Human intelligent machine teaming in single pilot operation: A case study

Narek Minaskan†[1], Charles Alban-Dromoy †[3], Alain Pagani[1], Jean-Marc Andre[4], and Didier Stricker[1,2]

[1] German Research Center for Artificial Intelligence, Augmented Vision department, Kaiserslautern, Germany;
[2] Technical University of Kaiserslautern, Kaiserslautern, Germany
[3] Universit ´e de Bordeaux, CATIE, Talence, France
[4] Bordeaux INP-ENSC IMS UMR CNRS 5218 Talence, France

Abstract. With recent advances in artificial intelligence (AI) and learning based systems, industries have started to integrate AI components into their products and workflows. In areas where frequent testing and development is possible these system have proved to be quite useful such as in automotive industry where vehicle are now equipped with advanced driver-assistant systems (ADAS) capable of self-driving, route planning, and maintaining safe distances from lanes and other vehicles. However, as the safety-critical aspect of task increases, more difficult and expensive it is to develop and test AI-based solutions. Such is the case in aviation and therefore, development must happen over longer periods of time and in a step-by-step manner. This paper focuses on creating an interface between the human pilot and a potential assistant system that helps the pilot navigate through a complex flight scenario. Verbal communication and augmented reality (AR) were chosen as means of communication and the verbal communication was carried out in a wizard-of-Oz (WoOz) fashion. The interface was tested in a flight simulator and it's usefulness was evaluated by NASA-TLX and SART questionnaires for workload and situation awareness.

Keywords: Human-computer interaction · Augmented reality · Human-machine interaction.

1 Introduction

Human-computer interaction (HCI) has developed certain common guidelines for traditional interfaces. However with rise of AI and fast-paced evolution of learning-based systems, new types of human-machine interaction (HMI) and interfaces (HMI^2) is needed since these systems do not have the traditional attributes of a computer system, namely due to uncertainty [2, 23]. The models provided by the learning systems may turn out to be inaccurate and will need

† These authors contributed equally to this work.

to be updated possibly through human intervention. There the operator must be provided with an interface to change the parameters and update the model.

Naturally each industry will need their own AI-enabled assistant and interaction models which complicates interface design since following a global paradigm is not possible. This paper focuses on the field of aviation and the problem of assistant system for single pilot operations (SPO). With the increasing numbers of commercial flights in the upcoming years, a shortage of pilots is to be expected. One solution to this problem is reducing the numbers of pilots in the cockpit. Three strategies proposed as a solution include: Single pilot in cruise (SPIC), reduced crew operation (RCO), and single pilot operation (SPO). The major difference between these there is that, SPIC and RCO propose a reduction in number of crew for a long-haul flight whereas in SPO the pilot is alone for the entire duration of the flight. This immediately implicates that for a SPO, highest levels of safety must be maintained at all times.

A major concern in SPO is total pilot incapciation (e.g. due to heart failure) and has been assesd in American airline pilots at 0.045 and impairment rate of 0.013 per 100000 flying hours [8, 9]. There are several solutions in the case of overloading and total or partial incapacitation such as assistant by automated systems, assistance on board the aircraft, or assistance from operator on the ground. For a SPO however, increasing the automation in the cockpit will put an extra burden on the pilot for monitoring the systems and will reinforce the paradox of automation [4, 18]. Nevertheless, an assistant system for SPO should have the characteristics of a co-pilot. A set of functional requirements for such a system were described by Cummings et al. [7]. which includes verbal and nonverbal communication.

For nonverbal communication, AR has been used for many decades in aviation to enhance navigation in aircraft [17]. The purpose of utilizing AR is to increase the pilot's situation awareness (SA) [21] in critical segments of flight such as take-off and landing [15, 20]. Moreover, new vision technologies such as synthetic vision system (SVS) or enhanced vision system (EVS) implemented on head-mounted displays (HMD) can decrease the workload of the pilot and increase their SA [22]. A study by Bailey et al. [3] shows that pilots are capable of handling abnormal situations safely with acceptable performance conditions, but with decrease flight performance and unacceptable workload.

The goal of this paper is to create an assistant system interface that not only is inline with human autonomy team (HAT) [16, 19] but also adds a level on intelligence to the system, thus creating a human-intelligent machine team (HiMT) that serves as a replacement for the co-pilot and mimics cognitive abilities of the human (cognitive assistant). The system is assessed by measuring the SA and workload [10] of the pilots in a complex flight scenario. For the purpose of simplicity, the role of AI and verbal communication is played out by a WoOZ. The paper is structured as follows: Section 2 describes the design of the experiment, flight scenario, the levels of communication and hypothesis followed by section 3 which describes the technical implementation. In sections 4 and 5, the methodology and results of the experiment are presented. The results of the experiment as discussed in section 6, followed by conclusion in section 7.

2 Experiment

The first activity carried out in the design of the experiment and the interface was a preliminary interview with the pilots to understand the tasks carried out and whether they would like to delegate tasks (under normal and abnormal circumstances) with the AI assistant system. In the results, it was clear that pilots do not wish a total handover to the system, rather wish to be presented with sufficient information which helps them with decision making in different phases of the flight and in time constraint.

2.1 Flight scenario

According to accidents statistics [1], 49% of accidents occur during landing and approach phase of the flight. Together with the risk associated with runway (RW) safety [11], a relevant scenario must be chosen which represents the risks of landing and approach. The flight scenario chosen for the experiment is commonly known as the Bremen scenario (Fig. 1) which represents the approach and landing phase of the flight. This scenario was used in the project Future Sky Safety "Human Performance Envelope" [13, 5] with normal two pilot crew in A320 flight simulator. As in this paper the scenario is used in a SPO scenario with AI cognitive assistant, it will allow a comparison of results for workload and situation awareness with the baseline two pilots crew.

The scenario starts at top of decent (TOD), at 32000 feet to Bremen airport, and plays out as follows:

1. The aircraft has 50 minutes left, which is around 2000 kilograms. The initial approach is on runway 27.
2. The air traffic controller (ATC) requests a Go-Around due to a truck stuck on the runway.
3. A shift in wind direction.
4. Electrical failure occurs (AC BUS 1 fault) druing downwind.
5. Approach and landing on new runway. Possibility of running out of fuel.

The complexity in decision making in this scenario is due to time pressure from the amount of fuel remaining which, a high workload, and degraded SA.

2.2 Communication levels of the assistant and hypothesis

Based on the flight scenario, the experiment was divided into four levels of communication:

– No assistant. The pilot flies the aircraft alone without any kind of assistant.

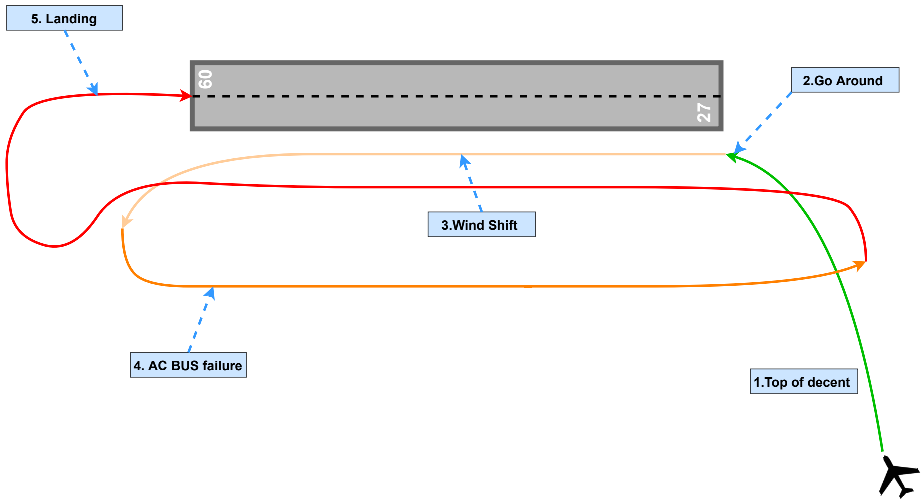

Fig. 1. A representation of Bremen scenario played out in A320 flight simulator for SPO.

- Assistant on request. The assistant is active only if the pilot asks a question or requests help. It does not provide any explanation for the provided information.
- Proactive simple assistant. Provides information through verbal communication and visual cues by AR. It does not provide any explanation for the provided information.
- Proactive evolved assistant. Provides information through verbal communication and visual cues by AR, and provides explanation to the pilot for the provided information and mimics reasoning.

With regards to the flight scenario and communication levels, the main hypothesis are the following:

- AI assistant has not effect on pilot workload and mental demand.
- A proactive assistant provides better situation awareness, with explanation

To verify the hypothesis, NASA-TLX and SART questionnaires were used to assess the cognitive workload and situation awareness of the pilots, and to understand which level of communication for the assistant was the most useful.

3 Implementation

Since the role of the AI assistant is played out by WoOz, the AR interface is implemented on a server-client architecture (Fig. 2).

A restriction imposed on the assistant system is that it doesn't directly communicate with the ATC; however, it is capable of understanding the communication between the pilot and ATC as well as the commands issued by the

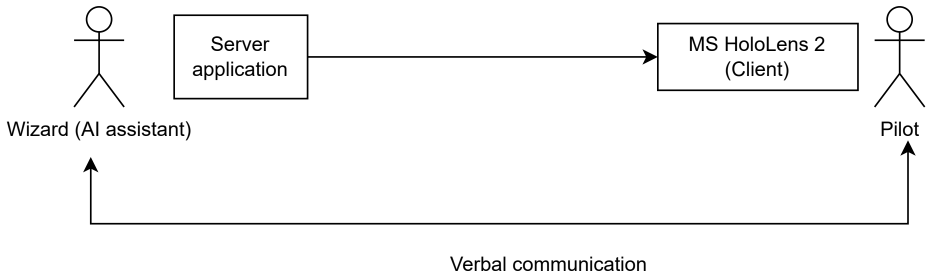

Fig. 2. An overview of AR application architecture.

ATC. The server application for the wizard (AI assistant) was created with C# windows forms and MQTT [6] networking library (Fig. 3).

The application for the server is consisted of two forms. The main form (Fig. 3 left) contains error messages for different phases of the flight, checklists, and units, such as flight control unit (FCU), and electronic centralized aircraft monitoring (ECAM). Some messages are displayed with two colors, either amber or red. indicating the severity of the error or situation which is meant to draw the pilot's attention to an problem or make him aware of an existing one.The second form (Fig. 3 right) is for passing FCU calculations to the pilot.

The AR application was developed with Unity engine together with MQTT broker [14] for receiving messages over the network and Microsoft Mixed Reality Toolkit [12] for enabling interactions in AR environment. The client application contains 3D and 2D visual cues which were enabled and disabled by the wizard through the server application. For the 3D format, the messages where shown in an extended panel (Fig. 4 top left) with the purpose of helping the pilot to keep track of the issues going on at the given time, or highlighted the necessary buttons on the overhead panel to help them speed up processes (Fig. 4 bottom). The inputs for the FCU were displayed directly above the unit, with the numbers on top of their respective fields on the panel (Fig. 4 top right). The 2D cues appeared in the middle-up part of pilot's field of view. To avoid repetitive calibration for each pilot, the 3D cues were pinned to the specific location on the world map generated by HoloLens, using world anchors so that the scene needs to be set up only once.

For verbal communication, the main difference between the simple and evolved assistant lies in the explanation for the information provided. For both versions of the assistant, the pilot can issue commands with keywords such as: "Perform ..." or "Compute ...". Lastly, in the case of the assistant of request, the simple version of the assistant was used.

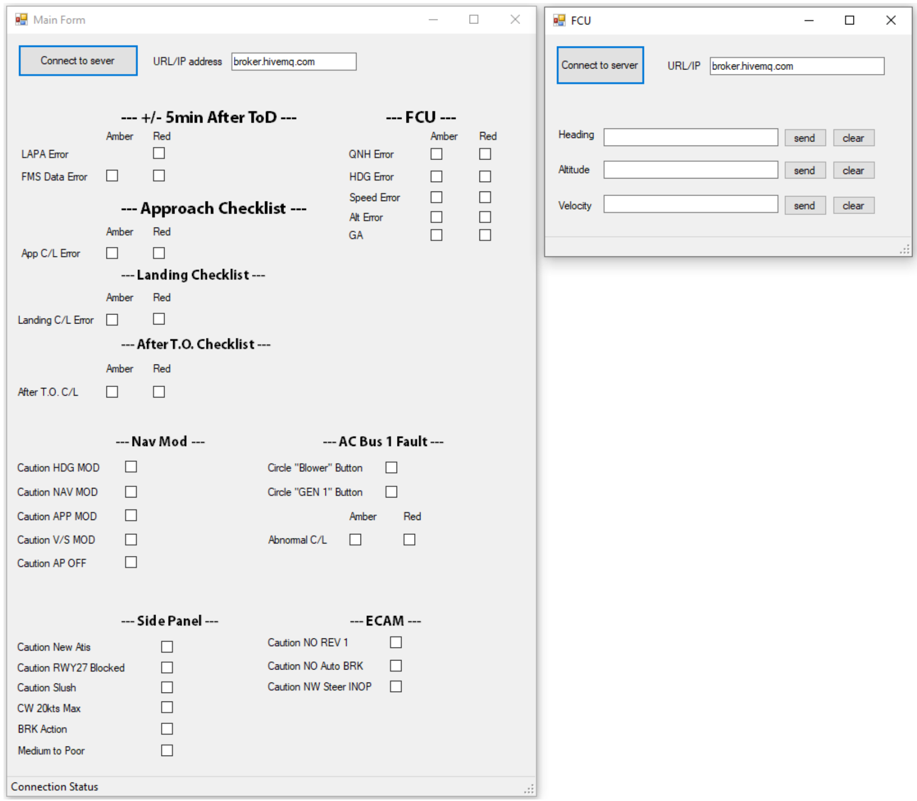

Fig. 3. The server application for wizard (AI assistant) consists of two forms. Left: Main form consisting error messages with indication in color. Left: Calculations for FCU.

4 Methodology

Twenty-four pilots were recruited to participate in the experiment of which two could not finish the experiment properly due to technical problems. The 22 pilots had a mean age of: $36,72$ y.o (SD=$10,88$), a 2975 mean flight hours (fh) experience on A320. 21% of the pilots were female and 61% were first officers, and were evenly distributed over the 4 levels. Ten pilots participated in the baseline (M=30.9 y.o., SD=3.28, M= 3125fh, SD=1557), five in level 1 (M=28.2 y.o., SD=4.38, M= 2240fh, SD=1415), five in level 2 (M=35 y.o., SD=10, M= 2700fh, SD=1987), seven in level (M=35.71 y.o., SD=7.9, M= 3685fh, SD=2814) and five in level 4 (M=48.4 y.o., SD=11.93, M= 3670fh, SD=1616).

The scenario was played out on a cockpit demonstrator located in Bordeaux INP premises together with A320 simulation with the software Prepare 3D (developed by Lockheed Martin). Before each flight session, the pilots were briefed about the simulator specifications (e.g. touch screen) and the flight scenario. For

Fig. 4. An example of AR cues shown to the pilot.

levels two,three, and four, the pilot was introduced to the assistant "Jack", and how they communicate with it and the capabilities it had (e.g. cannot take over, or push buttons). the pilots were presented with a paper summary of all the points. For pilots to familiarize themselves with the simulator, touch screens, and the assistant, a training scenario was conducted which consisted of take-off from Bordeaux, and landing. For experiment scenario, the pilots were briefed on the destination airport (Bremen), and are given the amount of time they need to get to know the destination airport on paper charts or electronic flight bag (EFB).

5 Results

Before data analysis, data was cleaned and assumptions for normality were tested, and violations of normality assumptions 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 four levels compared to the Baseline. All analyses were conducted using Jamovi 1.6.23 statistical software.

5.1 Workload NSA-TLX

To assess pilots' workload during the flight scenario a digital version of the NASA-TLX was used. By incorporating a multi-dimensional rating procedure, the questionnaire derives an overall workload score based on a weighted average of ratings on six subscales [10]: Mental demand, physical demand, temporal demand, performance, effort, and frustration.

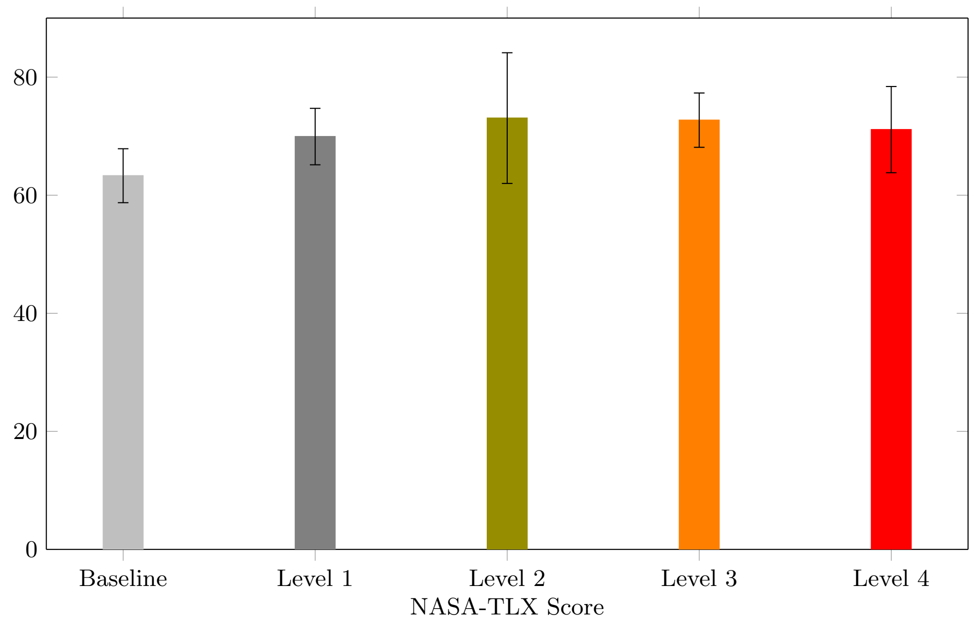

Fig. 5. Overall score of NASA-TLX for workload.

The overall workload did not significantly change between the different levels as shown in Fig. 5, $\chi^2(4) = 3.48, p = 0.482$. The means of the baseline (pilot and copilot) and the level 4 (Evolved Assitant) are close to each other. Baseline ($M = 63, SD = 14.5$) and level 4 ($M = 70.2, SD = 15.7$). The cognitive load score did not increase significantly, only 7.02 points for those with the evolved assistant (Level 4) compared to level 1 ($W = 1.386, p = 0.865$), and the cognitive load score increased by 10.1 points for those in level 2 ($M = 73, 1; SD = 24.8$) (IA on request) compared to the baseline, but this difference is not significant ($W = 1.732, p = 0.737$).

Ratings regarding mental demand (required mental and perceptual activity), showed no significant changes over different levels (Fig. 6) , $\chi^2(4) = 3.28, p = 0.512$.

For the performance (pilot's perception on how successful they executed their tasks), there was increase in level 2 (assistant on request) compared to rest of the levels (Fig. 7). On the other hand, perception of performance decreased in level 3, and 4 ($\chi^2(4) = 3.70p = 0.448,$)

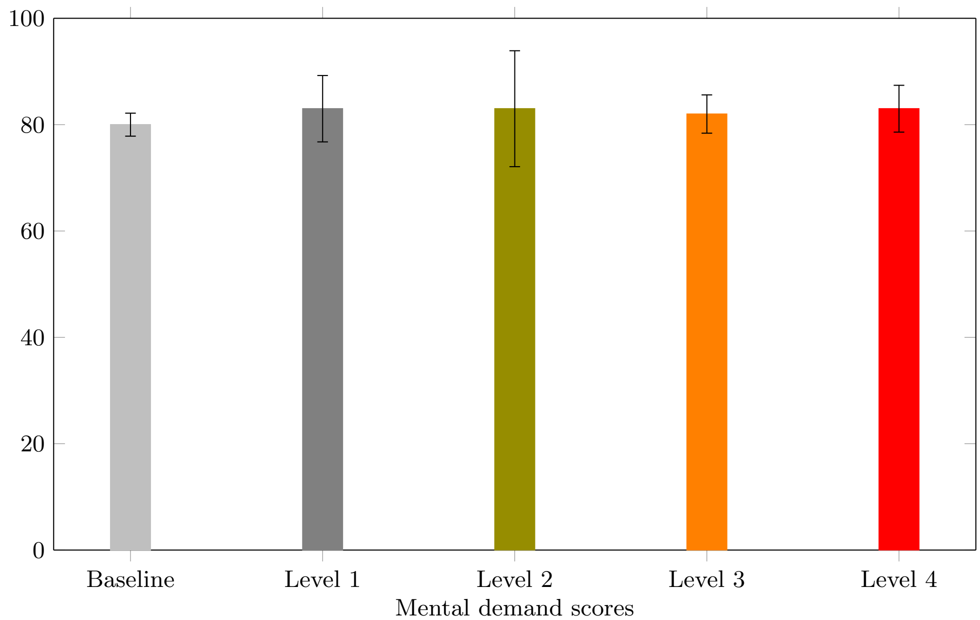

Fig. 6. Score of mental demand for baseline and four levels of assistance.

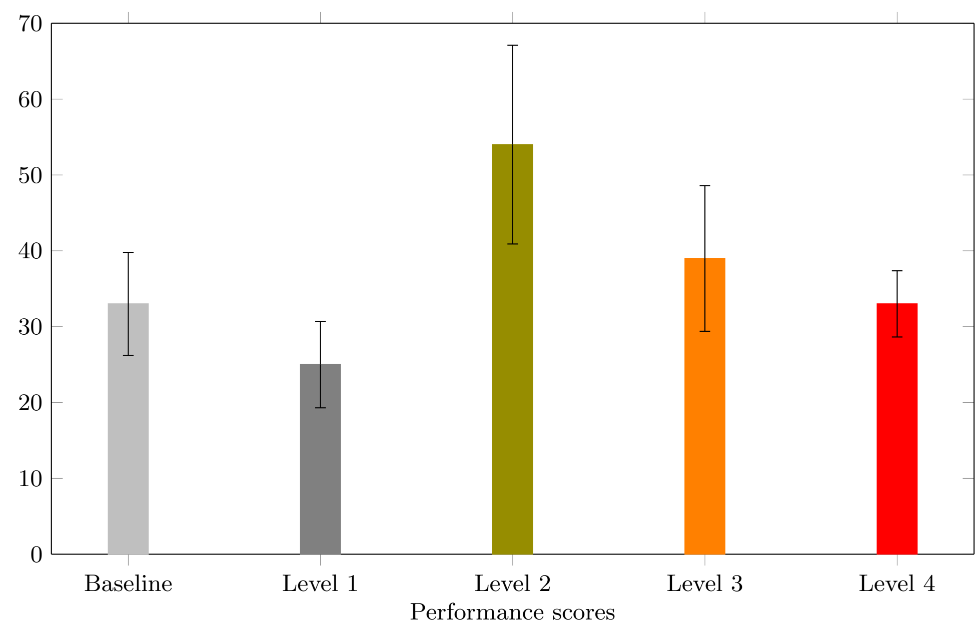

Fig. 7. Score of perception of performance for baseline and four levels of assistance.

5.2 Situation awareness

The overall levels of situation awareness for the four levels are presented in Fig.8. The lowest level of situation awareness was in level 1 followed by level 2. The highest level of situation awareness was in level 4 ($M = 22, SD = 2.12$). The level of situation awareness is even higher than the baseline ($M = 15.1; SD = 3.78$) but the difference between the levels was not significant.

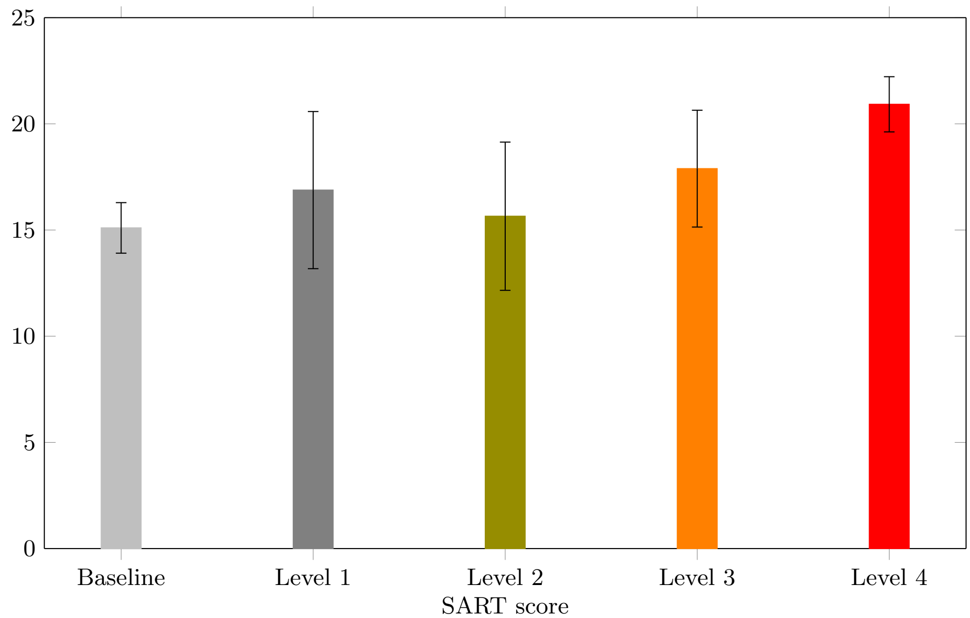

Fig. 8. Score of SART situation awareness.

The attentional demand (Fig. 9) was slightly lower in level 4 compared to the baseline, but the results were not significantly different ($\chi^2(4) = 5.07, p = 0.280$).

6 Discussion

The purpose of the experiment was to evaluate the impact of an AI assistant system, with verbal and AR interface in a SPO scenario. By taking the results into account, a proactive assistant interface provides better SA, even though the difference is not significant. Compared to the baseline (2 pilot crew), the SA was higher in level 3 and highest in level 4, which suggests an interface with explanations about the information provided by the assistant system can improve pilot's SA. Also in levels 3 and 4, the pilot and the assistant system had a better communication and exchanges which induces better understanding. On the other hand, the SPO situation increased the increased the pilots' workload in some cases. This occurs due to SPO scenario not being a familiar situation for

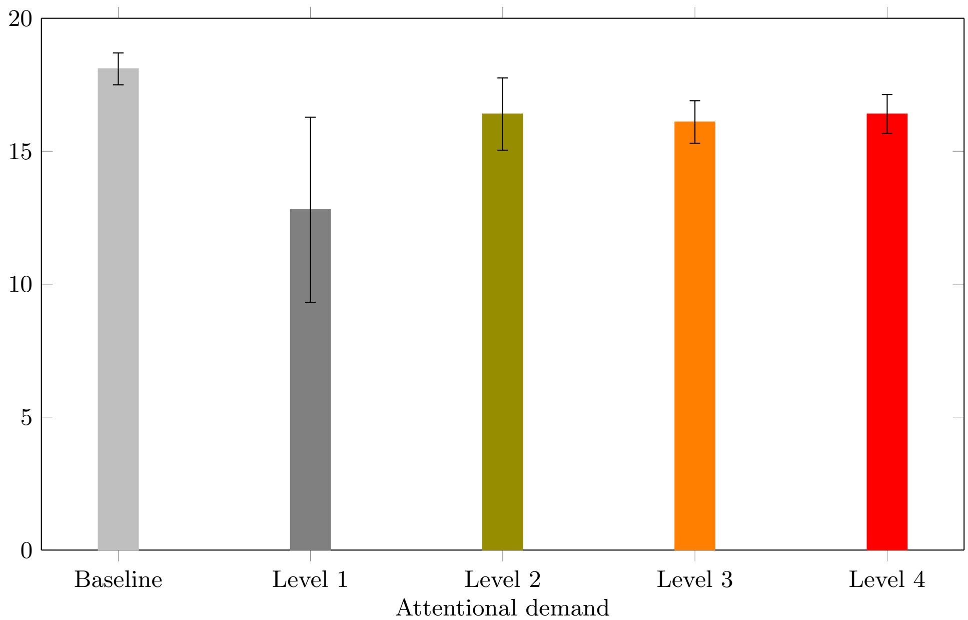

Fig. 9. Attentional demand.

the pilots. Moreover, the use of virtual assistant through AR glasses was also new to them. There is still a long way towards creating a safe and functional assistant system for SPO, however the usefulness of AR and verbal communication was backed up by the data. The future research will focus on replacing the WoOz with a real AI, and systematically increasing it's functionality.

7 Conclusion

AI systems do not necessarily follow the common guidelines for interface design. In this paper, a scenario of SPO was tested to study the collaboration of human pilot with a AI system which was played out as a WoOz together with AR as means of providing visual cues. Three types of communication styles were tested: On request, Simple proactive, and evolved proactive. The scenario was played out in a A320 simulator located in BordeauxINP 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 a proactive interfaces (level 3 and 4) resulted in a better SA and understanding of the situation. Future work will focus on replacing the WoOz with a real AI, with capabilities such as pilot monitoring. Moreover, natural language processing units should be used to enable verbal communication and interface with the AI.

Designing a dedicated AR headset for commercial cockpits is also of importance, as AR played a positive role in assisting the pilots in levels 3 and 4.

Acknowledgements 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 Th´eodore Letouz´e for taking care of the technical aspects of the simulator and Turkan Hentati for help with statistical analysis.

References

1. Airplanes, B.C.: Statistical summary of commercial jet airplane accidents. Worldwide Operations **2008** (1959)
2. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E.: Guidelines for human-AI interaction (may 2019). https://doi.org/10.1145/3290605.3300233
3. Bailey, R.E., Kramer, L.J., Kennedy, K.D., Stephens, C.L., Etherington, T.J.: 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)
4. Bainbridge, L.: Ironies of automation. In: Analysis, design and evaluation of man–machine systems, pp. 129–135. Elsevier (1983)
5. Biella, M., Wies, M., Charles, R., Maille, N., Nixon, J.: How eye tracking data can enhance human performance in tomorrow's cockpit. results from a flight simulation study in future sky safety. In: Joint AIAA and Royal Aeronautical Society (RaeS) Fall Conference on Modeling and Simulation for ATM (2017)
6. Craggs, I.: Mqtt for .net, https://github.com/eclipse/paho.mqtt.m2mqtt https://github.com/eclipse/paho.mqtt.m2mqtt
7. Cummings, M.L., Stimpson, A., Clamann, M.: Functional requirements for onboard intelligent automation in single pilot operations. In: AIAA Infotech@ Aerospace, p. 1652 (2016)
8. DeJohn, C.A., Wolbrink, A.M., Larcher, J.G.: In-flight medical incapacitation and impairment of us airline pilots: 1993 to 1998. Tech. rep., FEDERAL AVIATION ADMINISTRATION OKLAHOMA CITY OK CIVIL AEROMEDICAL INST (2004)
9. DeJohn, C.A., Wolbrink, A.M., Larcher, J.G.: In-flight medical incapacitation and impairment of airline pilots. Aviation, space, and environmental medicine **77**(10), 1077–1079 (2006)
10. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (task load index): Results of empirical and theoretical research pp. 139–183 (1988). https://doi.org/10.1016/S0166-4115(08)62386-9
11. ICAO: Icao accident statistics https://www.icao.int/safety/iStars/Pages/Accident-Statistics.aspx
12. Leigh, K.: Microsoft mixed reality toolkit, https://github.com/microsoft/mixedrealitytoolkit-unity https://github.com/microsoft/MixedRealityToolkit-Unity

13. Letouzé, T., Créno, L., Diaz-Pineda, J., Dormoy, C.A., Hourlier, S., André, J.M.: Mental representation impact analysis (meria), a method for analyzing mental representations for the design of hmi. a case study in aeronautics. Le travail humain **83**(1), 61–89 (2020)
14. Light, R.: Mosquitto mqtt broker, https://github.com/eclipse/mosquitto https://github.com/eclipse/mosquitto
15. Lim, Y., Gardi, A., Sabatini, R., Ramasamy, S., Kistan, T., Ezer, N., Vince, J., Bolia, R.: Avionics human-machine interfaces and interactions for manned and unmanned aircraft. Progress in Aerospace Sciences **102**, 1–46 (2018)
16. Matessa, M., Vu, K.P.L., Strybel, T.Z., Battiste, V., Schnell, T., Cover, M.: Using distributed simulation to investigate human-autonomy teaming. In: International Conference on Human Interface and the Management of Information. pp. 541–550. Springer (2018)
17. Nicholl, R.: Airline head-up display systems: human factors considerations. Available at SSRN 2384101 (2014)
18. Parasuraman, R., Molloy, R., Singh, I.L.: Performance consequences of automation-induced'complacency'. The International Journal of Aviation Psychology **3**(1), 1–23 (1993)
19. Shively, R.J., Lachter, J., Brandt, S.L., Matessa, M., Battiste, V., Johnson, W.W.: Why human-autonomy teaming? In: International conference on applied human factors and ergonomics. pp. 3–11. Springer (2017)
20. Spitzer, C., Ferrell, U., Ferrell, T.: Digital avionics handbook. CRC press (2017)
21. Taylor, R.M.: Situational awareness rating technique (sart): The development of a tool for aircrew systems design. In: Situational awareness, pp. 111–128. Routledge (2017)
22. Viertler, F., Hajek, M.: Evaluation of visual augmentation methods for rotorcraft pilots in degraded visual environments. Journal of the American Helicopter Society **62**(1), 1–11 (2017)
23. Yang, Q., Steinfeld, A., Rosé, C., Zimmerman, J.: Re-examining whether, why, and how human-ai interaction is uniquely difficult to design. In: Proceedings of the 2020 chi conference on human factors in computing systems. pp. 1–13 (2020)