Computers in Human Behavior 64 (2016) 449–454

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Workload perception in drone flight training simulators

Gabriel G. De la Torre*, Miguel A. Ramallo, Elizabeth Cervantes

University of Cadiz, Spain

A R T I C L E I N F O

Article history:
Received 9 April 2016
Received in revised form
14 July 2016
Accepted 22 July 2016
Available online 27 July 2016

Keywords:
Flight simulators
Workload
Training
Drone

A B S T R A C T

Workload perception was measured in a drone flight training Simulator computerized situation. There has been increasing research in recent years on the topic of Remotely piloted aircrafts (RPA). Eleven participants were tested for workload perception during a drone flight simulator training. Reliability, sensitivity and correlations were studied for the workload scale and its relationship with the simulator training tasks. Overall, there were clear effects of mental demand as showed in the workload perception during the training tasks. Reliability for the workload scale showed good score and sensitivity showed mental demand as the most important factor compared to the other parameters measured obtaining highest correlations with landing tasks and number of errors. In our results, we have seen how the AWT (adapted from NASA-TLX) showed good sensitivity in assessing the mental burden of participants. In our research, participants scoring higher in the mental demand subscale showed greater difficulty finishing training tasks, and also showed longer time delays in performing both training sections of the simulation. These types of tools measuring workload perception and virtual training systems can be used in future research, to see how this cognitive aspect affects piloting skills and its possible safety and training implications.

1. Introduction

Unmanned aerial vehicles (UAV), also known as remotely piloted aircrafts (RPA) or more commonly as drones, were initially developed for military purposes. After World War II, there were several countries working on this technology, in order to conduct surveillance without being seen by the enemy and risking human lives.

Until recently, drones had been limited to the military sphere, but nowadays, technological improvements, advances in communications, and battery technology, have ensured that small, low-cost UAVs allow civilians to work and conduct experiments with drones.

These aerial robots are a good solution because they can cover a wide area without touching the ground. Therefore, they can be used to explore, for example, the remains after a catastrophe (Astuti, Longo, Melita, Muscato, & Orlando, 2008). Their high mobility, the possibility of use in environments that are dangerous to humans (Kontitsis, Tsourveloudis, & Valavanis, 2003), and their ability to reduce operating time and improve the identification of causes and effects of crises, make them a useful tool for various types of tasks. These tasks include search and rescue missions using high definition imaging and thermal imaging (Rathinam et al., 2007); analysis of the gas composition within volcanoes (Astuti et al., 2008); surveillance operations including inspection and monitoring of the boundaries of rivers, bridges, and shorelines (Rathinam et al., 2007); fire monitoring in forests (Casbeer, Beard, McLain, Li, & Mehra, 2005); search for ground targets in unknown regions (Xie, Ye, Luo, & Li, 2012); and mapping (Templeton, Shim, Geyer, & Sastry, 2007) among other utilities. The still-expanding civil applications offer a multitude of solutions: review of high-voltage wiring, agriculture, mapping, measuring structures, anthropology, etc.

There has been increasing research in recent years on these small aerial vehicles, most of it with the object of interest being the use of algorithms or hardware specifications to make the drone capable of autonomous operation, or performing different types of tasks more effectively. This approach forgets that although there is no pilot on board these aircrafts, they require significant human interaction.

The fact that there is no human in the vehicle is misinterpreted by some as there is no human in the system; however, RPAs are complex systems that require a lot of human involvement and they

* Corresponding author. Department of Psychology, University of Cadiz, Puerto Real, 11510, Cadiz, Spain.
E-mail address: gabriel.delatorre@uca.es (G.G. De la Torre).

http://dx.doi.org/10.1016/j.chb.2016.07.040
0747-5632/

involve mixed human/robot systems to an extent (Kontitsis et al., 2003). Consequently, the study of human factors related to drone piloting or quadcopter operations may significantly contribute to the performance of tasks.

In literature regarding human factors in RPAs, a bulk of the existing research is based on UAV for military use, excluding the study of variables that can affect the operator of a small quadcopter. Similarly, state air safety agencies are increasingly setting rules on the use of these vehicles for safety and security reasons. This is a clear sign of the importance of the fact that we need to control the use of these vehicles in the airspace, as well as the importance of adequate training for future drivers, where issues related to human factors are included.

Research has shown that while the percentage of aircraft accidents attributed to mechanical failures has decreased dramatically over the past 40 years, the percentage attributable, at least in part, to human error has dropped to an even lower percentage (Shappel & Wiegmann, 2000). This situation shows the need to focus more on working with humans, and the need for people to remain indispensable in the use of UAVs. "Pilot error" is often the reason given for an aircraft accident; however, human error usually has an underlying cause. These causes can include high (or low) workload, fatigue, and poor knowledge of the situation or inadequate training among other causes of which one or some can slow performance and lead to an accident or failure of the objective (Manning, Rash, LeDuc, Noback, & McKeon, 2004).

While automation is being increasingly used in the working of these devices, we cannot forget that automation can also increase the workload of the operator and reduce situational awareness (Ruff, Narayanan, & Draper, 2002). Similarly, high levels of automation can also prevent the operator from quickly intervening to override automation if necessary (McCarley & Wickens, 2005). It is, therefore, very important to note that the automation of various functions should not eliminate human intervention in full (Hopcroft, Burchat, & Vince, 2006).

In human factors research, it appears that complex tasks are performed most successfully when the system is designed to support the needs of human beings instead of removing the human from the system (Abbott, Slotte, & Stimson, 1996]. In many cases, the goal of eliminating the human from the system has led to major system failures, specifically because the system was not designed to support interaction with the human (Casey, 1998). The combination of the strengths of humans and robots to achieve a cooperative task is becoming a popular paradigm (Bruemmer, Few, Nielsen, & Walton, 2007; Crandall & Cummings, 2007; Fong, Thorpe, & Baur, 2003). Adjusting the autonomy levels of the robot to allow human input is a good way to achieve an optimal combination in mixed human/robot teams. The underlying assumption is that the robot performance increases with more human input (Kaupp & Makarenko, 2008). If the human operator is an indispensable factor when working with UAVs, it would be essential to study variables that affect human performance, specifically in tasks where an operator interacts with a rotary UAV, in order to leverage the full potential of the operator to perform the task in the best possible way or even to prevent possible accidents or failures.

Human performance is affected by a variety of influences in both internal and environmental tasks. The functions that modulate human performance are equations derived from empirical data that are used to determine how human performance is affected by the combination and influence of factors found in specific conditions. Examples of these modulating functions are sleep quality and quantity, ambient temperature, stress, and workload (Aasman, Mulder, & Mulder, 1987).

Workload is defined as the combination of the demand for labor and the human response to this demand (Mouloua, Gilson, Kring, & Hancock, 2001). The assessment of workload is a key point in the research and development of systems for human-machine communication in order to ensure the safety, health, comfort, efficiency, and long-term success of the operator (Rubio, Díaz, Martín, & Puente, 2004). Workload levels vary considerably between extended periods of low workload and intense periods of high workload. An effective work design or schedule usually aims to avoid extremes of high or low demand, which can be a threat to the maintenance of skills (Sauer, Wastell, & Hockey, 1996). On the other hand, prolonged periods of high workload may result in reduced attention, increased stress, fatigue, reduced flexibility, and information processing deficits (Connors, Harrison, & Akins, 1985; Hockey, 1993).

Continuous periods of high workload increase fatigue, especially after multiple periods of total loss of sleep, long periods of sleepiness, or sleep fragmentation. This degrades the performance, productivity, safety, and effectiveness of the mission. Moreover, this loss of sleep combined with high workload reduces reaction time and decreases alertness (Kmeger, 1999). High physical and mental demands can also cause more errors due to increased fatigue and loss of concentration (Schuetz et al., 2008).

There are a number of tools to assess and predict mental workload. Most of these methods are divided into the following categories: (a) measures based on performance, (b) subjective measures, and (c) physiological measures (Meshkati, Hancock, & Rahimi, 1992). Of these, subjective measures are becoming increasingly important as assessment tools and have been widely used to assess the workload of the operator. The reasons for the frequent use of subjective methods include practical advantages (ease of implementation, no intrusion) and ongoing data, which support the ability of subjective methods to provide sensitive measures of the mental load of the operator (Rubio et al., 2004). There are several subjective tools for measuring mental workload; the most commonly used are the Cooper-Harper Scale, the Bedford Scale (Cooke & Mesa, 2006; Cooper & Harper, 1969; Roscoe, 1987; Roscoe & Ellis, 1990), the Subjective Workload Assessment Technique (SWAT; Reid & Nygren, 1988), and the NASA-Tax Load Index (NASA-TLX; Hart & Staveland, 1988). The information provided by these subjective scales can be a valuable source of information on mental workload in two fundamental ways. First, they can be used to identify specific sources of demand that a specific task may have. Second, they can reveal differences in workload between two or more individuals (Yurko, Scerbo, Prabhu, Acker, & Stefanidis, 2010).

For this study, we decided to use an adapted version of the NASA-TLX. We took into account the facts that the psychometric properties of this test are well documented, the test has been validated, and it has been previously used by the AMES Human Performance Research Laboratory Group at NASA as a tool for the subjective assessment of individual workload in real flights, flight simulation tasks (Battiste & Bortolussi, 1988; Corwin, 1989; Nataupsky & Abbott, 1987; Nygren, 1991; Shively et al., 1987; Tsang & Johnson, 1989; Tsang & Velazquez, 1996), and even unmanned vehicles (Byers, Bittner, Hill, Zaklad, & Christ, 1988). NASA-TLX has greater sensitivity compared to the computer versions of other scales (Hill et al., 1992) and has the ability to modify language and adjust questions to suit specific tasks and needs (Cao, Chintamani, Pandya, & Ellis, 2009).

Evidence suggests that human error is a major contributing factor to accidents in commercial aviation (Wiegmann & Shappell, 2001). Common errors that are also applicable to drone pilots include not initiating the appropriate maneuvers, failing to notice visual and auditory alerts, being unable to maintain good situation awareness, and poor decision-making. Identifying the cognitive factors and underlying neural circuitries that are predictive of pilot errors is a great challenge, as flying is a complex but necessary activity that takes place in a rapidly changing and uncertain environment. According to Hardy and Parasuraman (1997), a pilot's flying performance is dependent on domain-independent knowledge (e.g., cognitive functions), domain-dependent knowledge (e.g., procedural knowledge), pilot stressors (e.g., adverse weather conditions), and pilot characteristics (e.g., age, expertise). Numerous studies have linked flight functioning and cognitive performance, and different measurements of cognitive efficiency have been identified as crucial to piloting ability, including time-sharing (Tsang & Shaner, 1998), speed of processing (Taylor, Yesavage, Morrow, Dolhert, Brooks, & Poon, 1994), attention (Knapp & Johnson, 1996), and problem solving (Wiggins & O'Hare, 1995). The notion that specific aspects of cognition play a crucial role in the chain of events leading up to drone crashes suggests that the implementation of efficient screening of cognitive procedures for pilot selection, training, and the development of monitoring and instrumentation systems fitted for the human brain may decrease accident rates and maximize success.

2. Method

2.1. Participants

All the 11 participants, who volunteered to participate in the study, were adults with no previous experience in piloting planes or drones. They had no previous experience with radio control devices or toys. The average age of the study group was 31 years (SD 4.92); 45.5% of the sample was female and 54.5% male. None of the participants suffered from any serious sensory deficit that could prevent him/her from performing the experiment, and all of them had good physical and mental health.

2.2. Materials

An *AeroSIM RC* ©virtual pilot training system for Windows platforms (Fig. 1) was used for the experiment. We used a Phantom activated GPS quadcopter model option operated by a real handheld radio controller. All the participants completed the tutorial, which comprised 24 different tests encompassing five areas: gas (management of lifting power of the aircraft), displacement or movement (omnidirectional movement), steady or stationary flight (maintenance of the position and altitude), toward or forward (stabilization of the aircraft), and landing. The radio station used was a *DX5e Spektrum* 2.4 GHz. Each test had to be performed in a given time.

2.3. Axon Workload Test (AWT)

The AWT is a test adapted from NASA-TLX software (Hart & Staveland, 1988). This adaptation was done in Spanish and it was also adapted for the tasks of controlling a drone with handheld radio controllers. This questionnaire represents a method of multidimensional assessment that provides an overall workload score based on a weighted average of scores in six subscales: Physical Demand (PD), Time Demand (TD), Effort (EF) Performance (PF), Mental Demand (MD), and level of Frustration (FR). Of these, three relate to the demands imposed on the person (mental, physical, and time demands) and the others relate to the interaction of the person with the task (effort, frustration, and performance).

The test consists of two parts: score (ratings) and weight Participants obtain the score for each subscale after the completion of the task. A numerical score in the range of 0–100 is assigned to each subscale. Participants determine the weighting by selecting the workload subscale more relevant to them from a couple of options. The weighting is calculated from 15 pairs of combinations created from the six subscales. The test was reverse-translated by language experts in order to be effectively applicable for the Spanish-speaking participants, and the subscales were modified to fit the specific type of task that participants would be performing in the virtual simulation (Cao et al., 2009). The descriptions of the subscales and weighting options can be seen in Table 1.

2.4. Procedure

Before participating in the simulation study, all participants were asked for demographic data including general health status. In addition, a 1-h training session was provided to each participant, to teach them the use of controls in the handheld control system

Fig. 1. AEROSIM RC control screenshot. Credit: AEROSIM RC.

Table 1
Axon Workload Test scales (AWT).

Scales	Score	Text
Performance	Good/Bad	How good do you think you performed the task?
Time	Low/High	How stressful was the task in regard of the time of completion required?
Frustration	Low/High	How insecure, irritated, stressed or discomforted did you against relaxed, confident, satisfied or happy did you feel while performing the task? (the lower values represent the negative while the higher values represent the positive)
Mental Demand	Low/High	What do you think would be the difficulty level if the task you just performed? (lowest value is extremely easy while highest value represents the biggest difficulty or complexity.
Effort	Low/High	How much effort did the task required for you? Lowest value represents no effort at all while the highest value represents an extremely big effort.
Physical Demand	Low/High	How much of physical activity demand did the task required to be completed? (including pressing, pulling, turning, controlling, activating and holding) Lowest value represents minimum or no physical activity at all and highest value an extremely high physical demand to do the task.

and the basic operation instructions. After this briefing session, the 24 flight tests commenced. At all times, participants were accompanied by an instructor to resolve any doubts they may have about the controls or the specific type of task to perform. They were given two conditions: high autonomy and low autonomy.

In the high autonomy condition, after failing three times at the same task, participants moved to the next. However, in the low autonomy condition, they did not have this option and had to complete all the tests in the order established by the software *Aero Sim-Rc*. Every participant had this possibility during the training. An hour after completing the tutorial, we paused for participants to take the AWT, and right after its completion, they continued with the simulation tasks. Participants who completed the tutorial in the first hour of using the simulator filled out the AWT at the time they finished. In all cases, the recommended procedure was followed, and the AWT was taken not more than 15 min from the break/end of the simulator tasks (Moroney, Biers, Eggemeier, & Mitchell, 1992). Data regarding the number of tests successfully completed until the pause/end of tutorial (PT1), the number of successful tasks after the pause (PT2), the time in minutes for both periods of simulation tasks, and errors by batch and total were collected. An error can be caused by not completing a task in the required time or by irreparable damage to the aircraft due to a crash.

3. Results

For the AWT, we followed the recommendations of Cha and Park (1997) and Park and Cha (1998) to measure the reliability using Cronbach's alpha and it showed a good level of reliability (Cronbach's $\alpha = 0.82$).

We also calculated the descriptive statistics for the AWT subscales and error rates for simulator tasks. We found that mental demand was the scale with higher scores ($M = 73.18$, $SD = 13.65$) and landing errors were more common among flight simulator tasks. ($M = 70.00$, $SD = 45.47$) (Table 2).

Following this analysis, we performed Pearson correlations between AWT subscale scores and flight simulator tasks and observed that the mental demand AWT scale had significant correlations with the landing task (Land) ($r = 0.66$, $p < 0.05$) and total errors (Tot) ($r = 0.71$, $p < 0.05$) in flight simulator tasks. There was also a strong correlation among the AWT scales in general (Table 3).

We tested sensitivity and specificity of the AWT mental demand variable, using a linear (step) regression analysis. The regression model was statistically significant χ^2 (9) = 108.88, $p < 0.001$. The model explained 55.5% (Nagelkerke R^2) of the variance (Table 4). As for the low versus high autonomy condition we found no significant differences between both.

Table 2
Descriptive scores for AWT Scales and Simulator Tasks Errors.

AWT scales	Mean (SD)	Simulator tasks	Mean (SD)
Mental Demand	73.18 (13.65)	Gas errors	4.72 (5.25)
Physical Demand	22.72 (20.78)	Stationary errors	4,90 (4.59)
Time Demand	57.72 (26.40)	Movement errors	14.54 (15.41)
Effort	61.81 (22.72)	Landing errors	70.00 (45.47)
Performance	52.27 (23.27)	Forward flight errors	39.72 (35.02)
Frustration Level	49.09 (28.26)		
Global Score	57.30 (19.39)	TOTAL	133.90 (76.36)

4. Discussion

The use of simulators for training of future pilots offers several advantages, the most important being able to train a pilot at a lower cost, in real-time, and the possibility of practicing in new situations. This is important due to the risk of inexperienced pilots causing serious accidents in actual flight conditions. In addition, it allows us to evaluate future pilots and study different data parameters, which would be risky and difficult in real flight situations.

Research literature shows the similarity of data obtained from actual flight and simulations, making clear the benefit and utility of these types of studies (Staffan, 2002; Veltman, 2002).

The use of simulators allows us to modify different conditions as mentioned before. In the case of simulators for drones or UAVs such as *RC Aerosim or Heli X6*, we can change and customize several parameters including wind speed, flight mode, or even duration time of batteries, providing us with a great number of variables we can control for research and training.

Of all the tasks of the simulator system used in this research, those related to landing were the ones that showed higher error rates. This was also the task with higher mental demand. Results obtained with drones are also consistent with previous findings with helicopter flight simulators and NASA TLX (García-Mas, Ortega, Ponseti, de Teresa, & Cárdenas, 2015) and combat helicopter flight simulators with NASA bipolar rating scale (precursor for NASA TLX) (Haworth, Bivens, & Shively, 1986). Hart and Hauser (1987) studied the mental burden on civil aviation pilots during routine missions, observing high peak workload during takeoff and landing. It is also known that takeoff and landing operations produce higher levels of stress (Endsley & Strauch, 1997).

Data obtained with simulators is frequently similar to data obtained in real flight situations (Staffan, 2002; Veltman, 2002). Hence, we see a need to continue with this type of research because of the importance in studying how humans perform when interacting with these new air vehicles. This is important for different reasons: First, an operator's good mental and physical health can affect the number of accidents, outcome of the task, or missions in military settings. Second, although the management of such UAVs is simpler than that of a plane, there are specific difficulties

Table 3
Correlations for AWT scales and Flight Simulator Tasks: AWT Scales: 1) MD: Mental Demand; 2) PD: Physical Demand; 3) TD: Time Demand; 4) PF: Performance; 5) EF: Effort; 6) FR: Frustration and 7)SC: Global Score. Flight Simulator Tasks: 1) gas, 2) Stat: Stationary flight; 3) Tran: translational flight; 4) Land: Landing; 5) fwd: Forward flight and 6) Tot: total errors.

	MD	PD	TD	PF	EF	FR	SC	Gas	Stat	Tran	Land	fwd	Tot
MD													
PD	−0.10												
TD	0.54*	−0.08											
PF	0.55	−0.35	0.47										
EF	0.71*	0.26	0.68*	0.27									
FR	0.41	−0.20	0.82**	0.43	0.50								
SC	0.74**	−0.13	0.91**	0.67*	0.75**	0.84**							
Gas	0.30	−0.46	0.15	0.08	−0.04	0.24	0.14						
Stat	0.20	−0.24	0.08	0.18	−0.01	0.41	0.21	0.55					
Tran	0.42	−0.15	−0.23	0.31	0.16	−0.18	−0.01	0.19	0.28				
Land	0.66*	−0.48	0.51	0.38	0.36	0.58	0.53	0.60	0.44	0.47			
Fwd	0.44	−0.23	0.57	0.27	0.28	0.32	0.43	0.56	0.12	−0.25	0.32		
Tot	0.71*	−0.47	0.53	0.43	0.37	0.50	0.54	0.75**	0.48	0.39	0.90**	0.64*	

Table 4
Linear Regression for mental demand of AWT.

Model	Wald χ^2	β	Se	P
1	0.55	−0.74,5	3.11	0.009

associated with drone piloting, such as the sensory isolation between pilot and machine. The operator can only see what the drone sees through a computer screen, whereas aircraft pilots have access to much more information through multisensory stimuli that help them to understand how the aircraft is performing and to know about the environment (Draper, Geiselman, Lu, Roe, & Haas, 2000). For example, operators of unmanned aircrafts or drones neither have access to the vestibular signals that manned aircraft pilots use to gain an understanding of the orientation of the aircraft nor do they have access to kinesthetic cues they use to gain an understanding of turbulence, weather conditions, movement of the aircraft, or gravitational forces (Draper et al., 2000).

While our study shows significant data about the importance of mental demand tasks performed with simulators for training of drone pilots, we do need to mention some limitations. One limitation was the sample size, as it was difficult to find interested commercial drone pilots, apart from casual users.

In future research, it would be interesting to compare the performance of virtual and actual flight situations, analyze system variables and relationships between human/machine mixed teams, and study aspects including extreme environments and the relationship between training hours and mental demand perception. Regarding the specific role of virtual simulation in a training system, it would be interesting to see possible implications and relationships of previous video gaming and virtual reality experience on task performance in these virtual computer drone flight training systems.

5. Conclusion

In our results, we have seen how the AWT (adapted from NASA-TLX) showed good sensitivity in assessing the mental burden of participants. These types of tools can be used in future research, to see how this cognitive aspect affects piloting skills and its possible safety and training implications. To our knowledge, this is the first available tool of this type in Spanish.

We obtained significant scores for perception of high mental demand in close relationship to task performance outcome, which is consistent with previous research (Schuetz et al., 2008; Weinger, Herndon, & Gaba, 1997). This can be a considered an important factor in the training of drone pilots. In our research, participants scoring higher in the mental demand subscale showed greater difficulty finishing training tasks, and also showed longer time delays in performing both training sections of the simulation. They also showed a higher number of errors in comparison to participants who obtained lower scores in the mental demand scale of the AWT.

Acknowledgments

This study would not have been possible without the 11 volunteers who participated in the study. In particular we would like to acknowledge Manuel Guillén from AeroSIM RC for his collaboration.

References

Aasman, J., Mulder, G., & Mulder, L. J. (1987). Operator effort and the measurement of heart-rate variability. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 29*(2), 161–170.

Abbott, K. A., Slotte, S. M., & Stimson, D. K. (June 1996). *Federal Aviation Administration Human Factors Team Report on: The Interfaces Between Flightcrews and Modern Flight Deck Systems*. Technical Report (pp. D1–D3). Washington. DC: Federal Aviation Administration [Online]. Retrieved from http://www.faa.gov/avr/afs/interfac.pdf.pp.

Astuti, G., Longo, D., Melita, C. D., Muscato, G., & Orlando, A. (2008). HIL tuning of UAV for exploration of risky environments. *International Journal of Advanced Robotic Systems, 5*(4), 419–424.

Battiste, V., & Bortolussi, M. (1988). Transport pilot workload: A comparison of two subjective techniques. In *Proceedings of the human factors society thirty- second annual meeting* (pp. 150–154). Santa Monica. CA: Human Factors Society.

Bruemmer, D., Few, D. A., Nielsen, C., & Walton, M. C. (2007). Mixed-initiative control for collaborative countermine operations. *IEEE Transactions on Robotics*.

Byers, J. C., Bittner, A. C., Hill, S. G., Zaklad, A. L., & Christ, R. E. (1988). Workload assessment of a remotely piloted vehicle (RPV) system. In *Proceedings of the human factors society thirty-second annual meeting* (pp. 1145–1149). Santa Monica. CA: Human Factors Society.

Cao, A., Chintamani, K. K., Pandya, A. K., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods, 41*(1), 113–117.

Casbeer, D. W., Beard, R. W., McLain, T. W., Li, S. M., & Mehra, R. K. (2005, June). Forest fire monitoring with multiple small UAVs. In *American control conference, 2005. Proceedings of the 2005* (pp. 3530–3535). IEEE.

Casey, S. (1998). *Set phasers on stun and other true tales of design. Technology and human error* (2nd ed.). Santa Barbara. CA: Aegean Publishing Co.

Cha, D., & Park, P. (1997). User required information modality and structure of in-vehicle navigation system focused on the urban com-muter. *Computers & Industrial Engineering, 33*, 517–520.

Connors, M. M., Harrison, A. A., & Akins, F. R. (1985). *Living aloft: Human requirements for extended spaceflight*.

Cooke, N. J., & Mesa, A. Z. (2006). Human factors of remotely operated vehicles. In *Proceedings of the human factors and ergonomics society 50th annual meeting*. USA: SA Technologies.

Cooper, G. E., & Harper, R. P., Jr. (1969). *The use of pilot rating in the evaluation of aircraft handling qualities (No. AGARD-567)*. Advisory Group for aerospace research and development Neuilly-Sur-Seine (France).

Corwin, W. H. (1989). In-flight and post-flight assessment of pilot workload in commercial transport aircraft using SWAT. In *Proceedings of the fifth symposium on aviation psychology* (pp. 808–813). Columbus. OH: Department of Aviation. Ohio State University.

Crandall, J. W., & Cummings, M. L. (2007, March). Developing performance metrics for the supervisory control of multiple robots. In *Proceedings of the ACM/IEEE international conference on human-robot interaction* (pp. 33–40). ACM.

Draper, M. H., Geiselman, E. E., Lu, L. G., Roe, M. M., & Haas, M. W. (2000, July). Display concepts supporting crew communication of target location in unmanned air vehicles. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 44, pp. 85–88). SAGE Publications. No. 13.

Endsley, M., & Strauch, B. (1997). Automation and situation awareness- The accident at Cali, Colombia. In International Symposium on Aviation Psychology, 9th, Columbus, OH (pp. 877–881).

Fong, T., Thorpe, C., & Baur, C. (2003). Robot, asker of questions. *Robotics and Autonomous Systems, 42*(3), 235–243.

García-Mas, A., Ortega, E., Ponseti, J., de Teresa, C., & Cárdenas, D. (2015). Workload and cortisol levels in helicopter combat pilots during simulated flights. *Revista Andaluza de Medicina del Deporte, 9*(1), 7–11, 2016.

Hardy, D. J., & Parasuraman, R. (1997). Cognition and flight performance in older pilots. *Journal of Experimental Psychology: Applied, 3*(4), 313.

Hart, S. G., & Hauser, J. R. (1987). Inflight application of three pilot workload measurement techniques. *Aviation, Space, and Environmental Medicine*.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183.

Haworth, L., Bivens, C., & Shively, R. (1986). An investigation of single-piloted advanced cockpit and control configurations for nap-of-the-earth helicopter combat mission task. In *Proceedings of the 1986 meeting of the american helicopter society* (pp. 657–672). Washington, DC: American Helicopter Society.

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L., & Christ, R. E. (1992). Comparison of four subjective workload rating scales. *Human Factors, 34*, 429–439.

Hockey, G. R. J. (1993). *Cognitive-energetically control mechanisms in the management of work demands and psychological health*.

Hopcroft, R., Burchat, E., & Vince, J. (2006). *Unmanned aerial vehicles for maritime patrol: Human factors issues (No. DSTO-GD-0463)*. Defence science and technology organization Edinburgh (Australia) Air operations division.

Kaupp, T., & Makarenko, A. (2008, May). Measuring human-robot team effectiveness to determine an appropriate autonomy level. In *Robotics and automation, 2008. ICRA 2008. IEEE international conference on* (pp. 2146–2151). IEEE.

Kmeger, G. P. (1999). *Sustained military performance in continuous operations: Combatant fatigue.rest and sleep needs*. Fort Rucker. AL: U.S. Army Aeromedical Research Laboratory. USAARL Report No. 91–19.

Knapp, C. J., & Johnson, R. (1996). F-16 class a mishaps in the US air force, 1975-93. *Aviation, Space, and Environmental Medicine, 67*, 777–783.

Kontitsis, M. P., Tsourveloudis, N. C., & Valavanis, K. P. (2003, June). A UAV based automated airborne surveillance system. In *11th mediterranean conference on control and automation* (Rhodes, Greece).

Manning, S. D., Rash, C. E., LeDuc, P. A., Noback, R. K., & McKeon, J. (2004). *The Role of human causal factors in US army unmanned aerial vehicle accidents (No. USAARL-2004-11)*. Army Aeromedical Research Lab Fort Rucker AL.

McCarley, J. S., & Wickens, C. D. (2005). *Human factors implications of UAVs in the national airspace*. University of Illinois at Urbana-Champaign, Aviation Human Factors Division.

Meshkati, N., Hancock, P., & Rahimi, M. (1992). Techniques in mental workload assessment. In J. Wilson, & E. Corlett (Eds.), *Evaluation of human work. A practical ergonomics methodology* (pp. 605–627). London: Taylor & Francis.

Moroney, W. F., Biers, D. W., Eggemeier, F. T., & Mitchell, J. A. (1992, May). A comparison of two scoring procedures with the NASA task load index in a simulated flight task. In *Aerospace and electronics conference, 1992. NAECON 1992., proceedings of the IEEE 1992 national* (pp. 734–740). IEEE.

Mouloua, M., Gilson, R., Kring, J., & Hancock, P. (2001, October). Workload, situation awareness, and teaming issues for UAV/UCAV operations. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 45, pp. 162–165). SAGE Publications. No. 2.

Nataupsky, M., & Abbott, T. S. (1987). Comparison of workload measures on computer-generated primary flight displays. In *Proceedings of the human factors society thirty-first annual meeting* (pp. 548–552). Santa Monica. CA: Human Factors Society.

Nygren, T. E. (1991). Psychometric properties of subjective workload techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors, 33*(1), 17–33.

Park, P. & Cha, D. (1998). Comparison of subjective mental workload assessment techniques for the evaluation of in-vehicle navigation system usability. Paper presented at Session T56 of the 5thWorld Congress on Intelligent Transport Systems 1998, Seoul.

Rathinam, S., Almeida, P., Kim, Z., Jackson, S., Tinka, A., Grossman, W., et al. (2007, July). Autonomous searching and tracking of a river using an UAV. In *American control conference, 2007. ACC'07* (pp. 359–364). IEEE.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload* (pp. 185–218). Amsterdam: Elsevier.

Roscoe, A. H. (1987). *The practical assessment of pilot workload. AGARD-AG-282*. Neuilly Sur Seine. France: Advisory Group for Aerospace Research and Development.

Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale assessing pilot workload in flight. A decade of practical use*. Royal Aerospace Establishment. Technical Report 90019. Farnborough. UK: Royal Aerospace Establishment.

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology, 53*(1), 61–86.

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence: Teleoperators and Virtual Environments, 11*(4), 335–351.

Sauer, J., Wastell, D. G., & Hockey, G. R. J. (1996). Skill maintenance in extended spaceflight: A human factors analysis of space and analogue work environments. *Acta Astronautica, 39*(8), 579–587.

Schuetz, M., Gockel, I., Beardi, J., Hakman, P., Dunschede, F., Moenk, S., et al. (2008). Three different types of surgeon-specific stress reactions identified by laparoscopic simulation in a virtual scenario. *Surgical Endoscopy, 22*(5), 1263–1267.

Shappel, S. A., & Wiegmann, D. A. (2000). *The human factors analysis and classification system–HFACS (No. DOT/FAA/AM-00/7)*. US Federal Aviation Administration, Office of Aviation Medicine.

Shively, R., Battiste, V., Matsumoto, J., Pepiton, D., Bortolussi, M., & Hart, S. G. (1987). In flight evaluation of pilot workload measures for rotorcraft research. In *Proceedings of the fourth symposium on aviation psychology* (pp. 637–643). Columbus. OH: Department of Aviation. Ohio State University.

Staffan, M. (2002). Similarities and differences in psychophysiological reactions between simulated and real Air-to-ground missions. *The International Journal of Aviation Psychology, 12*(1), 47–61. January 2002.

Taylor, J., Yesavage, J., Morrow, D., Dolhert, N., Brooks, J., III, & Poon, L. (1994). The effects of information load and speech rate on younger and older aircraft pilots' ability to execute simulated air-traffic controller instructions. *The Journal of Gerontology, 49*, 191–200.

Templeton, T., Shim, D. H., Geyer, C., & Sastry, S. S. (2007, April). Autonomous vision-based landing and terrain mapping using an MPC-controlled unmanned rotorcraft. In *Robotics and automation, 2007 IEEE international conference on* (pp. 1349–1356). IEEE.

Tsang, P. S., & Johnson, W. W. (1989). Cognitive demand in automation. *Aviation. Space. and Environmental Medicine, 60*, 130–135.

Tsang, P. S., & Shaner, T. L. (1998). Age, attention, expertise, and time-sharing performance. *Psychology and Aging, 13*, 323–347.

Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics, 39*(3), 358–381.

Veltman, J. A. (2002). A comparative study of psychophysiological reactions during simulator and real flight. *The International Journal of Aviation Psychology, 12*(1), 33–48.

Weinger, M. B., Herndon, O. W., & Gaba, D. M. (1997). The effect of electronic record keeping and transesophageal echocardiography on task distribution, workload, and vigilance during cardiac anesthesia. *The Journal of the American Society of Anesthesiologists, 87*(1), 144–155.

Wiegmann, D. A., & Shappell, S. A. (2001). Human error perspectives in aviation. *The International Journal of Aviation Psychology, 11*(4), 341–357.

Wiggins, M., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied, 1*, 305–320.

Xie, S., Ye, Z., Luo, J., & Li, H. (2012). Cooperative searching for ground targets with multiple UAVs in unknown region. *Journal of Convergence Information Technology, 7*(23).

Yurko, Y. Y., Scerbo, M. W., Prabhu, A. S., Acker, C. E., & Stefanidis, D. (2010). Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simulation in Healthcare, 5*(5), 267–271.