2009

Multidimensional Evaluation of Pilot`S Threat and Error Management Performance During Complex Flight Maneuvers

Ioana Koglbauer

K. Wolfgang Kallus

Reinhard Braunstingl

Wolfram Boucsein

MULTIDIMENSIONAL EVALUATION OF PILOT`S THREAT AND ERROR MANAGEMENT PERFORMANCE DURING COMPLEX FLIGHT MANEUVERS

Ioana Koglbauer and K. Wolfgang Kallus
Department of Psychology, Karl-Franzens University
Graz, Austria

Reinhard Braunstingl
Institute of Mechanics, Technical University
Graz, Austria

Wolfram Boucsein
Institute of Physiological Psychology, Wuppertal University
Wuppertal, Germany

The present research attempts a multidimensional threat and error management performance analysis of pilots flying according to visual flight rules, during the recovery from four unusual aircraft states: extreme pitch, overbanked attitude, full stall and spin. An anticipative training program was developed based on flight mechanical and psychophysiological analysis of an expert's performance. Training took place in a flight simulator and was preceded and followed by check flights with an aerobatic aircraft, a Pitts S-2B, supervised by an expert aerobatic flight instructor. In a between-groups design, a multidimensional assessment was applied, comprising psychophysiological measures of arousal, emotion, subjective, workload and anticipative comparison processes as complementary to technical performance criteria. Our results evidenced benefits of combined theoretical and practical anticipative flight instruction for the threat and error management in complex flight maneuvers.

Incident and accident analyses performed by the AOPA Air Safety Foundation (2003) and by the Boeing Commercial Airplane Group (2004) indicate that loss of aircraft control due to pilot failures in managing complex aircraft states was responsible for a large number of fatalities in general and commercial aviation during the last decades. Causes of complex aircraft states such as unusual attitudes, stalls and spins involve pilot related factors, as well as technical failures and environmental conditions that can not be entirely avoided or controlled. Therefore, it is essential to enhance the pilot's technical and non-technical skills in managing these states within safe psychophysiological boundaries of workload and arousal.

We evaluated the pilot task performance within a model of threat and error management (TEM) described by Helmreich, Klinect and Wilhelm (1999). The model was initially developed to capture specific crew behavior and situational factors during normal flight operations and provide countermeasures at individual and organizational levels. It proposes an observation based evaluation of the TEM process and uses classifiers for pilot's responses to threats and errors such as threat recognition, error avoidance, error detection, additional error and error management, or producing an unsafe condition. According to this model, outcomes of the TEM process are divided into: unsafe action, additional error, or recovery to safe flight.

For analyzing non-technical strategies involved in TEM, we used the theoretical framework of the "anticipation-action-comparison unit" (Kallus, Barbarino & van Damme, 1997) and the concept of the "situation awareness loop" (cf. Kallus & Tropper, 2007). According to the "anticipation-action-comparison unit" concept, goal-relevant future mental pictures are predicted, based on key elements of the situation, previous experience and mental models. Actions are initiated by anticipations of their effects, and comparisons of predicted and actual situational changes close the feedback loop. Situational changes may be pre-classified as match/mismatch evaluations of anticipated and real variations. Anticipatory processes involve different levels of information processing and are manifested on different levels of the central nervous system organization, from complex conscious planning processes to unconscious anticipatory eye-movements (Kallus & Tropper, 2007; Wilson, 2000). Integrated within this model, the concept of situation awareness is referred to as the perception and awareness of key situational elements in a timely manner, the comprehension and determination of their relevance to safety goals and the forecast of their future status (Endsley, 1988). Furthermore, the "situation awareness loop" fosters a controlled sequence of anticipation, perception, comprehension, projection and action in a feedback circuit. Besides situation awareness, anticipation and outcomes of the task management, an operator's workload is another dimension of performance. Workload describes the costs invested by a human operator to achieve a particular level of performance (Hart & Staveland, 1988). The workload or the amount of resources invested in the TEM process is seen as mediator among performance, task difficulty and the operator's skill (Wickens, 2001). Situation awareness, anticipation and workload related arousal can be inferred from physiological measures and subjective ratings (Boucsein 2007; Boucsein & Backs, 2000; Boucsein, Koglbauer, Braunstingl & Kallus; 2009; Kallus & Tropper, 2007; Wilson, 2000 and 2002). In the present research, we attempt to discriminate arousal types using the integrative neurophysiological model of different kinds of arousal provided by Boucsein and Backs (2009), which describes the structural and functional hierarchy within the brain, being involved in general activation, perception, information processing and response preparation. In a previous study we used this four-arousal model in evaluating the adaptive psychophysiological arousal and emotion regulation of an expert pilot during different cognitive, emotional and physical demands of real flight tasks (Boucsein et al., 2009). The present study aimed at the specificity of different types of arousal for anticipatory processes, task performance and post-task echoing involved in the pilot's TEM performance.

Methods

In-flight technical and psychophysiological aspects of the pilot's TEM performance were evaluated by means of an experimental design with two groups: a training and a control group, compared at two times of measurement: before and after simulator training. Statistical analysis of the anticipative simulator training influences was processed using a univariate general linear model. Since there were considerable differences between the training and control group with respects to flight performance and psychophysiological state during the initial flight, and the assumption of homogeneity of regression slopes assumption had not to be rejected, the initial values were entered as a covariate.

Participants and procedure

Twenty-eight male pilots, aged between 21 and 64 years ($M = 38.04$, $SD = 11.45$) volunteered for the study. The participants were randomly assigned to an experimental group ($N = 16$) and a control group ($N = 12$). All pilots held an actual PPL VFR license (private pilot license for flight according to visual flight rules), with their total flight experience ranging from

40 to 430 hours. None of the pilots had aerobatic nor instrument flight rules (IFR) experience. The two groups did not differ significantly in their total flight experience.

The flight task consisted of four flight maneuvers: extreme pitch, overbanked attitude, power-off full stall and left spin with two rotations. Applying the methodological approach for evaluating anticipative processes carried out in our expert case study (Boucsein et al., 2009), each maneuver was split into four phases. An anticipation phase of 15 sec mental preparation preceded each maneuver, followed by the onset, recovery and post-recovery phases. All pilots received theoretical instruction and ground briefing regarding the nature of the maneuvers, their possible eliciting conditions, and avoiding and recovery procedures. Afterwards, each group attended an initial flight, followed by simulator training, simulator test and post-training flight tests. The simulator session of the training group consisted of specific recovery exercises, while pilots of the control group received VFR terrestrial and radio navigation training with similar degree of difficulty. The real flight sessions, supervised by an expert flight instructor (the third author), were performed in a two-place tandem Pitts Special S-2B, a light aircraft certified in the aerobatic category. Training took place in a fixed-base, two-seater generic light aircraft simulator with the following psychological fidelity features: wide screen projection by means of a three-channel visual system to facilitate the simulation of peripheral vision, rudder pressure simulation and an aerodynamic model including stall/spin and unusual attitudes behavior of the aircraft. Generic maneuver representation and response behavior of the simulator were validated at the Institute of Mechanics, Graz University of Technology, with flight maneuver data recorded during a real flight with the Pitts S-2B, using a body fixed coordinate system by means of an inertial platform, which included an aviation-certified laser compass, a MEMS gyro, a GPS sensor and acceleration sensors for each of the three axes (Boucsein et al., 2009). Due to space restriction, only the initial flight and second trial of the final test flight are analyzed in the present paper.

Dependent measures

The TEM performance was evaluated by instructor ratings, ranging from 1 (not acceptable) to 4 (very good). The criterion for a non-acceptable performance was the pilot's failure to respond, manifested as safety-critical omission, wrong prioritization of action or major unsafe acts. Acceptable performance criteria were the presence of major errors that exacerbated the threatening potential of the flight situation or complicated the recovery process, but which were finally mastered by the pilot. Good TEM performance standards included pilot actions that slightly differed in timing and precision from the correct performance, which in turn was rated as very good. The recovery duration was also measured as an additional performance parameter. Self-ratings of the pilot were collected for performance, effort, frustration and task load using the NASA-TLX (Hart & Staveland, 1988). During the entire experiment the pilot's electrodermal activity (EDA) was recorded with the Varioport system (Becker Meditec, Karlsruhe, 2005) as skin conductance from the medial sites of the left foot, adjacent to the plantar area (Boucsein, 1992, Fig. 28). For EDA, skin conductance level (SCL), non-specific skin conductance reactions frequency (NS.SCR freq.) and mean amplitude of skin conductance reactions (SCR amp.) were evaluated using the EDA-Vario software (Version 1.8; Schaefer, 2007). The above mentioned phases of task management were marked with a trigger which was set manually by the flight instructor. The first ten seconds sequence between the 2nd to 11th second of each phase interval was used for psychophysiological evaluation.

Results

TEM Performance and Subjective Ratings

The analysis of instructor ratings yielded that pilots of the training group showed higher overall performance (M = 3.21, SD = .07) than the control group (M = 2.65, SD = .08). The group difference reached significance [F (1, 3.00) = 38.623, p < .05]. The maneuver effects or group by maneuver interactions were not significant. Distinct maneuver analyses indicated a significantly superior performance of the training group in recovering from the extreme pitch attitude as compared to the control group [F (1, 25) = 7.019, p < .05]. Similar results were found for the overbanked attitude [F (1, 25) = 4.304, p < .05] and for the power-off full stall [F (1, 25) = 17.019, p < .001]. The training group showed better performance during the spin recovery (M = 3.05, SD = .17) than the control group (M = 2.59, SD = .20), but the differences were not significant. The recovery duration was slightly reduced from the initial covariate adjusted mean (M = 17.78 sec), not only in the training group (M = 14.22 sec, SD = .44) but also in the control group (M = 14.75 sec, SD = .50). However, the group differences were not significant.

Self-ratings of performance as assessed by the NASA-TLX were significantly better in the training group [F (1, 2.848) = 16.724, p <0.05], while the effort ratings were significantly lower compared to the control group [F (1, 4.899) = 6.756, p < 0.05]. The anticipative simulator training seemed to have a moderate impact on the perceived mental demand, with lower scores in the training group than in the control group, reaching just marginal significance [F (1, 3.714) = 6.546, p = 0.06]. Subjectively experienced physical and temporal demands of the flight task were significantly lower in the training group than in the control group [F (1, 4.066) = 17.604, p < 0.5, and F (1, 4.178) = 63.107, p < 0.001, respectively]. Subjective ratings of frustration did not vary significantly between the groups. Self-ratings of the current physical state and ratings of the psychological state before and after the flight did not show significant differences between the groups.

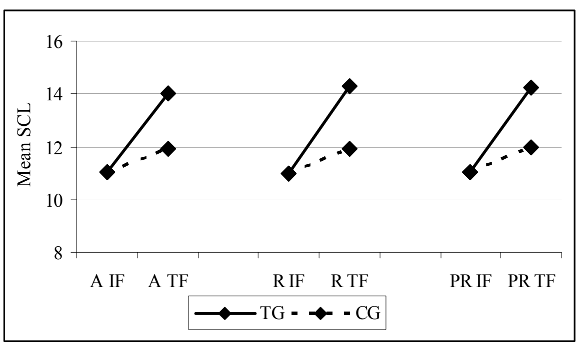

Figure 1. Covariate adjusted means of the Mean SCL (µS) of the training and control group during the initial (IF) and final (TF) test flight. (A= anticipation phase, R= recovery phase, PR= post-recovery phase).

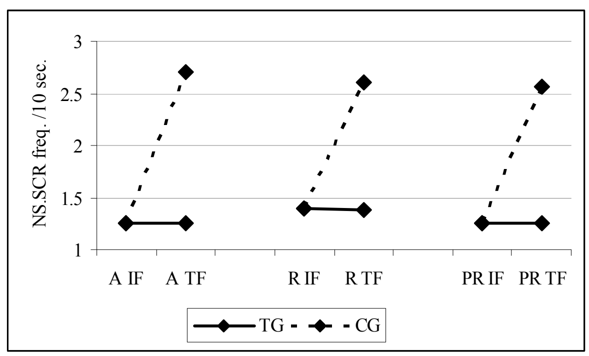

Figure 2. Covariate adjusted means of the NS.SCR freq. of the training and control group during the initial (IF) and final (TF) test flight. (A= anticipation phase, R= recovery phase, PR= post-recovery phase).

Results of electrodermal activity

The simulator training condition seemed to have significantly influenced the mean SCL during the anticipation phase, the training group showing significantly higher values than the control group [$F(1, 14.98) = 28.207$, $p < .001$]. As depicted in Figure 1, similar results appeared during the recovery and post-recovery phases [$F(1, 14.14) = 23.214$, $p = .001$, and $F(1, 22.54) = 24.142$, $p < .001$, respectively]. The mean SCL seemed to also to be influenced by the type of recovery maneuver during the anticipation [$F(3, 3.08) = 50.718$, $p = .005$], recovery [$F(3, 2.99) = 17.646$, $p = .05$] and post-recovery phase [$F(3, 2.98) = 26.984$, $p = .05$]. Group by maneuver interactions were not significant. The mean SCR amp was not significantly influenced by the simulator training condition, nor by maneuver or interactions between the two factors. The NS.SCR freq. (Figure 2) was higher in the control group than in the training group during the anticipation [$F(1, 3.04) = 10.740$, $p < .05$], recovery [$F(1, 2.99) = 47.010$, $p < .05$] and post-recovery phases [$F(1, 3.02) = 30.716$, $p < .05$]. No significant effects were found for the maneuvers and the interaction between experimental groups and maneuvers.

Discussion

Our interest was to determine multidimensional changes of performance under the influence of a specific recovery training in the simulator. The analysis of instructor ratings indicates that the anticipative training performed in a simulator with sufficient psychological fidelity significantly improves the pilot's flight performance in recovering from unusual attitudes, stalls and spins. In general, pilots of the training group improve their TEM performance quality, reaching a level between good and very good, which means that they successfully manage the maneuver threats, only slightly deviating in timing and precision from the correct performance. Pilots of the control group reach a mean performance between the levels of acceptable and good, meaning that their TEM performance generally involves minor and major errors that are, however, finally mastered by the pilots in the given situation. Self-ratings of performance follow the same trend, since pilots in the training group score their own performance significantly higher than pilots in the control group. These qualitative changes in performance are not paralleled by the duration of recovery. Furthermore, the associated workload, in terms of costs of performance, is significantly lower in the training group, since pilots who benefit of the anticipative training report significantly lower effort required by the recoveries than those who could not benefit. Although both groups have mean recovery durations of about 14 sec during the final test flight, subjective evaluations of temporal demand are significantly lower in the training group. Pilots of the training group evaluate the TEM tasks as less physically demanding than pilots of the control group, while the differences in evaluation of mental demand reach just marginal significance.

Autonomic nervous system (ANS) activity as reflected by EDA parameters will be interpreted within the framework of the four-arousal model provided by Boucsein and Backs (2009). The significant increase of mean SCL during all maneuver phases in the training group compared to the control group reflect an increase of general arousal, together with the cortical activation of the motor plans and conditioned behavior patterns permitting timely responses to the anticipated events. In turn, the lower NS.SCR freq. in the training group indicates a significantly decrease of negatively tuned affective responses during all phases and maneuvers. In contrast, the higher NS.SCR freq. in the control group reflects an activation of the affect arousal system, which is responsible for the elicitation of immediate responses such as flight/flight or freezing reactions. These subtle changes in ANS activity are not reflected in the subjective measures of psychological and physical state of the groups. Hence, recording of EDA in flight provides valuable information about the pilot's TEM, which complements performance and subjective ratings.

In conclusion, anticipative flight instruction involving hands-on simulator exercises and recovery procedures split into distinct anticipation-action-comparison units (Kallus et al., 1997) were demonstrated to improve the pilot's TEM performance capability as well as their neurophysiological adaptability to demanding maneuvers like unusual attitudes, full stalls and spins during real flight.

References

AOPA Air safety Foundation (2003). "Stall/spin: entry point for crash and burn?". In *Air Safety Foundation Special Reports*, Maryland.

Boeing Commercial Airplane Group (2004). Statistical summary of commercial jet airplane accidents, Worldwide Operations, 1994-2003. Airplane Safety Engineering, Seattle, Washington.

Boucsein, W. (1992). Electrodermal activity. New York: Plenum Press.

Boucsein, W. (2007). Psychophysiologie im Flug – eine Fallstudie. In P. G. Richter, R. Rau, & S. Mühlpfordt (Eds.), *Arbeit und Gesundheit. Zum aktuellen Stand in einem Forschungs- und Praxisfeld* (pp. 130-142). Lengerich: Pabst Science Publishers.

Boucsein, W., & Backs, R. W. (2000). Engineering psychophysiology as a discipline: historical and theoretical aspects. In: R.W. Backs, & W. Boucsein (Eds.). *Engineering Psychophysiology. Issues and Applications* (pp. 3-30). Mahwah, NJ: Lawrence Erlbaum Associates.

Boucsein, W., & Backs, R. W. (2009). Methods and models for the psychophysiology of emotion, arousal, and personality. In V. G. Duffy (Ed.), *Handbook of Digital Human Modeling* (pp. 35-1 – 35-18). Boca Raton, FL: CRC Press/Taylor & Francis.

Boucsein, W., Koglbauer, I., Braunstingl, R., & Kallus, K. W. (2009) The use of psychophysiological measures during complex flight maneuvers - an expert pilot study. In J. Westerink et al. (Eds.), *Probing Experience II*, Dordrecht: Springer (in preparation).

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society* 32nd Annual Meeting (Vol. 1, pp. 97-101). Santa Monica, CA: Human Factors Society.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.) *Human Mental Workload* (pp. 239-250). Amsterdam: North Holland Press.

Helmreich, R. L., Klinect, J. R., & Wilhelm, J. A. (1999). Models of threat, error, and CRM in flight operations. In *Proceedings of the Tenth International Symposium on Aviation Psychology* (pp. 667-682). Columbus, OH: The Ohio State University.

Kallus, K. W., Barbarino, M., & van Damme, D. (1997). Model of the cognitive aspects of air traffic control (HUM.ET1.STO1.1000-REP-2). Brussels: Eurocontrol.

Kallus, K. W., & Tropper, K. (2007). Anticipatory processes in critical flight situations. *Proceedings of the 14th International Symposium on Aviation Psychology*. Dayton, Ohio, April 23-26, 2007.

Wickens, C. D. (2001). Workload and situation awareness. In P. A. Hanckock, & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue* (pp. 443-450). Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, G. F. (2000). Strategies for psychophysiological assessment of situation awareness. In M. R. Endsley, & D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 175-187). Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *International Journal of Aviation Psychology*, 12 (1), 3-18.