Applied Computing and Informatics (2016) 12, 16–26

Saudi Computer Society, King Saud University
Applied Computing and Informatics
(http://computer.org.sa)
www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Computer-aided assessment of aviation pilots attention: Design of an integrated test and its empirical validation

Rosario Cannavò [a], Daniela Conti [b], Alessandro Di Nuovo [c,d,*]

[a] *Flight Simulation and Training Devices Unit, Cognitive Technologies and Services SRL, Catania, Italy*
[b] *Doctorate School of Neuroscience, University of Catania, Catania, Italy*
[c] *Faculty of Engineering and Architecture, Kore University of Enna, Enna, Italy*
[d] *School of Computing and Mathematics, Plymouth University, Plymouth, United Kingdom*

Received 27 December 2014; revised 3 May 2015; accepted 19 May 2015
Available online 24 May 2015

KEYWORDS

Software validation;
Computer assisted performance evaluation;
Attention assessment;
Aviation pilot selection and training

Abstract Attention has a key role in the flight performance of the aviation pilot, therefore it is among human factors commonly used in the evaluation of candidate pilots. In this context, our work aims to define a single integrated instrument able to measure all the distinctive attention factors and to assist the assessment and the training of aviation pilots.
In this paper, we present a battery of seven computerized tests, encompassing classical and innovative solutions inspired by the literature in the field, for the integrated measurement of the attention factors of aviation pilots. The computer software is validated by means of an experimental trial with 50 experienced aviation pilots and 50 untrained people as controls. Statistical analyzes confirm that

* Corresponding author at: B109 Portland Square Building, Plymouth University Campus, Plymouth PL4 8AA, United Kingdom. Tel.: +44 1752586320.
E-mail address: alessandro.dinuovo@plymouth.ac.uk (A. Di Nuovo).
Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.aci.2015.05.002
2210-8327

the instrument can effectively classify aviation pilots, and identify a subset of distinctive attention factors that could be used for monitoring their duty.

1. Introduction

In the typical routine, an aviation pilot has to switch rapidly between different sources of information to construct a mental situation model of what the airplane is doing and how to get it to go where he wants it to go. At the same time, he has to listen to Air Traffic Control (ATC) instructions, often delivered at sustained speed. Then, there is the requirement to interpret the instructions from ATC through the use of paper or electronic charts and to fly, or program the autopilot to fly, the desired route. At the same time, the pilot has to keep track of the weather at his current location, along his route and at his destination.

In this context, it is too crucial that the pilot develops Situation Awareness (SA) that is not specific to a particular type of aircraft and mission, but generalizes across many types of aircraft systems [9]. Many researchers recognize the essential role of attention in SA [8,16,18], as an example Schriver et al. [14] investigated the behavior of pilots with different experience levels and presented results that support the link between greater attention and more effective decision making of aviation pilots. Recently, Carretta [4] confirmed that attention is among factors that have consistently shown a relation to flying performance in a study of the Pilot Candidate Selection Method (PCSM), while King et al. [11] proved that tests of cognitive functioning, that include attention, can predict the training outcomes.

Previous work demonstrated that computerized tests are ideal to assess pilot aptitude [2] and, in particular, to measure attention factors [3]. In fact, many different computerized instruments to measure attention factors are available, but to the best of our knowledge none of these instruments was designed and tested specifically for aviation pilots with validation results published in the scientific literature. Moreover, commercially available computerized tests for aviation pilots usually measure attention indirectly, e.g. via simplified flight maneuver simulations, which involve several factors at the same time, and cannot give a specific attentional profile that can highlight strengths and weaknesses of the candidate pilot. Indeed, the attentional profile can be useful for ad-hoc training as theoretical and empirical evidence exists in support of the argument that both the control of attention and the ability to establish better attention management can be developed with training [10].

We also underline that the training of an aviation pilot is particularly difficult and expensive (the full process for a commercial license can cost up to a hundred thousand of US dollars) and flight schools have a limited number of seats available for each course, because the limited resources for the training (i.e. aircrafts or certified simulators). For this reason, the selection processes are of fundamental importance and aim to favor those candidates that have the highest attitude, i.e. that show the highest potential to be successful.

Starting with this background, our research is aimed to design a specialized instrument to effectively measure aviation pilot attention factors and to give the following contribution: (i) as an objective instrument to assist the selection of most promising candidates; (ii) to identify the factors that they should improve during the training; (iii) to select a subset of variables that have the highest impact on the classification of aviation pilots.

To this end, we designed and implemented a battery of seven computerized tests. The development of a novel battery was required to simplify the administration of the different tests and data collection, and to evaluate some additional factors, like the peripheral foveal vision, that were not studied before. Indeed, the battery has been designed to allow a multidimensional assessment of attention.

The rest of the paper is structured as follows. Section 2 describes the tests battery and details the software implementation. Section 3 reports and discusses the experimental results of the empirical validation of the proposed instrument through a trial with 50 experienced aviation pilots and 50 untrained people as a control group. Finally, Section 4 presents our conclusions.

2. Tests battery description and software implementation

In this section, we present the battery of seven tests designed to identify and assess the distinctive factors of the aviation pilot attention, and we give examples of the mapping with actual flight routine tasks.

Here we define attention as “a concentration of mental activity that allows you to take in a limited portion of the vast stream of information available from both your sensory world and your memory” [13].

Types of attention assessed include selective and divided attention, with auditory and visual stimuli. In particular, among visual stimuli, the emphasis is given to non-central stimuli. The tests differentiate among Central, Mid-peripheral and Far-peripheral focus of the visual attention. Peripheral vision is used to detect objects that are located at an eccentricity of more than nine degrees with respect to the foveal vision [15]. “Far-peripheral” vision exists at the edges of the field of view; “mid-peripheral” vision exists in the middle of the field of view. Test scenarios were designed using real-world pictures and videos to increase the level of ecological validity with real-life experience.

The software program presents a detailed introduction before beginning each test. Next a short interactive demo starts and then the subject is required to repeat until he is confident to go on with the real test. The battery of tests needs from 50 min to 1 h to be completed. The duration was carefully chosen by means of a preliminary pilot study ($N = 4$, 2 pilots, and 2 controls) in order to be challenging enough for the aviation pilots. Tests are presented to the subject automatically in the order shown in the following.

We define: *Reaction time*, which is the time elapsed from the appearance of the stimulus on the screen to the key pressed by the subject; *Error*, which occurs when the subject reacts without any stimulus; *Omission*, which occurs when the subject does not react to the stimulus. Note that in case of omission the reaction time is not calculated.

Stimuli are presented randomly in a window of 2 seconds and they stay on the screen for 2 seconds or until the key is pressed.

2.1. Test 1: simple reaction time

The test of simple reaction measures a purely psycho-physiological attention. The parameter identified in this way defines a useful baseline for the analysis of subsequent tests. The test is composed of two parts: the first part (**1A**), in which the visual stimulus is proposed in the central part of the screen, thus stimulating the use of the focused vision; in the second part (**1B**) of the test the same stimulus is proposed randomly in one of the four corners of the screen, to activate the Far-peripheral vision only.

For both subtests, the software records reaction times, errors and omissions.

2.2. Test 2: multiple search and memory

This test was to measure the attention shifting that, as an example, occurs when, in the course of a flight, the pilot has to execute steps out of the habitual task sequence. The attention's shifting factor can be measured via the multiple search and memory tests, in a version involving the visual-spatial channel, in analogy to the Toulouse-Piéron concentration test [17]. The task consists of the search and identification of random targets, among a set of similar stimuli represented by stylized hands, which show varying numbers of open fingers differently orientated. The stimuli are presented in three blocks of twenty stimuli, for which the subject has to find five targets. Targets' shapes are different in different blocks, to force the continuous change of attentional focus. Total time elapsed and the numbers of errors made by the subjects are recorded.

2.3. Test 3: color-word interference

The "resistance to distraction" factor is investigated by two different adaptations of the Stroop phenomenon, in which the attentional process counteracts and reduces the interference of automation. In the original version of this test, developed by Stroop in 1935, dominant color words written in a different color (e.g. red written in blue) are presented to the subject, which who must name the color in which the word is written (blue), exceeding the interference resulting from the habit of reading the words. Ultimately, the test tends to measure the subject's ability to overcome, with a particular attentional effort, the distraction induced from the semantic meaning of the presented stimulus, which must instead be processed according to the color appearance.

Our software implements an adaptation of this test, which is split into two sequential tasks:

3A. In the first task, which sets a baseline, the subject should recognize a color by pressing on the keyboard the button that is coupled with the color that appears on the screen.

3B. In the second task, which is the real test of interference, words that indicate a color appear filled with a different color (e.g. RED in blue), the subject should neglect the automatic recognition of the word semantic and press the button corresponding to the color in which the word is filled (i.e., in the example, blue).

To clearly identify the buttons to be pressed, colored stickers were applied to the keyboard.

Correct answers, errors, and reaction times to individual stimuli are recorded.

2.4. Test 4: ground interference

This test is designed to measure the ability to discriminate a target from an active background through peripheral vision, resembling the actual scanning of the cockpit instrumentation during the flight duty. An active background gives the interference by producing a distracting effect, which must be suppressed to correctly identify the target. The screen is divided into four quadrants and in each one a video is played during the whole duration of the test. The subject is required to identify the target among five different stimuli proposed. The stimuli are proposed in the peripheral area of the screen and they represent human figures, compatible with the position, size, and behavior with those that appear in the background video.

The software measures the reaction times to stimuli proposed as targets, along with the number of errors and omissions.

2.5. Test 5: divided attention

This test focuses on two parallel tasks: a visual search and auditory recognition. This simulates the usual duty of the aviation pilot, as he should pay attention to the onboard instrumentation and follow instructions from the ATC. This is an example of divided attention, which is the highest level of attention and refers to the ability to respond simultaneously to multiple tasks or multiple task demands. The investigation of this kind of attention is realized by means of a multi-tasking test. Combining two stimuli of different type in the same test we can evaluate the subject's ability to vary the attentional focus and, therefore, to distribute his attention to the management of many different stimuli.

In this test, the subject should press one of the two keys as soon as he perceives the target, which can be among the visual or auditory stimuli that are presented simultaneously.

Visual stimuli are different pictures that can appear randomly in one of the four corners of the screen, this way the subject must activate the Far-peripheral vision, while the background is neutral. In the meanwhile, to distract the subject, random sentences that resemble possible ATC speech, are spoken. The auditory target is the word "go", pronounced by the same voice that reads the ATC sentences.

The software records reaction times, number of correct responses, omissions and errors for visual (**5V**) and auditory (**5A**) stimuli.

2.6. Test 6: digit span

This test aims to assess the domain-general working memory capacity, as this has been identified among the fundamental components of pilots' attention [12], which is involved in the integration of incoming information into the situation model of the flight. We recall that a poor performance in this kind of sequential task is a predictor of failure in those learning tasks that require perception, retention and reproduction of symbols or sounds in a given order sequence [1].

To this end, the software proposes the classic digit span test: series of increasing length of digits are presented to the subject and he is asked to repeat them by pressing the appropriate number keys. The test is divided into two parts: (i) the direct, in which the subject should retype the digits as they appear on the screen; (ii) the inverse, in which the subject is required to retype from the last digit appeared to the first one. The backward part starts right after the forward one is ended.

At each step of the test, the number of digits that sequentially appear on the screen is increased by one. The test begins with three digits and continues up to eight. To access to the following step, the subject should recall all the digits at the current step. If the subject does not correctly recall all the digits, another sequence of the same length is shown. After two faults, the test is finished.

For this test, the software calculates the omissions as the total number available of digits (eight) minus the maximum number of digits correctly recalled by the subject. The error is the total number of errors made (i.e. number of wrong sequences inputted) by the subject.

2.7. Test 7: global vision

The objective of the final test is to evaluate the quality and timing of discrimination of an attentional process that uses the three visual channels and their interrelation. This test extends the ground interference by presenting moving targets that can appear randomly anywhere in the screen.

Table 1 Summary of the data recorded by the software.

Test and variable name	Test short ID	No. of stimuli	Omissions	Errors	Median reaction time (MEDTIME)
Simple reaction time					
Central	1A	40	✓	✓	✓
Peripheral	1B	40	✓	✓	✓
Multiple search	2	20	✓	✓	✓ Total TIME
Color-word interference					
Color discrimination	3A	40	✓	✓	✓
Color-word	3B	40	✓	✓	✓
Ground interference	4	16	✓	✓	✓
Divided attention					
Auditory	5A	16	✓	✓	✓
Vision	5V	22	✓	✓	✓
Digit span					
Direct	6D	8	✓	✓	–
Inverse	6I	8	✓	✓	–
Global vision	7	18	✓	✓	✓
Central	7C	6	✓	–	✓
Mid-peripheral	7M	6	✓	–	✓
Far-peripheral	7F	6	✓	–	✓

The active background interferes with the task and its distracting effect must be 'suppressed' by the subject to correctly identify the target. The target is to be identified within 5 different moving stimuli proposed. In this case, the shape of the stimuli is not the only discriminant factor, but the trajectory must be taken into account: they can move randomly vertical and diagonal. The starting position of the stimuli can more randomly in vertical and diagonal directions of the screen, which is divided into Central, Mid-peripheral and Far-peripheral.

The software records the response time to the targets, the number of errors and omissions (see Table 1). Omissions are grouped according to their starting position: Central (**7C**), Mid-peripheral (**7M**) and Far-peripheral (**7F**). Note that errors occurring without stimuli cannot be counted in a specific area.

3. Validation experiment: results and discussion

3.1. Participants

The number of subjects that completed all the tests in the battery is $N = 100$, divided into two groups: 50 aviation pilots (all males) and 50 untrained healthy people (all males) as a control group. Subject ages were in the range of 20–40 years old. Pilots were selected among subjects that have completed their training and experienced at least 1200 real or simulated hours of flight. Controls were recruited, after a preliminary selection from a wider sample, excluding those who underwent particular trainings (e.g. sports or a professional car driving experience), but with at least an education history that allows them to be enrolled in a pilot training course, i.e. a high school degree. The participants gave written informed consent to use the data collected during the trial for research purposes. Descriptive statistics of age and experience are summarized in Table 2.

3.2. Setting and procedure

The experimental study was conducted in computer rooms, where the battery was administered to several subjects at the same time. All subjects were in good health, rested and comfortably seated in front of a computer monitor at such a distance that the lateral area of the monitor coincided with the peripheral area of their vision. Finally, in order to avoid auditory interference all subjects wore a headset during the whole duration of the test. Before each session, a collective briefing was done to answer the general questions of the test battery.

The statistics presented in the results section were computed with the SPSS 21.

3.3. Trial results

First, we performed a *t*-test comparison to verify the hypothesis that pilots and controls are two independent samples. Results of *t*-test analysis are summarized in Table 3. Negative values of t are in favor of the pilots group. All tests have at least one variable that rejects the hypothesis (i.e. that the two samples come from the same group) with a statistical significance ($p < 0.01$). The best result is achieved from the seventh test (Global vision) as all of its variables well differentiate between the pilots and the controls.

In order to investigate whether experience has some influence on the performance, we also executed a pairwise Pearson correlation analysis using the aviation pilots' data, but none of the variables significantly correlates to experience. This preliminary test confirms that the proposed battery can be a suitable measure for identifying candidate pilots, since it is not influenced by the flight experience.

As further study, the linear discriminant analysis was applied to classify the pilots against the controls. Table 4 reports the classification matrix, showing a very good discrimination between the two groups (92% overall), along with the cross-

Table 2 Descriptive statistics.

Group	Variable	Min	Max	Mean	St. Dev.
Pilots	Age	21	39	31.56	3.32
	Experience	1203	4300	1806.46	626.03
Controls	Age	20	40	29.84	6.88
	Experience	0	0	0.00	0.00

Table 3 *t*-Test comparison for independent samples.

Test	Omissions		Errors		Median times	
	t	*p*	*t*	*p*	*t*	*p*
1A	−0.601	0.549	−0.345	0.730	**−5.284**	< 0.001
1B	−2.181	0.030	−1.083	0.280	**−7.051**	< 0.001
2	−2.157	0.032	−2.041	0.042	**−3.147**	0.002
3A	**−4.995**	< 0.001	**−4.952**	< 0.001	**−3.459**	0.001
3B	**−4.636**	< 0.001	**−4.541**	< 0.001	−1.998	0.047
4	−2.298	0.022	**−3.055**	0.002	−0.208	0.835
5A	−1.994	0.047	**−2.706**	0.007	**2.915**	0.004
5V	−2.260	0.025	−0.603	0.547	−0.836	0.404
6D	**−3.692**	< 0.001	−2.192	0.029	–	–
6I	**−5.544**	< 0.001	−0.465	0.642	–	–
7	**−6.646**	< 0.001	**−3.701**	< 0.001	**−8.202**	< 0.001
7C	**−4.742**	< 0.001	–	–	**−7.688**	< 0.001
7M	**−5.519**	< 0.001	–	–	**−5.360**	< 0.001
7F	**−5.399**	< 0.001	–	–	**−4.686**	< 0.001

Statistical significance (p) is included. Note when $p < 0.01$ the cell is evidenced in bold.

Table 4 Results of the discriminant analysis.

Classification matrix				Cross-validation classification matrix			
	Pilots	Controls	% Correct		Pilots	Controls	% Correct
Pilots	47	3	94	Pilots	43	7	86
Controls	5	45	90	Controls	10	40	80
Total	52	48	92	Total	53	47	83

Cases in row categories are classified into columns.

validation with the leave-one-out procedure, which confirmed the good predictive capability of the tests battery (83% overall).

For further analysis, we applied the forward stepwise procedure to identify the core subset of variables and the discriminant model is reported in Table 5, which also reports the standardized canonical discriminant function coefficients. The subset was obtained after a forward stepwise procedure that used the Wilks' lambda as criterion used for controlling the stepwise entry of variables: at each step, the variable that minimizes the overall Wilks' lambda is entered until no further minimization is possible. The Wilks' lambda criterion is a measure of group discrimination.

From Table 5 we can note that five out of seven variables are measures of errors or omissions (i.e. low), which underline how accuracy of the answers can be more important than a faster reaction time for aviation pilots. A greater importance for the peripheral focus can be derived by the presence of 1B errors instead of 1A. In

Table 5 Standardized canonical discriminant function coefficients.

Variable	Coeff.
1B_ERRORS	0.495
3A_OMISS	0.405
3B_ERRORS	0.445
4_OMISS	−0.424
4_MEDTIME	−0.568
6D_ERRORS	0.359
7C_MEDTIME	0.912

light of this result, divided attention does not seem to be effective to discriminate pilots from common people. Therefore, the final model does not include any variable of the test no. 5.

4. Conclusion

In this paper, we presented a battery of seven tests specifically designed for the assessment of the attention factors of aviation pilots. The battery was implemented in a computer software and validated by an experimental trial with 50 experienced aviation pilots and 50 untrained people as controls. *t*-test analysis showed the potential of all the tests in the battery to distinguish between untrained normal people and experienced aviation pilots, while discriminant analysis confirmed the validity of the battery as an instrument that can effectively measure all the distinctive attention factors that characterize aviation pilots. Finally, the stepwise discriminant analysis identified several attention factors that can be used to assist the selection of candidate aviation pilots and to focus their training.

In practice, the software presented in our paper allows setting a benchmark for one of the distinctive cognitive functions of aviation pilots. Candidate pilots can be assessed against this benchmark and the results, along with other psycho-attitudinal tests, can be used to select the most promising candidate pilots. The results prove that our battery of tests addressed all the distinctive attention factors of the aviation pilots as they are significantly different than controls and they are enough to efficiently discriminate between the two groups. Furthermore, the software derives an attention profile of the candidate that can be used to train those factors that are not fully developed at the level of the aviation pilot. Indeed, the computer software presented here is currently used in the evaluation of candidate pilots in several flight schools where it is integrated with standard selection protocols and it is employed to develop personal training programs.

Further development will focus on the definition of a battery of tests that can be integrated with onboard instrumentation (e.g. with augmented cognition systems like the one proposed by Di Nuovo et al. [6,7]), in order to quickly evaluate the current attention level before the start of a flight duty, and to assure that a sufficient level is attained to fulfill with the highest safety requirements. While future work will study the application of the battery presented here in similar contexts, for instance, air traffic control, in which was also found that attention plays a major role [5].

References

[1] A. Baddeley, Working memory, Science 255 (1992) 556–559.
[2] D. Bartram, Validation of the MICROPAT Battery, Int. J. Select. Assess. 3 (2) (1995) 84–95.
[3] L.C. Boer, Taskomat: evaluation of a computerized test battery, Int. J. Select. Assess. 3 (2) (1995) 105–114.
[4] T.R. Carretta, Pilot candidate selection method, Aviat. Psychol. Appl. Human Fact. 1 (1) (2011) 3–8.
[5] K. Conzelmann, D. Keye, Which aspects of a semistructured interview, besides cognitive ability tests, contribute incrementally to predicting the training success of air traffic controller trainees?, Int J. Select. Assess. 22 (3) (2014) 240–252.
[6] A.G. Di Nuovo, R.B. Cannavo, S. Di Nuovo, An agent-based infrastructure for monitoring aviation pilot's situation awareness, in: IEEE Symposium on Intelligent Agent (IA), IEEE, 2011, pp. 1–7.
[7] A.G. Di Nuovo, R.B. Cannavò, S. Di Nuovo, An intelligent infrastructure for in-flight situation awareness of aviation pilots, in: Foundations of Augmented Cognition. Directing the Future of Adaptive Systems, Springer Berlin Heidelberg, 2011, pp. 598–607.
[8] M.R. Endsley, Theoretical underpinnings of situation awareness: a critical review, Situation Awareness Anal. Meas. (2000) 3–32
[9] M.R. Endsley, D.J. Garland, Pilot situation awareness training in general aviation, in: H. Factors, E. Society (Eds.), Human Factors and Ergonomics Society Annual Meeting Proceedings, Proceedings 2 – Training, 2000, pp. pp. 357–360.
[10] D. Gopher, M. Well, T. Bareket, Transfer of skill from a computer game trainer to flight, Hum. Factors: J. Hum. Factors Ergon. Soc. 36 (1994) 387–405.
[11] R.E. King, T.R. Carretta, P. Retzlaff, E. Barto, M.J. Ree, M.S. Teachout, Standard cognitive psychological tests predict military pilot training outcomes, Aviat. Psychol. Appl. Human Factors 3 (1) (2013) 28.
[12] E.I. Knudsen, Fundamental components of attention, Annu. Rev. Neurosci. 30 (2007) 57–78.
[13] M.W Matlin, Cognition, eighth ed., Wiley, 2012, p. 640.
[14] A.T. Schriver, D.G. Morrow, C.D. Wickens, D.A. Talleur, Expertise differences in attentional strategies related to pilot decision making, Hum. Factors: J. Hum. Factors Ergon. Soc. 50 (6) (2008) 864–878.
[15] H. Strasburger, I. Rentschler, M. Juttner, Peripheral vision and pattern recognition: a review, J. Vision 11 (2011) 13.
[16] K. Sulistyawati, Y.P. Chui, Y.M. Tham, Y.K. Wee, Evaluation of process tracing technique to assess pilot situation awareness in air combat missions, in: Proceedings of the 7th International Conference on Engineering Psychology and Cognitive Ergonomics, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 824–833.
[17] E. Toulouse, H. Pieron, Prueba perceptiva y de atención, TEA Ediciones SA, Madrid, 1986.
[18] C.D. Wickens, J.S. McCarley, A.L. Alexander, L.C. Thomas, M. Ambinder, S. Zheng, Attention-situation awareness (A-SA) model of pilot error, in: David C. Foyle, Becky L. Hooey (Eds.), Human Performance Modeling in Aviation, 2008, pp. 213–239.