC teammate will be adjusted to pilot current task. Explanations will be given to increase the SA and to improve pilot's decision making. Anticipation of critical situations will also be added to step 4.

6 CONCLUSION AND FUTURE WORK

In this paper, cognitive computing and HiMT were introduced as a possible solution for SPO in contrast to more automation. To this end, an experiment was designed utilizing AR and vocal communication together with a complex flight scenario, to test two levels of CC teamate: CC Assitance on request and Proactive CC teammate. The experiment took place in a flight simulator located in Bordeaux INP premises, where pilots were recruited to pilot an A320 aircraft alongside with one of the levels. Afterwards, they were asked to fill out the NASA TLX and SART questionnaires for evaluation of mental workload and SA. The statistical analysis showed that in step 3, the pilot and the proactive CC teammate had a better collaboration with more exchanges, which induced a better understanding, less attentional demands and better attentional supply with a higher SA. The future work will include an additional step 4, where the CC teammate will be adaptive to the situation at hand and will anticipate critical situations. Moreover, Faulhaber et al. [9] confirm that the absence of the Pilot Monitoring affects the Pilot Flying's scanning behavior. In their experience the participants spent significantly more time scanning secondary instruments at the expense of primary instruments when flying alone. Future work should also tackle major challenges regarding eyetrackers, to examine if a CC teammate could reduce the Pilot Flying scanning time behavior or not.

ACKNOWLEDGMENTS

This project has received funding from the Clean Sky 2 Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 831891. The authors would like to thank Théodore Letouzé for taking care of the technical aspects of the simulator and Turkan Hentati for help with statistical analysis.

REFERENCES

[1] B. C. Airplanes. Statistical summary of commercial jet airplane accidents. *Worldwide Operations*, 2008, 1959.

[2] R. E. Bailey, L. J. Kramer, K. D. Kennedy, C. L. Stephens, and T. J. Etherington. An assessment of reduced crew and single pilot operations in commercial transport aircraft operations. In *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, pp. 1–15. IEEE, 2017.

[3] L. Bainbridge. Ironies of automation. In *Analysis, design and evaluation of man–machine systems*, pp. 129–135. Elsevier, 1983.

[4] I. Craggs. Mqtt for .net, https://github.com/eclipse/paho.mqtt.m2mqtt.

[5] M. L. Cummings, A. Stimpson, and M. Clamann. Functional requirements for onboard intelligent automation in single pilot operations. In *AIAA Infotech@ Aerospace*, p. 1652. 2016.

[6] C. A. DeJohn, A. M. Wolbrink, and J. G. Larcher. In-flight medical incapacitation and impairment of us airline pilots: 1993 to 1998. Technical report, FEDERAL AVIATION ADMINISTRATION OKLAHOMA CITY OK CIVIL AEROMEDICAL INST, 2004.

[7] C. A. DeJohn, A. M. Wolbrink, and J. G. Larcher. In-flight medical incapacitation and impairment of airline pilots. *Aviation, space, and environmental medicine*, 77(10):1077–1079, 2006.

[8] S. Evans and S.-A. Radcliffe. The annual incapacitation rate of commercial pilots. *Aviation, space, and environmental medicine*, 83(1):42–49, 2012.

[9] A. K. Faulhaber, M. Friedrich, and T. Kapol. Absence of pilot monitoring affects scanning behavior of pilot flying: implications for the design of single-pilot cockpits. *Human factors*, p. 0018720820939691, 2020.

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

[11] ICAO. Icao accident statistics.

[12] K. Leigh. Microsoft mixed reality toolkit, https://github.com/microsoft/mixedrealitytoolkit-unity.

[13] R. Light. Mosquitto mqtt broker, https://github.com/eclipse/mosquitto.

[14] M. Matessa, K.-P. L. Vu, T. Z. Strybel, V. Battiste, T. Schnell, and M. Cover. Using distributed simulation to investigate human-autonomy teaming. In *International Conference on Human Interface and the Management of Information*, pp. 541–550. Springer, 2018.

[15] R. Parasuraman and D. H. Manzey. Complacency and bias in human use of automation: An attentional integration. *Human factors*, 52(3):381–410, 2010.

[16] R. Parasuraman, R. Molloy, and I. L. Singh. Performance consequences of automation-induced'complacency'. *The International Journal of Aviation Psychology*, 3(1):1–23, 1993.

[17] S. J. Selcon and R. Taylor. Evaluation of the situational awareness rating technique(sart) as a tool for aircrew systems design. *AGARD, Situational Awareness in Aerospace Operations 8 p(SEE N 90-28972 23-53)*, 1990.

[18] S. J. Selcon, R. M. Taylor, and E. Koritsas. Workload or situational awareness?: Tlx vs. sart for aerospace systems design evaluation. In *Proceedings of the Human Factors Society Annual Meeting*, vol. 35, pp. 62–66. SAGE Publications Sage CA: Los Angeles, CA, 1991.

[19] R. J. Shively, J. Lachter, S. L. Brandt, M. Matessa, V. Battiste, and W. W. Johnson. Why human-autonomy teaming? In *International conference on applied human factors and ergonomics*, pp. 3–11. Springer, 2017.

[20] R. M. Taylor. Situational awareness rating technique (sart): The development of a tool for aircrew systems design. In *Situational awareness*, pp. 111–128. Routledge, 2017.