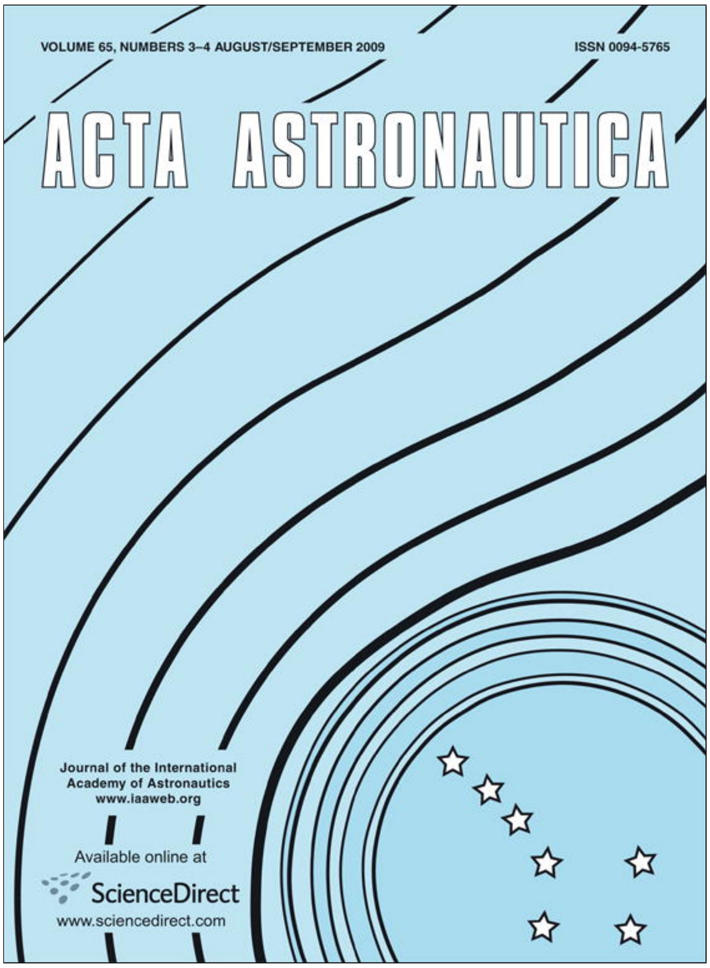
VOLUME 65, NUMBERS 3–4 AUGUST/SEPTEMBER 2009
ISSN 0094-5765
ACTA ASTRONAUTICA
Journal of the International Academy of Astronautics
www.iaaweb.org
Available online at
ScienceDirect
www.sciencedirect.com

PERGAMON

Available online at www.sciencedirect.com
ScienceDirect

Acta Astronautica 65 (2009) 325–329

ACTA ASTRONAUTICA

www.elsevier.com/locate/actaastro

Crew performance monitoring: Putting some feeling into it

N. Pattyn[a, b, c, *], P.-F. Migeotte[d], J. Morais[c], E. Soetens[a], R. Cluydts[a], R. Kolinsky[c]

[a] *Department of Cognitive and Biological Psychology, Vrije Universiteit Brussel, Belgium*
[b] *Department of Behavioral Sciences, Royal Military Academy, Brussels, Belgium*
[c] *Unité de Recherches en Neurosciences Cognitives, Université Libre de Bruxelles (U.L.B.), Belgium*
[d] *Laboratoire de Physique Biomédicale, U.L.B., Belgium*

Received 5 December 2007; accepted 21 January 2009
Available online 7 May 2009

Abstract

Two hypotheses have been invoked so far to explain performance decrements in space: the microgravity hypothesis and the multiple stressors hypothesis. Furthermore, previous investigations of cognitive performance did not specifically target executive functions. The aim of this study was to investigate the impact of operational stress on cognitive control, towards both neutral and emotionally loaded material, using both psychometric and physiological indicators (autonomic nervous system activity computed through cardio-respiratory recordings). We applied the same design in a study on student pilots ($N = 12$) in baseline conditions and right before a major evaluation flight and on astronauts ($N = 3$) before, during and after a short-duration spaceflight. To address the problem of scarcity of subjects, we applied analytical methods derived from neuropsychology: comparing each astronaut treated as a single subject to a group of carefully matched controls ($N = 13$). Results from both student pilots and astronauts showed that operational stress resulted in failing cognitive control, especially on emotionally loaded material that was relevant to the subjects' current concern. This impaired cognitive control was associated with a decreased physiological reactivity during mental tasks. Furthermore, for astronauts, this performance decrement appeared on the last data-collection before launch and lasted for the two in-flight measurements. These results thus allow us to conclude that: (i) performance testing including an emotional dimension seems more sensitive to operational stress, (ii) decreased heart rate reactivity was associated with impaired cognitive control and (iii) microgravity is not the sole causal factor of potential performance decrements in space, which are more likely due to the combination of multiple stressors.

1. Introduction

Considering the potential cost of human error in operational settings, it is undisputable that measuring cognitive performance is a relevant challenge. The need for a robust remote assessment method for cognitive performance has been specifically ranked as a research priority by several space agencies [1,2] for the preparation of exploration missions. As indicated in the previous reviews [3], the experimental evidence from cognitive performance during spaceflight reveals a lack of investigation on executive functions. However, both anecdotal reports from astronauts [4,5] and analysis from real-life crew errors during missions [6] indicate a potential involvement of cognitive control in the required performance of astronauts, as well as a probable failure of this dimension as being one of the major sources of human errors in-flight. Two hypotheses have been invoked so far to explain performance

---

[*] Corresponding author at: Department of Cognitive and Biological Psychology, Vrije Universiteit Brussel, Belgium. Tel.: +32 2 629 15 94; fax: +32 2 629 24 89.
E-mail address: npattyn@vub.ac.be (N. Pattyn).

0094-5765/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.actaastro.2009.01.063

decrements in space: the microgravity hypothesis and the multiple stressors hypothesis [7]. However, the range of assessment methods for cognitive performance in space has been somehow restricted so far [2]. A full description of cognitive performance requires the addition of subjective and psychophysiological aspects to the traditionally measured behavioural aspects. The present study aimed at performance investigation from this integrative point of view. We validated the method in a known stressful situation, and applied the method during a short-duration spaceflight.

The validation was performed on military student pilots (SPs) by comparing baseline recordings to results before a major evaluation flight. In such a situation, real-life performance is never free of emotional appraisal, which is also true during spaceflight. We thus included an emotional dimension in the testing. We targeted cognitive control through Stroop-like interference paradigms [8]. Since real-life performance is never free of emotional and/or motivational appraisal, emotional material was included to improve the ecological validity. This procedure of "emotional Stroop" has previously shown to be sensitive to real-life stress [9]. In addition, autonomic responses to the presentation of cognitive tasks were assessed through cardio-respiratory measurements. Systemic physiological responses, as indicators of mental workload and stress, were assessed through heart rate, heart rate variability and respiratory responses. To address the problem of scarcity of subjects in space research, we applied an analytical method derived from neuropsychology: comparing a single subject (the astronaut, former jet fighter pilot) to a group of carefully matched controls (jet fighter pilots) with adapted statistical testing [10].

2. Method

Subjects: SPs ($N = 12$) and jet fighter pilots ($N = 13$) from the Belgian Air Force. Three astronaut subjects before, during and after a short-duration (11 days) spaceflight.

Apparatus: The LifeShirt (VivoMetrics, Inc.) was used to record ECG and respiration for SPs. RR-interval (RRI), respiratory frequency (F_resp), tidal volume (TV), inspiratory time over total breath cycle time (Ti/Ttot) and respiratory sinus arrhythmia (RSA), calculated through the peak–valley method, were computed. ECG and respiration timings for the astronaut subjects were recorded through a modified version of the Portapres (TNO Biomedical Instruments), RRI and RSA were computed through in-house developed Matlab algorithms. Cognitive testing was presented on an IBM A21p laptop and subjects responded by pressing appropriately labelled keys. Reaction times (RT) and error rates (ER) were recorded.

Procedure: All sessions began with a 5 min rest recording. Cognitive testing included a colour-word, emotional and numerical Stroop task. For SPs, there were three categories of emotional words: general (e.g. "death"), pilot-specific (e.g. "crash") and SP-specific (e.g. "evaluation"). SPs were recorded during two sessions: a baseline session and right before their first major evaluation flight. A detailed description of the testing methodology has been given elsewhere [11]. For the astronaut subjects, there were two categories of emotional words: general (e.g. "death") and mission-specific (e.g. "depressurisation"). RT and ER were measured. Data were recorded 44 and 9 days before launch (L-44 & L-9), on 5th and 8th inflight days (FD5 and FD8) and after the return (R) on days 4 and 25 (R + 4 and R + 25). The control group performed one session, equivalent to L-44, excepted for one "best-match" control subject, who performed the tests in a similar time course as the astronaut subject. This measure was included to assess the repetition effect of the protocol.

3. Results

SPs: Only the most significant variations between baseline and pre-test are reported here. Fig. 1 shows the modulation of ER for neutral and general, pilot-specific and SP-specific emotional material according to recording session. ER increases with increasing specificity of

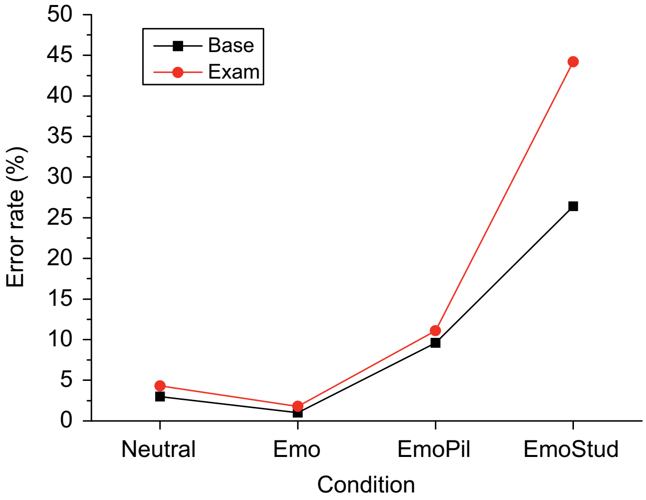

Fig. 1. SPs: Error rates for the different conditions of the emotional Stroop task (neutral, general emotional, pilot-specific emotional and student-pilot-specific emotional words) during the baseline and the pre-exam recording.

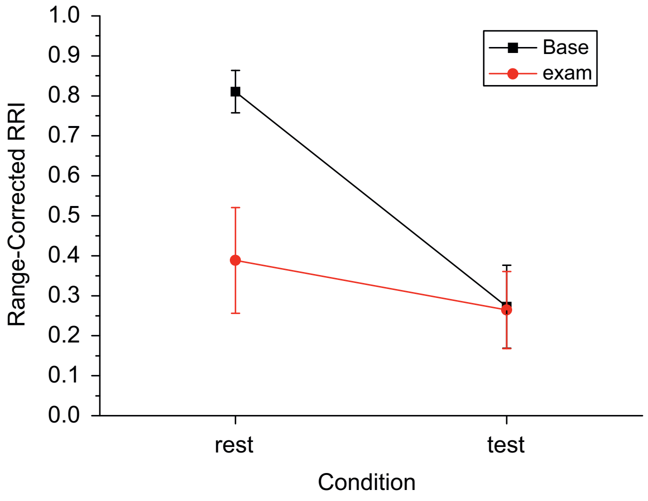

Fig. 2. SPs: Reactivity (range-corrected RRI variation between rest and test recordings) for baseline and exam session.

the emotional material and is modulated by the session. The ANOVA revealed a significant effect of session [$F(1, 11) = 7.18$; $p = 0.021$], emotion [$F(3, 33) = 55.61$; $p < 0.001$], as well as a significant interaction [$F(3, 33) = 3.78$; $p = 0.02$]. The only significant contrast between both sessions was for SP-specific emotional stimuli [$F(1, 11) = 13.82$; $p = 0.003$]. Physiological results have been reported in detail elsewhere [12] and showed an effect of stress on rest recordings, as a decreased RRI ($M_{\text{base}} = 851$ ms; $M_{\text{pre-test}} = 772$ ms) and RSA ($M_{\text{base}} = 189$ ms; $M_{\text{Pre-test}} = 133$ ms), which was significant for both RRI [$F(1, 11) = 11.17$; $p = 0.001$] and RSA [$F(1, 11) = 10.35$; $p = 0.002$]. Furthermore, the initial reactivity between rest recording and the first cognitive test, being the mixed colour-word and emotional Stroop test, appeared to be different between baseline and pre-test session for RRI. We applied the range-correction procedure as within-subject standardization to these RRI values, which yielded the results shown in Fig. 2. ANOVA showed a significant effect of test [$F(1, 11) = 22.38$; $p = 0.001$], no effect of session [$F(1, 11) = 2.29$; $p > 0.1$] and a significant interaction between session and test [$F(1, 11) = 10.62$; $p = 0.01$].

Spaceflight: The detailed physiological results will not be reported here. For the behavioural results, the very high inter-individual disparity of results across the three subjects confirmed the adequacy for a single subject analysis. Due to lack of space, we will not report all the behavioural results here (evolution of RT and analysis compared to controls and evolution of colour and emotional interference effects across data-collection), but we will focus on the most significant variation, which was shared by the three subjects: the increase of ER in-flight (already showing on the last measurement before launch), especially on emotionally loaded material.

The summary of ER for all conditions of the Stroop task is presented in Table 1.

During the first data collection of the astronauts, the results seem to show a floor effect for ER for Subject 1: his performance is nearly perfect. This is also the case for R + 4 and R + 25. For Subject 2 on the other hand, the first data-collection showed very slow responses and a very high ER, as illustrated in Table 1. Table 1 presented a very large inter-individual differences in the quality of performance from the three astronaut subjects: results from Subject 2 suggest L-44 was not an appropriate baseline measurement, Subject 3 shows an overall high ER and Subject 1 shows an exceptionally good performance. However, when considering intra-individual variations across subjects, all astronauts made more mistakes in-flight, whereas Table 1 presents the measurements where they displayed the fastest RT. Again, the different data-collection points were compared with the control group and the statistical significance was tested according to the adapted method [10]. Where the overall ER differed between controls and astronauts, these differences were investigated by additional testing per condition. It is noteworthy that both control conditions, neutral words and non-words, seem to elicit high ER for the control group as well as for the astronauts, whereas the results for the incongruent condition show a low ER for the control group and not a single error throughout all recordings for Subject 1. When comparing his data to the controls, the overall ER is significantly higher at L-9, FD5 and FD8. These higher ER are mainly due to the peaking errors in response to emotional stimuli: ER for general emotional words at L-9 differ significantly from controls, which is also evidenced at FD5 for both general and specific emotional words, and only for specific emotional words at FD8. Furthermore, this subject also showed a higher ER for simple negative priming at L-9 and FD8. As mentioned before, the L-44 recording suggests an inadequate performance measurement for Subject 2, because of both the very high RTs and ER. Apart from this erratic data collection, the evolution exhibited by Subject 2 shows similarities with Subject 1: peaking ER for in-flight recordings, partly due to increased ER for specific emotional stimuli. Results for Subject 3 again show that in-flight recordings are those eliciting the highest ER. ER for general emotional material peaks at the last measurement before launch, and the first one after the return, while ER for specific emotional material peak for the in-flight recordings. These higher ER are mainly due to

Table 1
Average ER (%) for all the conditions of the Stroop task for the jet fighter pilots control group ($N = 13$) and each data-collection point for the astronauts.

	GenEmo	MisEmo	Neutral	Cong	Incong	InNegP	NegP	XXX	Overall
Controls	**0.75**	**1.00**	**2.33**	**0.25**	**0.50**	**0.25**	**0.50**	**2.33**	**0.99**
Subject 1									
L-44	0.00	0.00	0.00*	0.00	0.00	0.00	0.00	2.00	0.25
L-9	15.00*	0.00	0.00*	0.00	0.00	0.00	3.00*	2.00	2.50*
FD5	15.00*	15.00*	2.00	0.00	0.00	0.00	0.00	2.00	4.25*
FD8	0.00	15.00*	0.00*	0.00	0.00	0.00	3.00*	2.00	2.50*
R + 4	0.00	0.00	0.00*	0.00	0.00	0.00	0.00	2.00	0.25
R + 25	0.00	0.00	0.00*	0.00	0.00	0.00	0.00	2.00	0.25
Subject 2									
L-44	0.00	5.00*	7.00*	2.00*	2.00	0.00	7.00*	11.00*	4.25*
L-9	10.00	0.00	0.00	0.00	1.00	3.00*	0.00	0.00	1.75
FD5	0.00	10.00*	0.00	2.00*	3.00*	7.00*	3.00*	6.00*	3.88*
FD8	0.00	15.00*	6.00*	7.00*	4.00*	3.00*	3.00*	5.00*	5.38*
R + 4	0.00	0.00	0.00	2.00	1.00	0.00	0.00	2.00	0.63
R + 25	0.00	0.00	0.00	0.00	4.00	0.00	0.00	2.00	0.75
Subject 3									
L-44	0.00	0.00	6.00*	3.00*	5.00*	7.00*	4.00*	3.00	3.50*
L-9	10.00*	0.00	0.00	3.00*	7.00*	7.00*	6.00*	6.00*	4.88*
FD5	0.00	10.00*	6.00*	5.00*	7.00*	15.00*	7.00*	10.00*	7.50*
FD8	0.00	10.00*	3.00	7.00*	11.00*	17.00*	13.00*	15.00*	9.50*
R + 4	10.00*	0.00	6.00*	5.00*	4.00*	0.00	6.00*	10.00*	5.13*
R + 25	0.00	0.00	0.00	7.00*	4.00*	7.00*	4.00*	6.00*	3.50*

GenEmo, general emotional words; MisEmo, mission-related emotional words; Neutral, neutral words; Congruent, congruent colour words; Incongruent, incongruent colour words; InNegP, inverse negative priming; NegP, simple negative priming; XXX, non-words. Significant variations (0.01 level) of the astronauts when compared to controls (printed in bold) are marked (*).

emotional stimuli, as shown on Fig. 3. In addition, in-flight data showed a decreased physiological variability and decreased RT variability. The potential link between these findings was further examined by computing the variability range for each parameter as a proportion of the variability range on the first measurement (L-44). An overall variability score for physiological data was computed by averaging the ranges for RRI and RSA, and for psychological data by averaging the RT ranges. These aggregated variability scores, along with overall ER, and ER for the emotional Stroop are depicted in Fig. 3. The increased ER occurred simultaneously with markedly reduced variance in both psychological and physiological measures. This is confirmed by a significant negative correlation ($\rho = -0.77$; $p < 0.05$) between ER for emotional material and physiological variability.

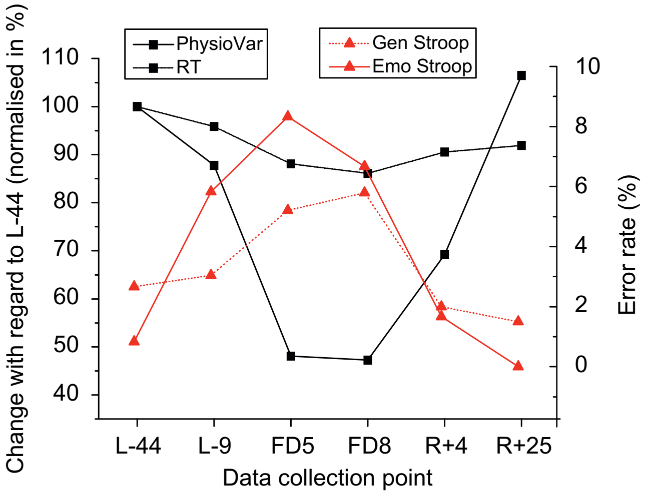

Fig. 3. Astronauts: Left *Y*-axis: change from the value at L-44 (%) for the aggregated "Physio" variability score (full black line, squares) and the overall RT (dotted black line, squares), for each data-collection point. Right *Y*-axis: Stroop error percentage, averaged over all conditions of the Stroop task (Gen Stroop, dotted red line, triangles) and for emotional words only (Stroop Emo, full red line, triangles), for each data-collection point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Specificity of the presented emotional material, or, in other words, its relevance to current concern was a key feature in the interference effect observed for ER in the emotional Stroop task. This was true for both

SPs under stress and for the astronaut subjects during spaceflight and on the last measurement before launch. These results suggest a failure of cognitive control related to relevant emotional material under the influence of stress and thus shows that emotional influence cannot be neglected when quantifying cognitive performance in stressful situations. This confirms the added value of including an emotional dimension to operational performance testing, as suggested by the previous research [12,13]. For SPs, physiological recordings during baseline and stress sessions showed two different effects. First, the rest values differ, with decreased RRI and RSA under stress. Second, the reactivity in RRI due to cognitive testing is decreased in the stress condition. According to these results, stress does not only affect rest levels, but also the dynamic range of heart period reactivity. Reactivity to mental task can be conceived as a phasic response, and the effect of pre-exam stress can be seen as a tonic response. Furthermore, the RRI values show an interaction between these phasic (=arousal) and tonic (=activation) responses. This interaction and the occurrence of a performance decrement under stress, indexed as increased ER indicating failing cognitive control, are findings which all fit with the model of information processing and stress described by Sanders [14]. This confirms the added value of the applied integrative approach to performance testing in operational conditions. Results form the astronaut subject show that the microgravity hypothesis does not seem sufficient to explain our results, since some of the cognitive effects already loom at L-9. This confirms previous reports of performance decrements on the last data-collection before launch, thereby discarding microgravity as the sole causal agent. The observed relationship between the performance decrement, indexed as increased ER on emotional material, and a decreased physiological variability is consistent with the results from SPs. To qualify this altered physiological variability in microgravity, we refer to the concept of “functional reserve” [13], which is decreased in space, causing a lowered reactivity of the autonomic system. The functional reserve of the organism is decreased, implying the usual range of variability is inaccessible, thus causing an impaired functioning. The necessity of variability and flexibility for optimal functioning has already been acknowledged. However, it is the first time this altered variability during space flight is evidenced as associated to performance decrements, and thus as their potential source.

These results thus allow to conclude that: (i) even in situations with small number of subjects, robust analytical methods are applicable, (ii) performance testing including an emotional dimension seems more sensitive to operational stress, (iii) decreased heart rate reactivity was associated with impaired cognitive control in situation of operational stress and (iv) microgravity is not the sole causal factor of potential performance decrements in space, which are more likely due to the combination of multiple stressors.

Acknowledgements

This research was supported by a Prodex Grant no. 90030 by the Euro Space Society and by a grant from the Belgian Department of Defence.

References

[1] BPCR: Bioastronautics Critical Path Roadmap, National Aeronautics and Space Administration, 2004.

[2] HUMEX: Study on the Survivability and Adaptation of Humans to Long-duration Exploratory Missions, European Space Agency, 2000.

[3] J.G. Casler, J.R. Cook, Cognitive performance in space and analogous environments, International Journal of Cognitive Ergonomics 3 (1999) 351–372.

[4] B.J. Bluth, Pilots of Outer Space, Society 21 (1984) 31–36.

[5] P.W. Burgess, Real-world multitasking from a cognitive neuroscience perspective, in: Attention and Performance, vol. XVIII, 2000.

[6] A.P. Nechayev, V.I. Miyasnikov, S.I. Stepanova, O.P. Kozerovko, The aspects of psychophysiological analysis of erroneous actions performed by astronauts, Aviakosmicheskaya i Ecologicheskaya Meditsina 32 (1998) 11–18.

[7] D. Manzey, Monitoring of mental performance during spaceflight, Aviation Space and Environmental Medicine 71 (2000) A69–A75.

[8] C. Ray, Examination stress and performance on a color–word interference test, Perceptual and Motor Skills 49 (1979) 400–402.

[9] J.R. Crawford, D.C. Howell, P.H. Garthwaite, Payne and Jones revisited: estimating the abnormality of test score differences using a modified paired samples *t* test, Journal of Clinical and Experimental Neuropsychology 20 (6) (1998) 898–905.

[10] N. Pattyn, P.F. Migeotte, R. Kolinsky, J. Morais, M. Zizi, Investigating human cognitive performance during spaceflight, Journal of Gravitational Physiology 12 (1) (2005) P39–P40.

[11] N. Pattyn, P.F. Migeotte, X. Neyt, R. Cluydts, Reactivity revisited: a dual physiological effect of stress and mental challenge on cardio-respiratory parameters and respiratory sinus arrhythmia, Psychophysiology, 2007, submitted for publication.

[12] E. Hudlicka, M.D. McNeese, Assessment of user affective and belief states for interface adaptation: application to an air force pilot task, User Modeling and User-Adapted Interaction 12 (2002) 1–47.

[13] A.F. Sanders, Towards a model of stress and human-performance, Acta Psychologica 53 (1983) 61–97.

[14] R.M. Baevsky, V.M. Baranov, A.G. Chernikova, I.I. Funtova, A.V. Pashenko, J. Tank, Results of cardiorespiratory system autonomic regulation investigations during long term International Space Station missions: experiment “PULSE”, in: Joint 9th European Symposium on Life Sciences Research in Space & 26th Annual International Gravitational Physiology Meeting, 26th June–lst July 2005, Cologne, Germany, 2005.