

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



Contents lists available at ScienceDirect

Biological Psychology

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

A methodology to compensate for individual differences in psychophysiological assessment[☆]



Bernd Johannes^{a,*}, Anthony W.K. Gaillard^{b,1}

^a German Aerospace Center (DLR), Aerospace Medicine, Linder Höhe, D-51147 Köln-Porz, NRW, Germany

^b TNO Soesterberg, Priv.: Kolonieweg 16, 3768 EX Soest, The Netherlands

ARTICLE INFO

Article history:

Received 16 January 2013

Accepted 8 November 2013

Available online 6 December 2013

Keywords:

Individual differences

Autonomic response pattern

Autonomic space

Heart rate

Pulse transit time

Skin conductance

Finger temperature

Pattern-specific integration

ABSTRACT

The main methodological drawback to use physiological measures as indicators of arousal is, the large interindividual variability of autonomic responses hindering the direct comparability, between individuals. The present methodology has been tested in two cohorts ($n_1 = 910$, $n_2 = 845$) of, pilot applicants during a selection procedure. Physiological data were obtained during two mentally, demanding tasks and during a Flight Simulator Test. Five typical Autonomic Response Patterns (ARP), were identified by cluster analyses. Autonomic spaces were constructed separately for each group of, subjects having the same typical ARP, on the basis of their normalized eigenvectors. The length of the, vector sum of scores on autonomic space dimensions provided an integral index for arousal, labeled, Psychophysiological Arousal Value (PAV). The PAV still reflected the changes in mental load during the, tests, but equalized physiological differences among ARP-groups. The results obtained in the first, cohort were verified in the second cohort.

© 2013 The Authors. Published by Elsevier B.V. All rights reserved.

1. Introduction

Psychophysiological measures have been used in a wide area as an index for arousal, or indirectly as a measure of mental load and stress (see Gaillard, 2008). Although promising results have been obtained, the reliability of these measures has been criticized. Several approaches have been developed to make measures more robust. One way pursued is to combine the data obtained from different variables into an overall arousal scale (Baevisky, 1997, 2002; Moran, Montain, & Pandolf, 1998; Steptoe & Vögele, 1991). This approach, however, has been criticized by several authors (e.g., Fahrenberg & Foerster, 1982). Two disadvantages have been raised. Firstly, the common correlation matrices of the different physiological measures show too much inconsistency (Haynes & Wilson, 1979; Wenger & Cullen, 1972), due to large individual differences. Secondly, the intensity of a particular reaction depends on the initial state (Wilder, 1950) of a subject or more correct on the initial localization of the subject in the autonomic space (Berntson,

Cacioppo, & Quigley, 1991, 1993). To solve these problems, several researchers tried to identify patterns of psychophysiological responses by means of multivariate methods, such as pattern classification analysis (Christie & Friedman, 2004; Kreibitz, Wilhelm, Roth, & Gross, 2007) or by cluster analysis (Allen, Boquet, & Shelley, 1991; Speisman, Osborn, & Lazarus, 1961; Stephens, Christie, & Friedman, 2010).

Another approach is to investigate the mechanisms underlying the changes in heart rate. Cacioppo and coworkers (Berntson, Cacioppo, Quigley, & Fabro, 1994; Cacioppo, 1994; Cacioppo, Uchino, & Berntson, 1994) focused on source analyses “beyond” heart rate, which resulted in a three dimensional autonomic-response model for the autonomic control of heart rate (see also Backs, 1998, 2001). The present approach (see also Johannes & Salnitski, 2004; Johannes, Salnitski, Soll, Rauch, & Hoermann, 2008) is an extension of the method developed by Cacioppo and coworkers. The main difference with our scaling approach is that the number of coordinates and the number of included end organs are extended. Our method is based on the assumption that an orthogonal vector model is able to assess the sum of the autonomic mechanisms affecting physiological variables. This results in an “autonomic space” in which the eigenvectors¹ are the

[☆] This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

* Corresponding author. Tel.: +49 2203 6013130; fax: +49 2203 695211.

E-mail addresses: bernd.johannes@dlr.de, bernd.johannes@hotmail.com (B. Johannes).

¹ Present address: Postbus 23, 3769 ZG Soesterberg. Tel.: +31 35 6022 474.

¹ In a factor analysis the eigenvector represents the power (or strength) of a particular factor, based on the interconnectedness and the common variance of the contributing variables which load on that factor. To limit the number of factors, in

dimensions of the space model and the eigenvector scores represent the loads of all measures to be integrated. The aim of the present work is to develop a methodology, which is in addition able to account for individual differences. This methodology identifies individual autonomic response patterns (ARPs) in psychophysiological variables and uses pattern-specific integration.

In earlier studies (Johannes, Eichhorn, & Fischer, 1994; Johannes, Salnitski, Thieme, & Kirsch, 2003) we demonstrated that ARPs obtained during a mentally demanding test, are affected differently, as a function of environmental demands, coping styles and individual states. We also showed that the number of typical ARPs is limited. In the present study we will show that there are five typical ARPs in healthy subjects. These ARPs can be used for pattern-specific scaling, when it can be demonstrated that the eigenvectors of different ARP-groups are similar, even when the overall correlations between subjects are low (Haynes & Wilson, 1979). If reliable ARPs can be identified, one can normalize the scores of the eigenvectors separately for each group of subjects showing the same typical ARP. In this way, pattern-specific vector units are obtained which can serve as reference to make comparisons among subjects, even when their ARPs are different. The vector length in this vector space is assumed to represent the sum of autonomic reactivity in a particular subject. In this way the intensity of the response can be compared between subjects.

2. Methods

The selection procedure of pilots and other air personnel at the German Aerospace Center (DLR) (Goeters, 1998a, 1998b; Hoermann, 1998a, 1998b) is a well-organized and documented procedure. A quality management system, certified according to ISO 9000, guarantees the continuous control of quality. A Flight Simulator Test (FST) is a standard part of the selection procedure. Only the applicants, who successfully passed all tests on the first day, were admitted to perform the FST on the second day. The physiological assessment was used for research purposes only, not for the selection. The psychophysiological assessment was ethically approved by the local committees for several studies (space experiments, bed rest studies, isolation studies). The subjects signed an informed consent form prior to the selection procedure. This included explicitly the voluntary participation in the additional scientific program.

2.1. Subjects

All subjects were ab-initio candidates for airline pilot training. The pilot applicants underwent a two-day selection process. There were two cohorts. Whereas the results of earlier studies (Johannes et al., 2008) were based on all available complete data sets, the present data were cleansed to exclude outliers which could have disturbed the grouping results. In Cohort 1 910 subjects (838 male, 72 female) remained and 845 (771 male, 74 female) in Cohort 2 (exclusion procedure is described in detail in section 4.1). The cohorts did not significantly differ with respect to age (21.15 years, SD 2.43). The subjects were slim (BMI 23), nonsmokers and sportive active. The herein newly presented cohort 2 consisted of inhabitants of the FRG (809, 95.7%), Austria (20, 2.4%), Switzerland (5, 0.6), and 8 other European countries (11, 1.1%).

2.2. Procedures

The assessment of the physiological measures took place in two separate sessions. On the first day subjects had to perform two mentally demanding cognitive tasks in the TTS and on the second day they performed the FST.

2.2.1. Two task session (TTS)

The subjects were examined in groups of five in an air-conditioned room maintained at 21 °C. Computer monitors were located in front of them on small tables. The headsets were used to record each subject's voice commands and to give audio instructions. The preparation phase took 15–20 min, including the task instructions. The two tasks were applied in a fixed order (MANOMETER task, MATRICES-task). To maximize the loading effect the MANOMETER-task was always presented first, because this task was found to be more challenging in earlier research (Johannes et al., 1994) focusing on task types and task intensities as recommended by Allen et al. (1991) or Steptoe and Vögele (1991). Several researchers have recommended an "active" rest (Piferi, Kline, Younger, & Lawler, 2000). Therefore after each task a

most applications factors having an eigenvector lower than one are omitted from further analyses, not yet herein. Therefore we will use the original term eigenvector.

3-minute video-clip (with accompanying tranquil music) was presented to induce relaxation. Each test phase lasted 5–6 min, depending on the subject's working speed.

The mental load was manipulated differently in the two tasks. In the MANOMETER-task time pressure was enhanced by increasing the number of gauges to be pursued. All pointers had to show in one half of the screen direction (right, left, upper, lower half) as given in the upper part of the display. In this case subjects had to check if the system is "Okay!". If at least one of the pointers differed more than 90 degrees from the given direction subjects had to state "Error!". The pace of the presentation varied in such a way that the error rate remained on the same level (between 20 and 25%), the number of gauges increased from five to nine.

The MATRICES-task consisted of cognitive solving problems similar to those of the Raven test (Raven, 1971a, 1971b). The series started with easy tasks. Ten cognitive problems had to be solved by the subjects. An adaptive testing procedure was used: if the correct solution was found, the next task had the next higher level of difficulty. If a task was not solved correctly, the task was on the same level. If the subject failed again to find the solution of this task, one task of the next lower level of difficulty was given.

2.2.2. Flight Simulator Test (FST)

The FST took place in an air-conditioned room. Two subjects were tested simultaneously by two instructors. The test consisted of three training tasks, followed by three test tasks. During the training tasks, the instructor answered all questions and provided information and help as comprehensively as possible. The FST comprised challenging flight exercises under instrument flight rules (IFR) and was described in detail in Johannes et al. (2008).

2.3. Variables

2.3.1. Psychological variables

In a first phase of the pilot selection procedure several biographic and psychological data were obtained. Among them, the Temperament Structure Scales (TSS, Hoermann & Maschke, 1996; Maschke, 1987) was applied for personality assessment. After the FST was completed, subjective statements from the participants were recorded on paper protocols. In the cohort 2 the subjects answered additionally the NASA Task Load questionnaire (TLX, Hart & Staveland, 1988) and the instructors evaluated the perceived excitation of the subjects by means of a nine-step Likert scale. The task performance during the FST was evaluated by visually analyzing a chart plot of the track, the altitude profile and the speed profile. Using standardized criteria the visually detected deviations were transformed into Stanine values for five dimensions and combined into one integrated task result. A final overall score was given as an expert rating (Sus, 1993). The instructors evaluated the subjects for "resilience" using a nine-step Likert scale.

2.3.2. Psychophysiological variables

Only the main points of the measurement system and the data-analysis are described here; see Johannes et al. (2008) for more details. A compact light weight version of the HealthLab system (Koralewski Industrie Elektronik oHG (KIE), Hamburg, Germany, www.koralewski.de) was used for the measurements.

The selection of the physiological measures was based on a series of previous studies (Johannes et al., 1994) in which several measures were examined on usability and validity, in particular for applications in the field. More systemic set of measurements were used in other application studies (e.g. Ledderhose et al., 2010) with enhanced laboratory possibilities for the baseline assessment. But herein blood pressure was successfully applicable only during the TTS. The variables of blood pressure proved to be highly relevant for the pattern (ARP) differentiation. In particular, ARP 5 showed a hypertensive pattern, characterized by high blood pressure (and low PTT values). Due to technical and safety limitations impedance cardiography and blood pressure cuff (arm or finger) are not appropriate in field applications like flying a plane, docking a space craft, or the FST. Therefore, during the FST we used PTT which is assumed a reliable correlate of blood pressure (Obriest, Light, McCubbin, Hutcheson, & Hoffer, 1980; Payne, Symeonides, Webb, & Maxwell, 2006; Steptoe, Smulyan, & Gribbin, 1976; Weiss, Del Bo, Reichel, & Engelman, 1980).

During the TTS and the FST electrocardiogram (ECG), skin resistance, finger skin temperature (FT) and pulse wave were registered continuously. In addition, during the TTS oscillographic blood pressure (BP) was monitored and respiration registered by means of a resistance belt, fitted to the subject's chest. The pulse transit time (PTT) served as an indicator of blood pressure changes during the FST. The PTT was calculated as the interval between the R-peak of the ECG and the time point of the highest slope of the first front of the pulse wave. The tonic parameter skin conductance level (SCL) was calculated from the skin resistance between the finger sensor (dry Ag/AgCl electrode) and the ground electrode of the ECG. Finger temperature was registered by a thermo sensor (type FS-03/M) integrated together with the SCL electrode and the plethysmography sensor into a compact finger sensor. The oscillographic BP cuff (MobiGraph) was fitted to the subject's right arm. The calculation of the physiological measures (means and standard deviations (SDs)) was performed by means of the NEURON-32 software package of the HealthLab system. Artifacts were automated detected, counted and excluded by the firmware of the HealthLab system. The artifact recognition system was verified in former applications and of

sufficient quality for laboratory and cockpit situations and healthy subjects. In addition during the statistical analyses of the primary measures all outliers (≥ 3 SD) were excluded.

For the identification of the ARPs 14 measures were used. For the scaling approach the following seven physiological measures (Systolic BP, SD_Syst; Diastolic BP; SD_Diast, SD_PTT, SD_FT, SD_SCL) were discarded, because they were inappropriate for application (blood pressure measures) or did not fulfill criteria, such as robustness (non-sensitivity to artifacts) and reliability. Whereas the Mean of SCL, FT and PTT still are reliable under field conditions, the SDs of these measurements are too much contaminated with artifacts, excluding them for an automated analyses.

2.4. Methodology

To correct the individual differences during a psychophysiological assessment on the basis of their ARPs the following steps were taken; the steps are described in detail in the supplement.

First, for the identification of ARPs (see also 2. in the supplement) physiological responses were obtained during two moderate mentally loading tasks and a relaxing condition (Two Task Session, TTS, described in detail in the Method section). Using cluster analysis five typical ARPs were identified. On the basis of their typical ARP subjects were assigned into five ARP groups.

Secondly (see also 3.1. in the supplement), the eigenvectors of the physiological data obtained by Exploratory factor analysis (FA) in the FST were analyzed separately for each ARP-group. Using Confirmatory FA (see 3.2. in the supplement) it was shown that the eigenvector structures of the ARP-groups were identical. Separately for each ARP-group the eigenvector scores were normalized (see 3.3. in the supplement).

Thirdly, for each subject the normalized eigenvector scores were integrated into a vector sum (see 4. in the supplement), which provided an index of psychophysiological arousal (Psychophysiological Arousal Value, PAV). The lack of differences between the ARP-groups in the PAV-scores demonstrated that the present method is able to compensate for individual differences in physiological reactivity.

2.5. Study set-up

To evaluate the above presented methodology physiological measures were collected during the TTS and during the FST. The typical ARPs obtained during the TTS (see 5.1. in the appendix) as well as the PAV calculation (see 5.2. in the appendix) will be verified in a second group of subjects (Cohort 2).

We tested the following hypotheses:

1. The number of typical ARPs among subjects is limited.
2. All the ARP-groups have the same structure of the eigenvector space.
3. The integral PAV does not differ among ARP-groups.
4. The PAV still reflect the changes in mental demands.
5. The typical ARPs found in the first group of subjects (Cohort 1) can be replicated in another group (Cohort 2).
6. The integration method, calculated in Cohort 1 can be used to equalize the physiological differences among ARP-groups in Cohort 2.

2.6. Statistics

All statistical analyses were run using SPSS for Windows 15.0 (Bühl & Zöfel, 2000) and the related AMOS 7.0 software. For all tests a significance level of .05 was used. Due to the large size of the cohorts in several cases a less liberal alpha was used (.001) as indicated in the results section.

To compare the cluster solutions both within and between the two cohorts the contingency coefficient (cc), the lambda (λ), and the kappa (κ) were used.² Differences among ARP groups in the raw data as well as in the PAV data were analyzed with multivariate analysis of variances, repeated measurements. For the content validation of the PAV, correlations were calculated with subjective ratings of both the participants and the investigators.

3. Results

3.1. Description of cohorts

Before the data analyses, out of the 1044 subjects 922 were selected which had no outliers (mean ± 3 SD) and a complete set of the 14 final measures. The single linkage cluster analyses used first (see supplement, 2.) excluded another 12 subjects, so that from 922 subjects 910 subjects remained in the final Cohort 1. *The identical*

data selection procedure was applied to Cohort 2. Of 1076 cases only 845 cases remained within this cohort after excluding any missing and outlier data.

Table 1 presents the means and standard deviations (SDs) of the physiological measures of Cohort 1 and Cohort 2. The cohorts responded in a similar way; although a few significant differences were found. In general, the task and rest conditions revealed normal distributions for the means of heart period duration, systolic and diastolic blood pressure and PTT in both cohorts. The SDs of these parameters were not normally distributed in both cohorts, which is mostly the case in this type of data.

3.2. Results of the pattern detection

In both cohorts, cluster analyses were run with 14 variables using the Ward method. A detailed analysis of the differences between the cluster solutions was done for each cohort, verifying 4, 5 and 6 cluster solutions. *K*-means cluster analyses were consequently run with 4, 5 and 6 given clusters in both ascending and descending order.

The results of *k*-Means clustering of Cohort 1 in ascending order were compared when the results obtained with descending order. Table 2 presents the comparison of the results of the final *k*-Means clustering of the Cohort 1 and the validation Cohort 2. The highest repetition score was found for the 5 cluster solution.

Fig. 1 shows the ARPs for the 5 cluster solutions. The structures of Cohort 1 and Cohort 2 for their independent 14 variable *k*-means solutions appear quite similar. While a clear pattern difference could be found between the solutions with 5 clusters and 4 clusters a further differentiation into 6 clusters did not result in any additional new pattern. The differences between single measures were small and did not provide a new structure. The solutions for 4 cluster and 6 cluster are not presented in the figure. Normal distribution of the data was tested separately for each cluster. In general the mean values were normally distributed except for the SCL (in all clusters) and the FT (in cluster 1, 2 and 5). In contrast, only half of the SDs was normally distributed. Although the homogeneity within these clusters was not ideal, they were much higher than in the data across cohort. The subjects of Cohort 2 were classified into ARP-groups by the discriminant functions (see supplement 5.1.) calculated on TTS data of Cohort 1. Fig. 2 presents the physiological data of TTS (left side) and FST (right side) for Cohort 2. The differences between ARP-groups were analyzed by means of the General Linear Model (repeated measures), revealing significant differences between ARP-groups clusters for the TTS and the FST. In the analyses of variance of the FST only the test period was included in which there was no interaction between applicant and instructor (only for preparation (P) and test tasks (T)).

The FST protocol evoked significant changes in all ARP-groups for the five physiological variables. The means in the task phases during the FST were larger than in the preparation phases. The training and exercise phase prior the test period provided still lower measures as illustrated in Fig. 2 by exercise task 3 (E3). The analyses of variance confirmed that under high load ARP-groups differ in physiological responses.

Since the subjects were classified on the basis of differences in their physiological parameters obtained in the TTS, the group differences for those conditions are to some extent tautological. The similarity and significant differences in the physiological raw data during the FST next day supports validity and reliability of the classification in the ARP-groups.

3.3. Results of the integration method

The data of test task 1 of the FST were factor analyzed for eigenvectors. The correlation matrix of the seven variables (see

² Whereas the cc describes any relationship between frequency distributions of different categories, the λ and the κ describe symmetric relationships like correlation coefficients. The latter one requires the same number of clusters in compared groups and represents the "classical" statistic in former studies.

Table 1
The mean (M) and SD of the physiological measures obtained during the TTS cross the experimental phases, separately for Cohort 1 (n=910) and Cohort 2 (n=845). The p-levels refer to the differences between the means of the cohorts. A restrictive alpha-level of .001 is applied. The abbreviations in parentheses are also used in the figures.

	Cohort 1		Cohort 2		p
	M	SD	M	SD	
Heart Period Duration (ms) (HPD)	671.59	88.73	682.18	96.82	.017
Heart Period Standard Deviation (ms) (SD_HPDP)	34.19	13.66	34.53	23.80	.621
Root of Mean Successive Square Differences of HPD (ms) (RMSSD)	33.14	17.66	37.55	23.80	.000
Pulse Transit Time (ms) (PTT)	253.16	23.28	257.06	24.43	.001
Skin Conductance (μS) (SCL)	10.04	6.71	9.78	6.42	.417
Finger Temp. (°C) (FT)	30.57	3.88	31.17	3.61	.009
Finger Temp. Standard Deviation (°C) (SD_FT)	.52	.36	.57	.45	.000

Table 2
Confirmation of k-means cluster solutions in Cohort 2 (based on 14 measures) between Cohort 1 clusters for external classification (rows) and Cohort 2 clusters as internal classifications (columns); all contingency coefficients (cc) and symmetric Lambdas (λ) and Kappa's (κ) were highly significant (p < .001).

	Cohort 2 k=4	Cohort 2 k=5	Cohort 2 k=6
Cohort 1 (k=4)	cc = .728; λ = .510; κ = .556		
Cohort 1 (k=5)		cc = .781; λ = .578; κ = .547	
Cohort 1 (k=6)			cc = .766; λ = .399; κ = .284

2.3.2.) had positive manifold (each column had a positive sum, see Nunnally & Bernstein, 1994, p. 470). The first factor was found to be general, since all variables had a positive load. The next four factors appeared to be bipolar, i.e. half of variables having a positive, the other half having a negative load. The analysis of the correlation

between factor scores and the measure with the highest load identified the direction of these factors as an “arousal” indicator. A five vector model appears to be the optimal solution. All communalities were .95 or higher. The explained variance was about 98.35%; the first rotated vector explained 30.96%, each of four others round

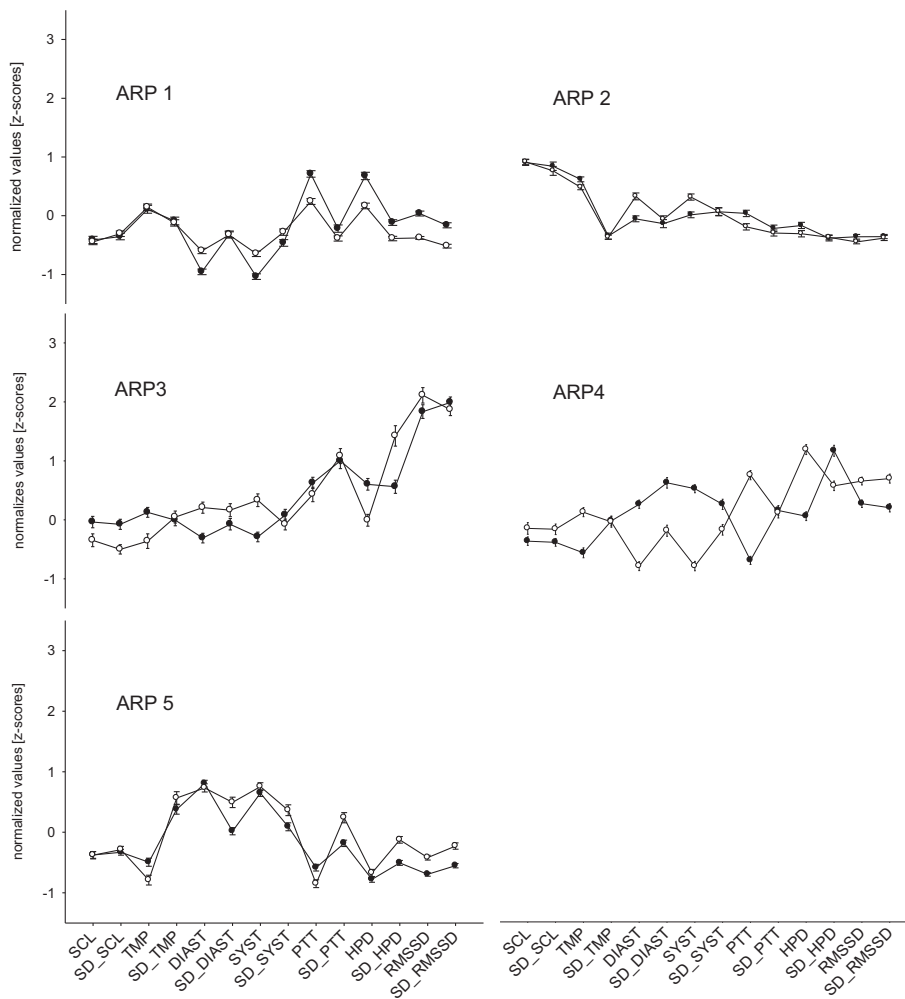


Fig. 1. Pattern of fourteen physiological measures obtained during the TTS in a 5-cluster solution (ARP1 to ARP5), separately for Cohort 1 (filled circles) and Cohort 2 (open circles).

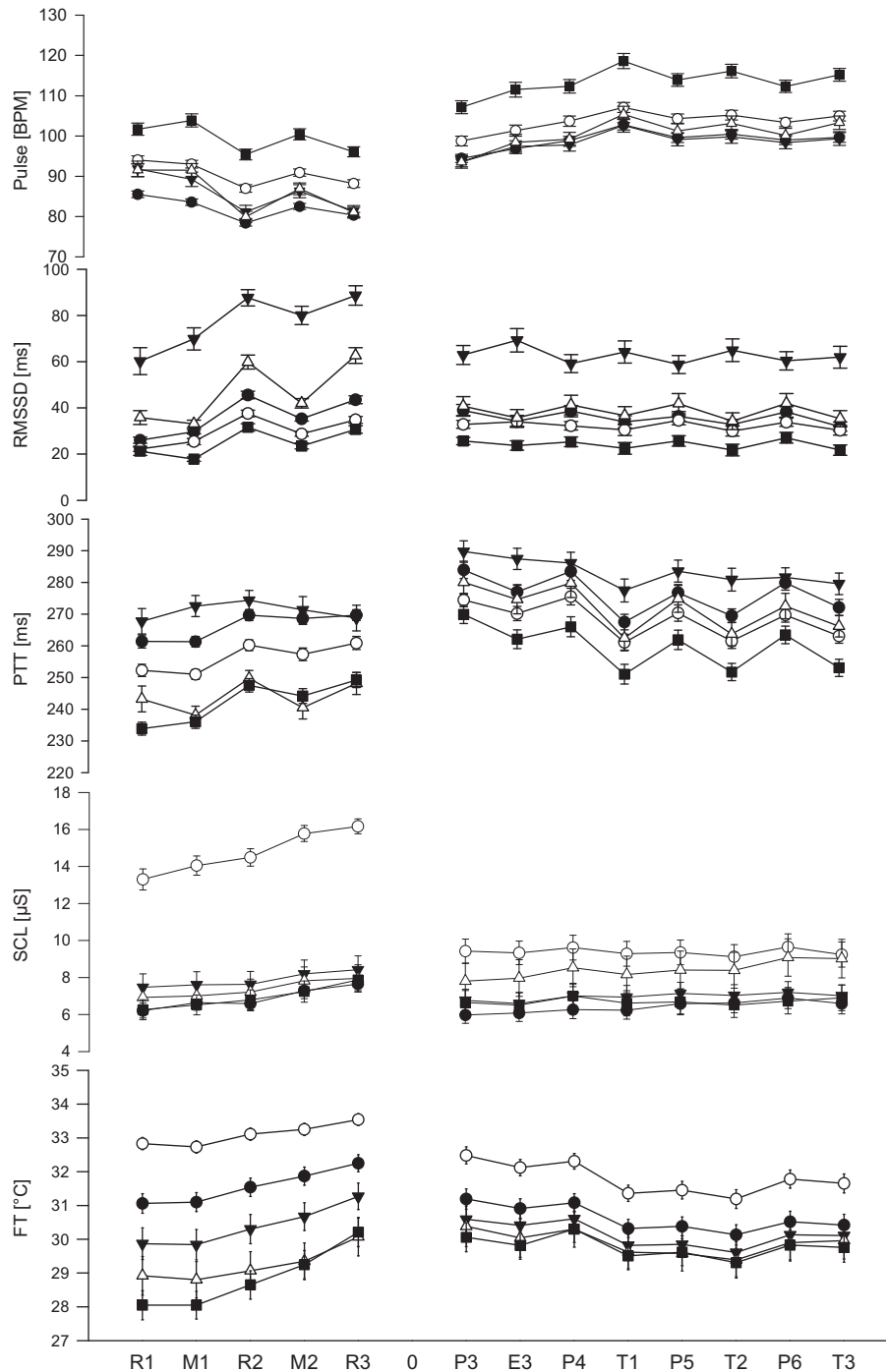


Fig. 2. Five physiological measures plotted separately for the five ARP-groups of Cohort 2 (ARP 1 filled circles –●–, ARP 2 open circles –○–, ARP 3 filled triangles –▼–, ARP 4 open triangles –△–, ARP 5 filled squares –■–) during the TTS (left side) and the FST (right side). The phases in the TTS were R1, R2, R3–rest, M1, M2–mentally loading tasks. The phases in the FST were as follows: P3–preparation of the third exercise, E3–the 3rd exercise task, three test tasks (T1, T2, T3) with preparation phases (P4, P5, P6) before.

about 17%. Each factor was mainly correlated to one physiological variable, $r > .95$. Only the variables HPD, RMSSD and PTT showed also correlations with the other dimensions (less $r = .2$), all other intercorrelations were lower than .05.

The factor scores (Fig. 3) were very different among the ARP-groups ($F(4) = 10.311, p = .000$) and showed a significant interaction between factors and ARP-groups ($(F(16, 1696) = 3.082, p = .000)$) indicating that the eigenvectors in the ARP-groups have different lengths. But Confirmatory FA demonstrated that the five ARP-groups in both cohorts had similar eigenvector structures. The model of the Confirmatory FA is presented in Fig. 4. Although the

strength of the eigenvectors was different, five eigenvectors having a similar structure were found in all the 10 ARP-groups.

Applying this model to the five ARP-groups of both cohorts provided slightly different parameters of correlation and co-variations, but generally the model fits the data of all groups as demonstrated by the chi-square testing (Table 3).

The above analysis allowed the separate normalization of the factor scores for each ARP-group and each factor. We integrated the normalized factor scores of the single cases, as the length of the resulting vector sum of the physiological measures representations on this vector frame.

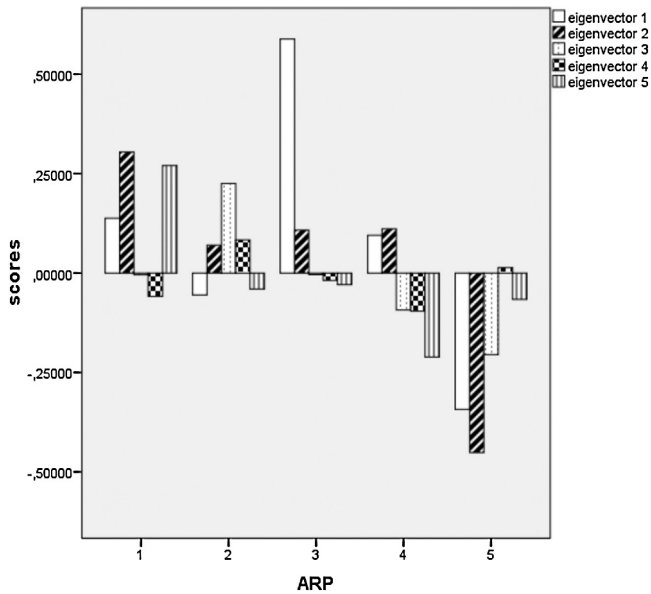


Fig. 3. Mean eigenvector scores of the five eigenvector model separately for the five ARP-groups.

Table 3

Chi-square tests of Confirmatory FA. For each ARP-group the same model of the factor structure was found in both cohorts to fit, supporting the identity of the factor structure as the basis of the integration approach. The fit of the model was to be rejected if the *p* was less than .05.

ARP-group	Cohort I			Cohort II		
	Chi-square	<i>p</i>	<i>n</i>	Chi-square	<i>p</i>	<i>n</i>
1	9.26	.598	138	19.04	.087	140
2	17.96	.117	159	17.97	.116	127
3	16.98	.151	74	13.21	.354	86
4	20.29	.062	97	5.87	.922	53
5	7.08	.852	146	14.96	.244	85
Total			614			491

These multiple regression functions and vector sums were applied to the FST data of Cohort 2 with regard to the assigned pattern classification. Fig. 5 demonstrates that the individual differences among the ARP-groups of the validation cohort disappeared by integrating the data on the basis of their PAVs. For the GLM analysis, the three test tasks (T) and their preparation phases (of Cohort 2 were included. No difference was found between ARP-groups in the PAVs in Cohort 1 ($F(4) = .318, p = .866$) and Cohort 2 ($F(4) = .975,$

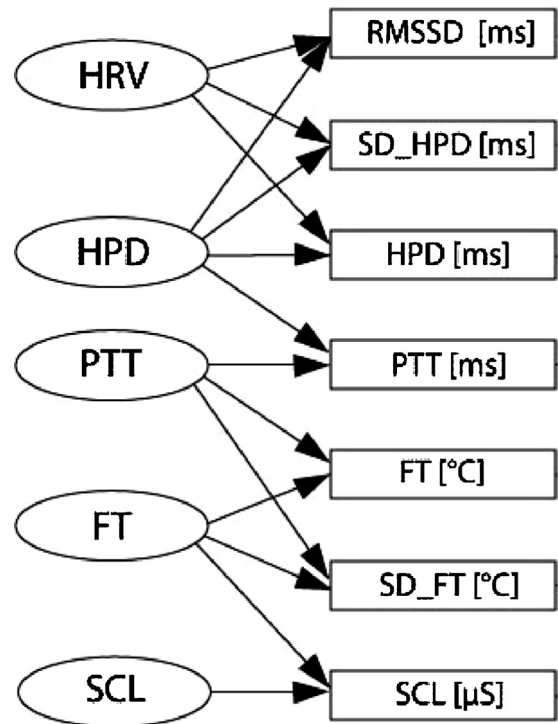


Fig. 4. In the applied model for confirmatory factor analysis five main factors explain the variance of seven physiological measures (rectangles). The respective measurement errors are not inter-related and therefore omitted.

$p = .421$). Larger PAVs were found in the Task phases of the FST than in the preparation phases in both Cohort 1 ($F(5, 377) = 78.64, p = .000$) and Cohort 2 ($F(5, 319) = 68.80, p = .000$), which demonstrates that the PAVs reflect the changes in reactivity induced by the changes in mental load. Moreover, no interactions were observed between the ARP-groups and FST phases (Cohort 1: $F(20, 1520) = .898, p = .590$; Cohort 2: $F(20, 1288) = .830, p = .678$). Thus all ARP-groups responded similar during the FST.

The ratings of the instructors were not correlated with the physiological parameters or with the integrated PAV. In contrast, significant correlations ($r = .3-.4, p = .000$) between these ratings were found with different performance ratings. Of the correlations between PAV-values and performance-ratings, only one significant but low correlation (“exactness in timing”; $r = .063, p = .044$) was found.

Subjective TLX ratings of the applicants, applied with a subgroup of 350 participants, did not correlate with any physiological

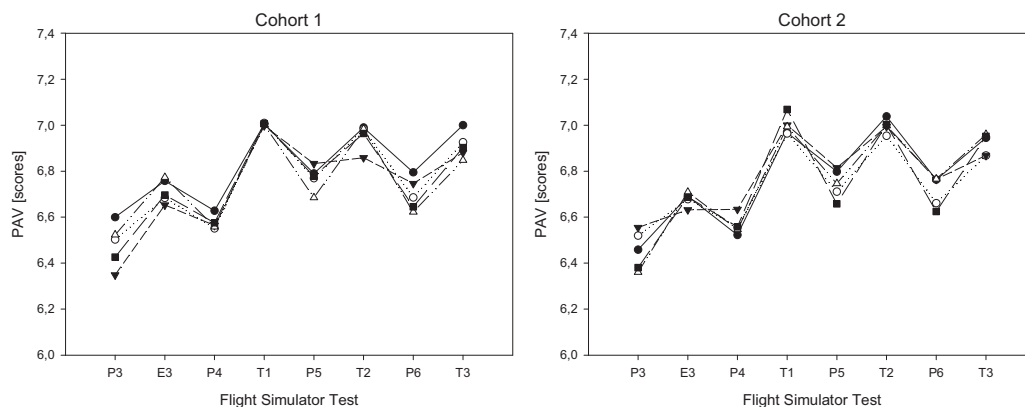


Fig. 5. The integrated Psychophysiological Arousal Value (PAV) during the FST, for Cohort 1 and for Cohort 2. The curves for the five ARP-groups are plotted (ARP 1 filled circles –●–, ARP 2 open circles –○–, ARP 3 filled triangles –▼–, ARP 4 open triangles –△–, ARP 5 filled squares –■–). See Fig. 2 for abbreviations.

measure. There were correlations between two subscales of the TLX (“frustration” ($r = -.159, p = .005$) and “performance” ($r = -.321, p = .000$)) with the “resilience” rating given by the instructors. The personality traits did not differ between the two cluster groups, except in one scale (spoiltness, between cluster 3 and 5). Compared to participants failing the selection participants, which successfully passed the selection, showed higher values in PAV in the test tasks of the FST ($F(1) = 4.966, p = .027$). This effect was stronger the longer the aspirants were working on the test ($F(2,336) = 3.193, p = .042$). Although the successful participants rated themselves as more resilient (Wilcoxon: $p < .001$), the rating about their level of excitation was lower (Wilcoxon: Task 1, $p = .007$; Task 2 $p = .000$, Task 3 $p = .001$). A more detailed psychological pass-fail analysis was recently provided by Hoermann and Goerke (2013).

4. Discussion

The present study demonstrates the usability, reliability and validity of a scaling methodology to calculate an index for psychophysiological arousal, which is not affected by individual differences in responsiveness (see Fig. 2) of the autonomic nervous system. The present results show that the number of typical response patterns to mental load can be limited (Table 2, see hypothesis 1). Cluster analysis demonstrated that a solution of five typical response patterns was optimal. *This restricted range facilitates the applicability in practical situations.* As is shown in Fig. 1, these patterns are quite similar in Cohort 1 and 2.

The present classification approach was first used in clinical studies (Johannes, Salnitski, Thieme et al., 2003; Thieme, 1996; Thieme, Johannes, & Gromnica-Ihle, 1995). The methodology was applied to detect the autonomic response pattern in different groups of patients. Differences were observed between the response patterns of persons with hypertension and rheumatic diseases as to normal controls. We assume that including cohorts with diverse clinical diagnoses may widen the applicability of the present methodology for clinical purposes.

The results support the notion of Fahrenberg and Foerster (1982), that two experimental conditions are sufficient to make a reliable estimate of an individual response pattern. Significant differences were found among the five ARP-groups in all measures, whereas each measure still reflected the changes in demands in both the TTS and the FST (see Fig. 2).

For each of the five groups of subjects having the same typical ARP an orthogonal autonomic space was constructed. Using Confirmatory FA it was demonstrated that the structures of the autonomic spaces were the same in the two cohorts for all ARP-groups (see also Fig. 4, Table 3, and hypothesis 2).

The present methodology provides a set of plausible autonomic correlates. The factor HRV appears to reflect the parasympathetic control of the heart as proposed by Berntson et al. (1991), the factor HPD could represent the “residual heart rate” (Bucks, 1998; Grossman & Svebak, 1987), the beta-adrenergic sympathetic overbalance heart control component. PTT is known to be to correlate with sympathetic changes (Obrist et al., 1980; Weiss et al., 1980) in blood pressure (Payne et al., 2006; Steptoe et al., 1976). FT correlates with sympathetically changes in vasoconstriction (Surwit & Fenton, 1980) and SCL reflects the sympathetic innervation of the sweat glands (Christie & Friedman, 2004; Dawson, Filion, & Schell, 1989). The PAV can be considered as an extension of the autonomic space model of Berntson et al. (1991). It is un-weighted (vector) sum of parasympathetic withdrawal and sympathetic innervation on different effector organs of the ANS.

In these autonomic spaces the eigenvectors have different lengths in the different ARP-groups but their structure and direction in the autonomic space are identical (comparable to cubes

which may be different in “length”, “height” and “depth”, but have all a rectangular form). This result represents the main verification, that the PAV model can be applied to all different ARP groups.

Since raw data shows large differences among ARP groups, it is not possible to make direct comparisons between the responses to mental load of different subjects (Fig. 2). Therefore the physiological measures were converted into so-called PAVs. This enables the integration of different physiological measures coming from different end-organs into one scale. In this way the response intensity of subjects with different ARPs can be compared. As can be seen in Fig. 5 the PAVs do not differ between ARP-groups (hypothesis 3), but they still reflect the changes in mental load during the FST (hypothesis 4). The integration of different physiological parameters also takes into account that subjects (with different ARPs) may respond to mental load with different physiological processes.

The reliability of the methodology was tested by replicating the study in a second group of subjects. As is shown in Fig. 1, in Cohort 1 the 5-cluster solution appeared to be optimal and the pattern structure was similar to the solution found in Cohort 2 for each ARP-group (hypothesis 5). In addition, the vector structure of the autonomic space and the PAV values mimicked the results found in Cohort 1 (see Fig. 5).

The developed scale of PAVs enabled comparisons between subjects of cohort 2 (hypothesis 6), which have different patterns of autonomic reactivity.

The correlations between subjective measures of load (ratings by participant and by instructor) and performance results were difficult to interpret. They might suggest an anchoring of the subjective load evaluation on the success during the test. The lack of correlation between subjective and objective assessments of load and arousal is not necessarily problematic, because it shows that the two types of measures are influenced by different underlying mechanisms which together will provide a better understanding of how people respond to mental load (e.g. Yeh & Wickens, 1988). Some data suggest opposite results. Applicants which passed successfully the FST showed a higher PAV than participants which failed. However, they evaluated themselves as lower excited and were evaluated by the instructors to be more “resilient”. That underlines the necessity to develop better objective methods for further research.

This appears as a step forward from an autonomic space of heart rate regulation (Berntson et al., 1991, 1993; Cacioppo, 1994; Cacioppo et al., 1994) towards an autonomic space of several target organs of the autonomic nervous system. The methodology presented here seems valuable for psychophysiological research, in particular for studies in the field.

The methodology can also be used to minimize the set of relevant physiological measures needed. In former analyses (Johannes & Salnitski, 2004) we started with 22 measures to construct an autonomic space with nine dimensions. In the present study we reduced the measures for scaling to most robust and applicable ones under field conditions. We ended in five dimensions based on only seven measures, which could be integrated into one index of arousal. On the basis of the present results a standardized analysis can be provided within the HealthLab system which may be used also by researchers with limited statistical expertise. *But this methodology can also be applied with most statistical packages. On the web site: www.dlr.de/arp_pav we provide as example the final raw data and a complete analysis script for SPSS.* This methodology could be used to monitor occupational work load (e.g. Veltman & Gaillard, 1996, 1998), it may provide more valid information about autonomic reactions than a single channel approach, e.g. with HR or HRV only.

It has been shown that extreme environments (space, high altitude) evoke not only changes in the intensity of autonomic responses but result also in pattern changes (Baevsky, Chernikova,

Funtova, & Tank, 2011; Fritsche-Yelle, Charles, Jones, & Wood, 1996; Johannes, Salnitski, Polyakov, & Kirsch, 2003). These different patterns could be considered as different functional states. One of the aims of the present study was to identify such changes comparing the patterns obtained before a particular application of the PAV method. Detecting a changed functional state allows now to make the arousal measurements comparable between these states.

The strengths of the present assessment method applicable under field conditions, was tested on large statistical samples. The present study shows that differences in individual response patterns can be taken into account, which improves the reliability of autonomic measures. In this way individuals may be better compared. The application of this method in the field may be limited, because only minimal body movements are permitted and environmental influences on the temperature measurement are excluded.

The mathematical model and the scaling approach have to be verified in more complex situations, where the data are modeled with a larger set of measurements. The detailed characteristics of the PAV scale have still to be analyzed. It is not yet known, which level of monotony and linearity or which kind of non-linearity is given (see also Johannes et al., 2007).

Acknowledgments

The first “Thank you” goes to Ines Rachow, leading psychological technical assistant for pilot selection in the DLR department in Hamburg, for her organizational work and technical assistance with the routine data assessment. We would like to thank the test instructors of the DLR Flight simulator test for their support and additional work with electrodes and signal verification. We have to thank Dr. Klaus-Martin Goeters, that time head of the department, who supported the study in a great way. The authors are deeply thankful to Ron White, Kurt Pawlik and Bob Hockey for methodological and statistical discussions. And last but not least, we thank the space agency German Aerospace Center, DLR, for supporting the present work as well as for providing the hardware and software development services, hence making the highly automated data assessment and extensive data processing/analysis possible.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.biopsycho.2013.11.004>.

References

- Allen, M. T., Boquet, A. J., & Shelley, K. S. (1991). Cluster analysis of cardiovascular reactivity to three laboratory stressors. *Psychosomatic Medicine*, *53*, 272–288.
- Backs, R. W. (1998). A comparison of factor analytic methods of obtaining cardiovascular autonomic components for the assessment of mental workload. *Ergon*, *41*, 733–745.
- Backs, R. W. (2001). An autonomic space approach to the psychophysiological assessment of mental workload. In P. A. Hancock, & P. A. Desmond (Eds.), *Stress, workload and fatigue* (pp. 279–289). New Jersey, London: Mahwah.
- Baevsky, R. M. (1997). Noninvasive methods in space cardiology. *Journal of Cardiovascular Diagnosis and Procedures*, *14*, 161–171.
- Baevsky, R. M. (2002). Analysis of heart rate variability in space medicine. *Human Physiology*, *28*, 202–213.
- Baevsky, R. M., Chernikova, A. G., Funtova, I. I., & Tank, J. (2011). Assessment of individual adaptation to microgravity during long term space flight based on stepwise discriminant analysis of heart rate variability parameters. *Acta Astronautica*, *69*, 1148–1152.
- Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1991). Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review*, *98*, 459–487.
- Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1993). Cardiac psychophysiology and autonomic space in humans: Empirical perspectives and conceptual implications. *Psychological Bulletin*, *114*, 296–322.
- Berntson, G. G., Cacioppo, J. T., Quigley, K. S., & Fabro, V. T. (1994). Autonomic space and psychophysiological response. *Psychophysiology*, *31*, 44–61.
- Bühl, A., & Zöfel, P. (2000). *SPSS Version 9 Einführung in die moderne Datenanalyse unter Windows*. München: Addison-Wesley.
- Cacioppo, J. T. (1994). Social neuroscience: Autonomic, neuroendocrine, and immune response to stress. *Psychophysiology*, *31*, 113–128.
- Cacioppo, J. T., Uchino, B. N., & Berntson, G. G. (1994). Individual differences in the autonomic origins of heart rate reactivity: The psychometrics of respiratory sinus arrhythmia and preejection period. *Psychophysiology*, *31*, 412–419.
- Christie, I. C., & Friedman, B. H. (2004). Autonomic specificity of discrete emotion and dimensions of affective space: A multivariate approach. *International Journal of Psychophysiology*, *51*, 143–153.
- Dawson, M. E., Filion, D. L., & Schell, A. M. (1989). Is elicitation of the autonomic orienting response associated with allocation of processing resources? *Psychophysiology*, *26*, 560–572.
- Fahrenberg, J., & Foerster, F. (1982). Covariation and consistency of activation parameters. *Biological Psychology*, *15*, 151–169.
- Fritsche-Yelle, J. M., Charles, J. B., Jones, M. M., & Wood, M. L. (1996). Microgravity decreases heart rate and arterial pressure in humans. *Journal of Applied Physiology*, *80*, 910–914.
- Gaillard, A. W. K. (2008). Concentration, stress and performance. In P. A. Hancock, & J. L. Szalma (Eds.), *Performance under stress* (pp. 59–75).
- Goeters, K. M. (1998a). Standards of selection: Legal and ethical issues. In K. M. Goeters (Ed.), *Aviation psychology: A science and a profession* (pp. 97–101). Aldershot: Ashgate.
- Goeters, K. M. (1998b). General standards of selection: Validity and utility analysis. In K. M. Goeters (Ed.), *Aviation psychology: A science and a profession* (pp. 103–112). Aldershot: Ashgate.
- Grossman, P., & Svebak, S. (1987). Respiratory sinus arrhythmia as an index of parasympathetic cardiac control during active coping. *Psychophysiology*, *24*, 228–235.
- Hart, S. G., & Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkate (Eds.), *Human mental workload*. Amsterdam, The Netherlands: Elsevier.
- Haynes, S. N., & Wilson, C. L. (1979). *Behavioral assessment: Recent advances in methods, concepts and applications*. San Francisco: Jossey-Bass.
- Hoermann, H. J. (1998a). Fundamentals of selection—basic concepts. In K. M. Goeters (Ed.), *Aviation psychology: A science and a profession* (pp. 47–53). Aldershot: Ashgate.
- Hoermann, H. J. (1998b). Selection of civil aviation pilots. In K. M. Goeters (Ed.), *Aviation psychology: A science and a profession* (pp. 55–62). Aldershot: Ashgate.
- Hoermann, H. J., & Goerke, P. (2013). Assessment of social competence for pilot selection—A combined analysis of self-report and behavior-oriented approaches. *International Journal of Aviation Psychology* (in press)
- Hoermann, H. J., & Maschke, P. (1996). On the relation between personality and job performance of airline pilots. *International Journal of Aviation Psychology*, *6*, 171–178.
- Johannes, B., Eichhorn, C., & Fischer, F. (1994). A complex experimental assessment for objective hierarchical description of hierarchical psychophysiological behavior as human regulatory phenotype. Proceedings of the IUPS-Meeting Gravitational Physiology. *Journal of Gravitational Physiology*, *1*, 73–74.
- Johannes, B., & Salnitski, V. P. (2004). Integration of different autonomic measures into common indicators of psychophysiological costs. In K. M. Goeters (Ed.), *Aviation psychology, practice and research* (pp. 327–342). Aldershot: Ashgate.
- Johannes, B., Salnitski, V. P., Polyakov, V. V., & Kirsch, K. (2003). Changes in the autonomic reactivity pattern to psychological load under long-term microgravity—Twelve men during 6-month spaceflights. *Aviakosm i Ekolog Med*, *37*(3), 6–16.
- Johannes, B., Salnitski, V. P., Soll, H., Rauch, M., & Hoermann, H. J. (2008). De-individualized psychophysiological strain assessment during a flight simulation test—validation of a space methodology. *Acta Astronautica*, *63*, 791–799.
- Johannes, B., Salnitski, V. P., Thieme, K., & Kirsch, K. (2003). Differences in the autonomic reactivity pattern to psychological load in patients with hypertension and rheumatic diseases. *Aviakosmicheskaya i Ekologicheskaya Meditsina*, *37*, 28–42.
- Johannes, B., Wittels, P., Enne, R., Eisinger, G., Castro, C., Thomas, J., et al. (2007). Non-linear function model of voice pitch dependency on physical and mental load. *European Journal of Applied Physiology*, *101*, 267–276.
- Kreibitz, S. D., Wilhelm, F. H., Roth, W. T., & Gross, J. J. (2007). Cardiovascular, electrodermal, and respiratory response patterns to fear- and sadness-inducing films. *Psychophysiology*, *44*, 787–806.
- Ledderhos, C., Rothe, S., Gens, A., & Johannes, B. (2010). Objective psychophysiological strain assessment of AWACS crews in simulators and real flight operations. *Annual Military Scientific Research Report 2009*, Federal Ministry of Defence, 76–77.
- Maschke, P. (1987). Temperament Structure Scales (TSS). European Space Agency, Tech. Rep. ESA-TT-1069. Oberpfaffenhofen: ESA.
- Moran, D. S., Montain, S. J., & Pandolf, K. B. (1998). Evaluation of different levels of hydration using a new physiological strain index. *American Journal of Physiology*, *275*, R854–R860.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, Toronto: McGraw-Hill.
- Obrist, P. A., Light, K. C., McCubbin, J. A., Hutcheson, J. S., & Hoffer, J. L. (1980). Pulse transit time: Relationship to blood pressure and myocardial performance. *Psychophysiology*, *16*, 292–301.

- Payne, R. A., Symeonides, C. N., Webb, D. J., & Maxwell, S. R. J. (2006). Pulse transit time measured from ECG: An unreliable marker of beat-to-beat blood pressure. *Journal of Applied Physiology*, *100*, 136–141.
- Piferi, R. L., Kline, K. A., Younger, J., & Lawler, K. A. (2000). An alternative approach for achieving cardiovascular baseline: Viewing an aquatic video. *International Journal of Psychophysiology*, *37*, 207–217.
- Raven, J. C. (1971a). *The standard progressive matrices* (13th ed.). London: Lewis and Co.
- Raven, J. C. (1971b). *The advanced progressive matrices* (2nd ed.). London: Lewis and Co.
- Speisman, J. C., Osborn, J., & Lazarus, R. S. (1961). Cluster analysis of skin resistance and heart rate at rest and under stress. *Psychosomatic Medicine*, *13*, 333–343.
- Stephens, C. L., Christie, I. C., & Friedman, B. H. (2010). Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis. *Biological Psychology*, *84*, 463–473.
- Stephens, A., Smulyan, H., & Gribbin, B. (1976). Pulse wave velocity and blood pressure changes: Calibration and applications. *Psychophysiology*, *13*, 488–493.
- Stephens, A., & Vögele, C. (1991). Methodology of mental stress testing in cardiovascular research. *Circulation*, *83*, 1114–1124.
- Surwit, R. S., & Fenton, C. H. (1980). Feedback and instructions in the control of digital skin temperature. *Psychophysiology*, *17*, 129–132.
- Sus, I. (1993). Entwicklung eines standardisierten Verhaltensbeobachtungsbogens als diagnostisches Instrument in der Eignungsauswahl von Verkehrsflugzeugführern. DLR-FB 93-55, DLR, Hamburg.
- Thieme, K. (1996). Die psychophysiologische Konstitution von Patienten mit systemischen Lupus erythematodes und Rheumatoider Arthritis. Thesis. Berlin: FB Psychologie der Humboldt-Universität zu Berlin.
- Thieme, K., Johannes, B., & Gromnica-Ihle, E. (1995). Stress reactivity and personality of patients with SLE and RA. *LUPUS*, *4*, 77.
- Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, *42*, 323–342.
- Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, *41*, 656–669.
- Weiss, T., Del Bo, A., Reichel, N., & Engelman, K. (1980). Pulse transit time in the analysis of autonomic nervous system effects on the cardiovascular system. *Psychophysiology*, *17*, 202–207.
- Wenger, M. A., & Cullen, T. D. (1972). Studies of autonomic balance in children and adults. In N. S. Greenfield, & R. A. Sternbach (Eds.), *Handbook of psychophysiology* (pp. 535–576). New York: Holt.
- Wilder, J. (1950). The law of initial values. *Psychosomatic Medicine*, *12*, 392.
- Yeh, Y.-Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *Human Factors*, *30*, 111–120.