Applied Human Factors and Ergonomics International

Methodological Issues of Pilots' Performance Assessment

Igor V. Bukhtiyarov[1], K. Wolfgang Kallus[b], Evgeniya S. Osokina[ac] and Christine M. Talker [b]

[a]Research Institute of Occupational Health
Federal Agency of Scientific Organizations, Moscow, Russia

[b]Department of Psychology
Karl-Franzens-University Graz, Graz, Austria

[c]CBR Laboratory of Cognitive Psychophysiology
National Research University – Higher School of Economics, Moscow, Russia

ABSTRACT

The paper provides a review of present-day studies on the problem of pilots' performance in various flight conditions, with a focus on their methodology. Conceptual frameworks of the studies (concepts of working capacity, functional state and mental workload) are discussed, and different objective and subjective measures and methods used are described. Eye-tracking is regarded with special attention as a promising tool able to examine the internal mechanisms of pilots' performance. The paper hints to the importance of systemic methodological approach to pilots' performance assessment and proposes the direction for further research in the field of aviation psychophysiology.

Keywords: Pilots, Performance, Working capacity, Functional state, Mental workload, Eye-tracking, Psychophysiological measures, Psychological measures, Methodology

INTRODUCTION

Although the character of piloting varies in different aviation environments depending on the type of propulsion, mission goals and tasks, probable stressors and hazards, crew size, technological equipment, flights regularity and duration, and many other conditions, the key principles of pilots' performance are common (Bates et al., 1997). On the whole, performance of a flight, reflects working capacity of a pilot, which represents "a pilot's characteristic, which depends on his / her physiological and psychological functions and determines his / her ability to perform a professional activity with a required quality within a certain period of time" (Kozlov, 2002, p. 60). Working capacity varies in the course of performance and depends on numerous situational and constant factors, such as functional state, mental workload, spatial disorientation (SD), hypoxia, G-forces, circadian rhythms, professional experience, personalaty traits and many others (Kozlov, 2002).

The most influential of these factors are functional states – fatigue, monotony, tension, stress, etc. These states are also referred to in the literature as mental, emotional, psychophysiological or critical psychological states (Bodrov, 2009; Borghini et al., 2012). The attribute "functional" points to the connection of such states with a person's ability to perform certain work or specific goal-oriented activity. Initially this concept has been developed in Russian physiology relative to energetic resources of a working organism. However, since the performance process cannot be separated from a human life in general, functional state includes not only physiological, but also psychophysiological and psychological aspects (Leonova, 1984). It reflects basic characteristics of functioning, on the one hand, and serves for adaptation to the environment, on the other hand (Ilyin, 2005). Being a product of interaction between the organism and the environment, functional state is a dynamic formation, which varies with time-on-task under the influence of different forces (Bodrov, 2009). A decrease in performance effectiveness in the case of negative functional states is a result of depletion of psychophysiological resources (Leonova, 1984).

In sum, working capacity is manifested, first, in performance effectiveness, and second, in functional state parameters – physiological and psychophysiological reactions of an organism and subjective psychological mode, reflecting "performance costs" (Bodrov, 2009). It determines the overall approach to pilot's working capacity and performance assessment methodology.

STATE-OF-THE-ART IN PILOTS' PERFORMANCE RESEARCHES: METHODOLOGICAL ISSUES

Every functional state manifests itself in a specific manner in constitutional, psychological and behavioral levels. Such manifestations can be identified with the help of many objective and subjective measures. The most prominent are the measures of functioning of different organism systems: central nervous system, cardiovascular system, respiratory system, motor system and others. Moreover, there are differences in emotional, motivational and cognitive spheres. In behavior functional states are tied with performance effectiveness, quantity of errors and lapses, intensity and tempo of working process (Leonova, 1984).

Aviation experts (Bodrov, 2009) recommend using the following measures for pilot's functional state assessment:

1. professional (direct) measures:
    a. measures of a flight performance and ground training (indices of effectiveness, reliability, accuracy and speed),
    b. measures of special working tests performance (indices of effectiveness, reliability, accuracy and speed),
2. functional (indirect) measures:
    a. subjective measures (psychological questionnaires and interview data),
    b. objective measures:
        i. physiological and psychophysiological measures (e.g., heart rate (HR), heart rate variability (HRV), electrodermal activity (EDA), skin conductance level (SCL), brain activity, blood pressure, temperature, tremor, tapping speed, flicker fusion frequency, reaction time),
        ii. psychological measures (e.g., attention, memory, reasoning),
        iii. biochemical measures (e.g., red blood cells, phagocytes, catecholamines),
        iv. measures of functional reserves of the organism (e.g., Letunov functional test).

Functional state shifts are accompanied with changes in different measures, which are often multidirectional. Evidently, regarding a single measure can lead to wrong identification of the functional states. Therefore, functional state identification requires a complex systemic approach including different methods. Integral analysis of a number of measures, considering not only their values, but also interconnections between them, dynamics and relation to performance effectiveness, is required (Leonova, 1984).

Interestingly, the same ideas and measures are often used in mental workload studies. In fact, the concept of mental workload is more prevalent in international aviation psychology and psychophysiology, compared with the concept of the functional state. Mental workload, composed of a flight task requirements and conditions – both external (weather, aircraft malfunctions, emergencies, etc.) and internal (pilot's properties), is also regarded as a fundamental factor contributing to pilot's performance (Bates et al., 1997). It is stated that pilot's mental workload assessment, like a functional state assessment, requires comprehensive analysis of physiological, psychological and behavioral data (Hart, Wickens, 1990).

The evident flight performance measures are errors in navigation process, flight parameters maintaining, time measurements, expert ratings, etc. (Borghini et al., 2012; Cheung et al., 2004; Kozlov, 2002). However, identical performance results may be observed in different functional states or different levels of mental workload (Borghini et al., 2012). For this reason the researchers combine them with other psychophysiological and psychological methods.

The most frequently employed physiological methods are electrocardiography (ECG), (EDA recording and electroencephalography (EEG). The measures provided by these methods are sensitive to both functional state and mental workload variations.

ECG measures – HR and HRV – are used in numerous aviation studies, dealing with problems of excessive and insufficient mental workload, stress, emotional arousal, fatigue, SD, training efficacy and many others. It has been shown that these measures tend to increase in situations requiring substantial mental effort and under the influence of stressors of different origin, and decrease with a decline of a level of arousal (Borghini et al., 2012; Causse et al., 2001; Cheung et al., 2004; De Rivecourt et al., 2008; Kallus et al., 2011; Previc et al., 2009; Willson, 2002).

EDA measures are less commonly used in aviation research in comparison with ECG measures. Nevertheless, there is often a correlation between HR and SCL: for instance, both HR and EDA response usually increase during the most demanding stages of a flight (takeoff and landing) and the most difficult flight maneuvers (Willson, 2002).

EEG is not widely distributed in pilot's performance investigations because of relative complexity of necessary experimental setting. Nonetheless, changes of brain activity in different power spectra mirror pilot's mental workload and fatigue. High mental workload during a flight is accompanied with EEG power increase in theta band and decrease in alpha band. It often occurs in cases of increased focused attention or increased time pressure – especially in conjunction with increased task demands, e.g., decision-making conflicts. The transition between mental workload and mental fatigue is characterized by an increase in brain activity in delta and alpha bands and its decrease in beta band (Borghini et al., 2012).

Despite having similar patterns of changes, "each of the psychophysiological measures provides unique information" (Wilson, 2002, p. 16). For example, HR is marked in the literature as the most sensitive measure of mental workload. It usually shows differences between resting state and task execution (De Rivecourt et al., 2008) and well suits for situations with a pronounced affective component (conditions of high responsibility, danger, etc.). HRV gives more precise information than HR in emotionally neutral conditions (Jorna, 1993). Moreover, it clearly reflects mental workload fluctuations at the intermediate level of cognitive effort (De Rivecourt et al., 2008).

Subjective measures in pilot's performance studies are collected either by means of standardized psychological tools or in the course of an interview.

Every functional state is attributed by a number of typical emotions, sensations and feelings. A state of fatigue is linked with tiredness and languor, monotony – with boredom, stress – with numerous negative emotions. Moreover, specific needs are inherent in different states (Leonova, 1984). Diagnostics of affective and motivational spheres can be conducted with a number of international and national psychological questionnaires and projective techniques. Besides, there are several psychological tools measuring workload. They are based on subjective evaluation of difficulties a person is faced with in the course of the performance. The most widely used tools of this category are NASA Task Load Index (NASA-TLX), Subjective Workload Assessment Technique (SWAT), Cooper-Harper rating scale, and bipolar rating techniques. They provide information about an appraisal of mental and physical workload, temporal demands, frustration level, time load, stress load (Bates et al., 1997). It is necessary to mention that shifts of psychophysiological measures (such as HR and EDA parameters) are not necessarily accompanied by an increase in subjective workload rates: these rates may depend on psychological factors such as familiarity with flight maneuvers, while objective reactions of an organism are not affected by them (Wilson, 2002).

Reconstruction interview is also a useful method enabling to obtain subjective data concerning not only the pilot's state during the flight, but also his / her self-ratings of performance (Kallus et al., 2011).

It is also fruitful to include into the studies some special tests, simulating particular performance tasks and examining separate processes engaged: motion, sensorimotor coordination, perception, attention, memory and others (Leonova, 1984).

Psychomotor tests are aimed at assessment of perceptual-motor abilities, which represent coordination between thinking and doing. There are specific batteries of psychomotor test for pilots, including eye-hand-foot coordination tasks, reaction time tasks, divided and selective attention tasks and others. These tests diagnose not only the motor activity, but also functioning of decision making and motion control mechanisms (Bates et al., 1997; Leonova, 1984). They often help to make the distinction between similar functional states. In this way, monotony causes subconscious increase in working tempo and an increase in reaction time in psychomotor tests, while fatigue, on the contrary, causes reaction time decrease. It should be noted that under different conditions some psychomotor tests are more affected by the functional state than others (Ilyin, 2005).

Psychomotor tests are related to low-level information processing (perception and attention). Besides, cognitive working tests, giving measures of focused attention, short-term memory, reasoning and judgment, are used. Psychomotor and cognitive information processing measures, obtained in different tasks, often correlate (Bates et al., 1997).

The above-mentioned methods, applied in a systemic manner, may make a rather diversified appraisal of the pilot's performance and working capacity. However, a serious drawback of such approach is that all objective measures are generalized and, therefore, are not able to shed light on pilot's behavior during the flight and the processes underlying performance.

PRACTICE AND PROSPECTS OF EYE-TRACKING APPLICATION IN PILOTS' PERFORMANCE RESEARCHES

Correct perception and processing of relevant information is a critical factor in a flight performance (Borghini et al., 2012). Since a pilot receives a major portion of information for navigating an aircraft through vision, the analysis of eye movements, regulated by attentional processes and determining visual information selectivity (Krauzlis, 2008), offers great opportunities for interpretation of pilot's behavior and understanding of the inner mechanisms of his / her performance.

Such analysis is available with the help of eye-tracking – an up-to-date psychophysiological technology, detecting point-of-regard by the "corneal-reflection" method and allowing eye movement activity recording and measurement of its basic components (saccades and fixations, forming a gaze, or a visual scanpath), as well as blinks and pupil size (Poole, Ball, 2006).

In recent years a considerable number of papers has been published dedicated to the issue of pilots' eye movements during the flight. In some of them only normal flight conditions are viewed; it is simply assumed that different stages of a flight are characterized by different levels of mental effort (De Rivecourt et al., 2008; Diez et al., 2011; Di Nocera et al., 2007). In the more complicated works aircraft systems failures (Sarter et al., 2007), external threats (Dehais et al., 2010; Thomas, Wickens, 2004), instrumental flight conditions (Causse et al., 2011) or SD conflicts (Cheung et al., 2004) are simulated. Furthermore, in certain studies pilots are sleep-deprived (Previc et al., 2009; Stern et al., 1994) or stressed (Causse et al., 2011).

The aforementioned papers report that, if a pilot is either initially in non-optimal functional state or faces some difficulties during the flight, performance deteriorates and eye-movement activity changes. In most cases fixation parameters alter and / or duration of fixations on different sources of information – so-called "dwell time" – changes. Dwell time is the first eye-tracking measure to change under conditions causing pilot's stress (Causse et al., 2011; Sarter et al., 2007; Thomas, Wickens, 2004), fatigue (Previc et al., 2009), SD (Cheung et al., 2004). It also changes during the performance of difficult maneuvers when cognitive workload increases (De Rivecourt et al., 2008, Diez et al., 2011).

The sources of information used by a pilot (for example, cockpit instruments) may be marked out as separate areas of interest (AOIs). As, according to “eye-mind hypothesis”, a human gaze indicates attention focus displacement (Poole, Ball, 2006, p. 213), dwell time and sequence of fixations on different AOIs show the pilot’s attention distribution during the flight. A lack of attention to task-relevant AOIs may lead to loss of important information and subsequent impairments in decision-making and performance (Chang et al., 2006).

Another eye-tracking measure widely used for the purposes of pilot’s performance and functional state assessment is spontaneous blink frequency, or blink rate (BR). This measure, in fact, does not necessarily require videooculographic eye-tracking devices and can be recorded using electrooculography (EOG) (Borghini et al., 2012; Willson, 2002). BR is known to increase as a function of time-on-task during the flight and is ordinarily enhanced in fatigued and drowsy pilots (Borghini et al., 2012; Stern et al., 1994). With an increase of visual or cognitive demands BR decreases. For example, in instrumental flight rules (IFR) conditions, requiring extremely careful instrument scanning, it is higher than in visual flight rules (VFR) conditions (Willson, 2002).

Pupil size is of particular interest within the context of pilot’s functional state evaluation. This measure is frequently regarded as an index of mental workload and emotional state: it dilates with an increased task difficulty and varies in dependence on emotional activation of arousal. Pilot’s pupils are enlarged in degraded flight conditions (Dehais et al., 2010) and SD circumstances (Cheung et al., 2004). They are also greater in size during the stage of landing, when pilot’s subjective anxiety and cognitive demands are high (Dehais et al., 2010).

In some studies eye-tracking is a single psychophysiological method used. This approach is appropriate when the authors regard only attention allocation strategies (Diez et al., 2011), but it is quite objectionable when pilot’s performance is explained in view of functional state or mental workload. For instance, if researchers initially assume that different stages of a flight are related to different levels of mental workload, and do not make any attempts to prove that experimentally (Di Nocera et al., 2007), their results cause doubts.

Moreover, many researches regarding unusual in-flight situations (challenging unanticipated autoflight-related events (Sarter et al., 2007) or unexpected off-normal events outside the aircraft (Thomas, Wickens, 2004), surprisingly, consider only gaze parameters and indices of flight performance effectiveness, ignoring other popular measures such as, at least, HR, HRV, SCL and psychological. Actually, it is impossible to draw any extensive inferences on the basis of the given simplistic methodology.

CONCLUSIONS

Eye-tracking is a highly valuable tool for aviation researches. It allows analyzing pilot’s performance at all phases of a flight from departure to landing and its dependence on different external conditions and pilot’s functional state.

However, in many cases the conclusions, made in studies applying eye-tracking, are not well-grounded, because the authors neglect other assessment methods and measures. It is necessarily to combine eye-tracking with both direct rating of flight performance effectiveness and psychophysiological and psychological diagnostics, since “eye movement metrics alone struggle to explain operator performance” (Chang et al., 2006, p. 1624).

REFERENCES

Bates, M. J., Colwell, C. D., King, R. E., Siem, F. M., Zelenski, W. E. (1997), “*Pilot Performance Variables*” (No. AL/CF-TR-1997-0059), ARMSTRONG LAB BROOKS AFB TX MANPOWER AND PERSONNEL DIV.

Bodrov, V. A. (2009), “*Professional Fatigue: Fundamental and Applied Problems*”. Moscow: RAS Institute of Psychology. (In Russian)

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F. (2012), “*Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness*”. Neuroscience Biobehavioral Reviews.

Causse, M., Baracat, B., Pastor, J., Dehais, F. (2011), “*Reward and uncertainty favor risky decision-making in pilots: evidence from cardiovascular and oculometric measurements*”. Applied Psychophysiology and Biofeedback, Volume 36, No. 4, pp. 231-242.

Chang, D., Fuchs, S., Milham, L., Bell-Carroll, M., Stanney, K. (2006), *“Exploring eye tracking measures to understand operator performance”*, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Volume 50, No. 16, pp. 1624-1628.

Cheung, B., Hofer, K., Heskin, R., Smith, A. (2004), *“Physiological and behavioral responses to an exposure of pitch illusion in the simulator”*. Aviation, Space, and Environmental Medicine, Volume 75, No. 8, pp. 657-665.

De Rivecourt, M., Kuperus, M. N., Post, W. J., Mulder, L. J. M. (2008), *“Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight”*. Ergonomics, Volume 51, No. 9, pp. 1295-1319.

Dehais, F., Causse, M., Pastor, J. (2010), *“Toward the definition of a pilot’s physiological state vector through oculometry: a preliminary study in real flight conditions”*. In: Proceedings of HCI Aero.

Di Nocera, F., Camilli, M., Terenzi, M. (2007), *“A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload”*. Journal of Cognitive Engineering and Decision Making, Volume 1, No. 3, pp. 271-285.

Diez, M., Boehm-Davis, D. A., Holt, R. W., Pinney, M. E., Hansberger, J. T., Schoppek, W. (2001), *“Tracking pilot interactions with flight management systems through eye movements”*, in: Proceedings of the 11th International Symposium on Aviation Psychology, pp. 1-6.

Hart, S. G., Wickens, C. D. (1990), *“Workload assessment and prediction”*, in: Manprint, Springer Netherlands, pp. 257-296.

Ilyin, E. P. (2005), *“Psychophysiology of Human States”*. Moscow: Piter. (In Russian)

Jorna, P. G. A. M. (1993), *“Heart rate and workload variations in actual and simulated flight”*. Ergonomics, Volume 36, No. 9, pp. 1043-1054.

Kallus, K. W., Tropper, K., Boucsein, W. (2011), *“The importance of motion cues in spatial disorientation training for VFR-pilots”*. The International Journal of Aviation Psychology, Volume 21(2), pp. 135-152.

Kozlov, V. V. (2002), *“Human Factor: History, Theory and Practice”*. Moscow: Poligraf. (In Russian)

Krauzlis, R. J. (2008), *“Eye movements”*, in: Fundamental Neuroscience, L. R. Squire (ed.). Oxford: Elseiver, pp. 775-792.

Leonova, A. B. (1984), *“Psychodiagnostics of Functional States”*. Moscow: Moscow State University Publishing House. (In Russian)

Poole, A., Ball, L. J. (2006), *“Eye tracking in human-computer interaction and usability research: current status and future prospects”*, in: Encyclopedia of Human-computer Interaction, C. Ghaoui (ed.). London: Idea Group Inc., pp. 211-220.

Previc, F. H., Lopez, N., Ercoline, W. R., Daluz, C. M., Workman, A. J., Evans, R. H., Dillon, N. A. (2009), *“The effects of sleep deprivation on flight performance, instrument scanning, and physiological arousal in pilots”*. The International Journal of Aviation Psychology, Volume 19, No. 4, pp. 326-346.

Sarter, N. B., Mumaw, R. J., Wickens, C. D. (2007), *“Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data”*. Human Factors: The Journal of the Human Factors and Ergonomics Society, Volume 49, No. 3, pp. 347-357.

Stern, J. A., Boyer, D., Schroeder, D. J. (1994), *“Blink rate as a measure of fatigue: A review”*. WASHINGTON UNIV ST LOUIS MO.

Thomas, L. C., Wickens, C. D. (2004), *“Eye-tracking and individual differences in off-normal event detection when flying with a synthetic vision system display”*, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Volume 48, No. 1, pp. 223-227

Wilson, G. F. (2002), *“An analysis of mental workload in pilots during flight using multiple psychophysiological measures”*. The International Journal of Aviation Psychology, Volume 12, No. 1, pp. 3-18.