Psychoneuroendocrinology 101 (2019) 186–192

Contents lists available at ScienceDirect

Psychoneuroendocrinology

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

Virtually stressed? A refined virtual reality adaptation of the Trier Social Stress Test (TSST) induces robust endocrine responses

Patrick Zimmer[a,1], Benjamin Buttlar[b,1], Georg Halbeisen[b], Eva Walther[b,*], Gregor Domes[a,*]

[a] Department of Biological and Clinical Psychology, University of Trier, Johanniterufer 15, 54290, Trier, Germany
[b] Department of Social Psychology, University of Trier, Universitätsring 15, 54296, Trier, Germany

ARTICLE INFO

Keywords:
Psychological stress
Virtual reality
Trier Social Stress Test
Free salivary cortisol
Hypothalamus-pituitary-adrenal axis
Alpha amylase

ABSTRACT

In recent years, virtual reality (VR) technology has found its way into nearly all fields of psychology. Previous studies indicated that virtual reality adaptations of the TSST are less potent in stimulating HPA-axis responses, with lower salivary cortisol responses recorded as compared to the in-vivo TSST. (TSST-IV). In the present experiment we tested the stress-induction potential of a refined version of the TSST-VR using a fully orthogonal experimental design in which ninety-three healthy males were either assigned to the TSST condition or a corresponding control condition in a real or virtual environment. We found a significant increase of endocrine, autonomic and self-reported stress markers in both stress conditions. Notably, we found a robust rise in salivary cortisol to the TSST-VR comparable to that observed in the TSST-IV. Despite subtle differences in response between virtual and in vivo settings, we conclude that VR adaptations of in-vivo stressors have the potential to induce real physiological and subjective reactions.

1. Introduction

Over the last decades a number of highly standardized laboratory stressors have been developed to induce psychosocial stress in the laboratory (e.g. the Socially-Evaluated Cold Pressor Test, SECPT, (Schwabe et al., 2013); Maastricht Acute Stress Test, MAST, (Smeets et al., 2012). Among these protocols, the Trier Social Stress Test (TSST; Kirschbaum et al., 1993) has become widely used in psychobiological stress research as it has been proven to evoke robust endocrine and cardiovascular responses in the majority of participants.

The TSST mainly consists of a short mock job interview and a mental arithmetic in front of an audience of two or three people. It thus induces the two main factors for robust HPA-axis activation: Social evaluative-threat and uncontrollability (Dickerson and Kemeny, 2004). A recent meta-analysis provided evidence that the TSST is quite robust against protocol variations (Goodman et al., 2017). As long as the protocol comprises both tasks in front of evaluative judges, most participants respond with a significant increase in free salivary cortisol resulting in an overall average two-fold increase over baseline.

Aside from adaptations for specific environments (e.g. MRI, EEG, groups) and populations (e.g. children, elderly), the TSST has been adapted for the use in virtual realities (VR). Using the TSST-VR has three main advantages: Firstly, it significantly reduces the resources needed for research as it makes the presence of extensively trained judges obsolete. Secondly, it offers maximum experimental control, as the agents reliably behave in a highly controlled and standardized way. Lastly, it provides an environment that easily allows for the manipulation of contextual factors (characteristics of the panel, features of the room etc.). It is thus not surprising that a number of preliminary studies have tried to validate their specific adaptation of the TSST-VR (Kelly et al., 2007; Kotlyar et al., 2008; Jönsson et al., 2010; Wallergård et al., 2011) and to provide evidence that the TSST-VR induces a comparable pattern and magnitude of psychobiological reactions as their in-vivo counterpart.

Despite the fact that the published studies on variations of the TSST-VR reported reliable subjective stress responses, most of them demonstrated less robust or lower stress responses of the HPA-axis, concluding that VR adaptations of the TSST are less potent in inducing psychobiological stress reactions. The explicit comparison to a comparable in vivo stressor, however, was not made in most of these studies (e.g. Jönsson et al., 2010; Ruiz et al., 2010; Fich et al., 2014; Montero-López et al., 2016). One recent study by Shiban et al. (2016) implemented a control in vivo condition, and found lower average HPA-axis responding and lower responder rates (using predefined response criteria; Miller et al., 2013) suggesting that the TSST-VR is a milder stressor compared to its in vivo original. One possible explanation is based on

* Corresponding authors.
E-mail addresses: walther@uni-trier.de (E. Walther), domes@uni-trier.de (G. Domes).
[1] Equal contribution.

https://doi.org/10.1016/j.psyneuen.2018.11.010
Received 30 August 2018; Received in revised form 26 October 2018; Accepted 6 November 2018

the assumption that emotional reactions in virtual environments are associated with the individual's feeling of presence (Diemer et al., 2015). Immersion and interactivity have been associated with the subjective presence in VR environments (Baños et al., 2004) and are limited by technical factors such as the graphics engine modeling the virtual environment and the extent to which the agents react to the participants' behavior. In the previous studies, these factors might have limited immersion and interactivity and in consequence may have mitigated the participant's sense of presence which potentially led to lower psychophysiological reactions to the stressful situation.

To overcome these limitations, we designed a virtual reality adaptation of the TSST, in which the virtual surroundings are precisely modeled after the actual laboratory setting. Furthermore, an eye-tracking device was used for real-time feedback of eye-to-eye contact between the participant and the virtual judges. In addition, we modified the VR judges to match their real counterparts as closely as possible, thus maximizing interactivity and sense of presence.

We conducted a standard TSST (Kirschbaum et al., 1993) in VR and in vivo and carefully parallelized both conditions. Similarly, we administered a comparable, but non-stressful placebo version of the TSST (Het et al., 2009) in vivo and in VR. This orthogonal design permitted us to assess the effects of the social stress induction in vivo and in VR independently and therefore detect potentially differential outcomes. We hypothesized that in this rigorous experimental design that uses a refined TSST-VR, similar physiological and psychological stress reactions to a social evaluative stressor will be found in vivo and in VR.

2. Methods

2.1. Participants and design

The experimental design comprised two between-subjects factors: *Strain* (stress or control) and *Reality* (VR or in vivo). An a priori calculation of required sample size for the two-way interaction resulted in a minimum of $N = 84$ for a power of $1\text{-}\beta = .95$ and an effect of $d = 0.8$—an effect size which can be expected in combinations of public speaking and cognitive tasks like the TSST (Dickerson and Kemeny, 2004).

Participants were recruited by on-campus advertisement and were included into the study if they had a BMI between 19 and 26 kg/m^2 and an age between 18 and 50 years. Further exclusion criteria were (a) acute or chronic somatic or psychiatric disease, (b) regular intake of medication, (c) psychotherapeutic treatment during the last year, (d) nicotine intake of more than five cigarettes per day, and (e) regularly working night shifts (Niu et al., 2011). Participants were asked to refrain from physical exercise and alcohol at least 24 h prior to testing and to refrain from consuming anything but water two hours prior. The study was approved by the ethics committee at the University of Trier and conducted in line with the Declaration of Helsinki. All participants gave informed written consent and were paid 30€ for their participation.

Ninety-three male participants ($M = 25.02$; $SD = 4.41$; age range: 19–45) enrolled for the study and were randomly assigned to one of the four conditions: Stress-VR ($n = 29$; age range: 20–45; $M = 24.93$; $SD = 4.63$), stress in vivo ($n = 21$; age range: 21–44; $M = 26.05$; $SD = 4.80$), control-VR ($n = 22$; age range: 18–33; $M = 22.82$; $SD = 3.72$), and control in vivo ($n = 21$; age range: 19–32; $M = 24.30$; $SD = 3.61$). A one-way ANOVA revealed no significant age differences between the groups ($F(3, 88) = 2.17$ $p = .097$, $\eta_p^2 = .07$). Five participants of the stress-VR group and one of the stress in vivo group had to be excluded due to technical errors in the VR procedure, resulting in a total sample of N = 87. Furthermore, due to technical errors, one person had to be excluded from the heart rate (HR) data analysis and another four people from the skin conductance level analysis.

2.2. Apparatus

The VR environment was generated using the Steam Source engine (Valve Corporation, Bellevue, Washington, USA) and controlled by the VR simulation software CyberSession 5.6 (VTPlus GmbH, Würzburg, Germany). A Head-Mounted Display (HMD; Oculus Rift DK2, Oculus VR LLC, Menlo Park, CA, USA) and headphones were used. Heart rate and skin conductance were monitored and recorded with Brain Vision Recorder (Version 1.20.0801, Brain Products GmbH, Gilching, Germany). Further technical specifications can be found in the supplementary methods published online with this article.

2.3. Measures

2.3.1. Saliva sampling and analysis

At seven time points throughout the experiment, participants were asked to give saliva samples, using Salivettes (Sarstedt, Nümbrecht, Germany), to determine salivary cortisol and alpha amylase (sAA) levels. After the experiment saliva samples were stored at – 20 °C until biochemical analysis was carried out by the University Laboratory. For details of biochemical analyses, see supplementary methods online.

2.3.2. Heart rate

Heart rate (HR) was recorded using a finger-pulse-plethysmograph (Becker Meditec, Karlsruhe, Germany). The sampling rate was 100 Hz. Brain Vision Analyzer (Version 2.1.1.964, Brain Products GmbH, Gilching, Germany) was used to export RR-intervals. ARTiiFACT (Kaufmann et al., 2011) was used to correct artifacts and export mean HR of the different experimental segments.

2.3.3. Skin conductance level

Skin conductance level (SCL) was recorded using two Ag/AgCl surface electrodes (Ø = 8 mm^2) that were covered with isotonic electrode gel and placed on the thenar and hypothenar area of the non-dominant palm (Dawson et al., 2016). The sampling rate was 100 Hz. Again, Brain Vision Analyzer was used to export SCL to Ledalab (Benedek and Kaernbach, 2010) which allowed conduction of artifact correction and exporting of mean SCL of the different experimental segments.

2.3.4. Subjective measures

At five time points, participants rated their subjective feelings of stress on visual analogue scales with a range of 0 (not at all) to 100 (very much) (cf. von Dawans et al., 2012).

2.4. Procedure

Experimental sessions were scheduled to start at 3.30 p.m. or 5.30 p.m. to control for the circadian rhythm of cortisol (Kudielka and Wüst, 2010). After giving informed consent, participants filled out the first subjective stress ratings (VAS1) and gave the first saliva sample (S1) before being lead into the VR laboratory. After application of the equipment for the physiological measurements, participants in the VR conditions put on the HMD and the TSST (or Placebo TSST (Het et al., 2009), began.

After the baseline measurements—that doubled as a period of acclimatization to the new situation—in either the real laboratory or the virtual environment (which was an exact replicate of the real laboratory), participants received instructions on the following task either by the experimenter or via headphones and written on the screen. In the stress conditions, they were told that they would have to do a job interview in front of a panel of judges who would shortly enter. In the control conditions, participants were told that they would have to talk about a self-chosen topic in an empty room. Both conditions were conducted in accordance with their respective original protocols (Kirschbaum et al., 1993), and (Het et al., 2009) although minor

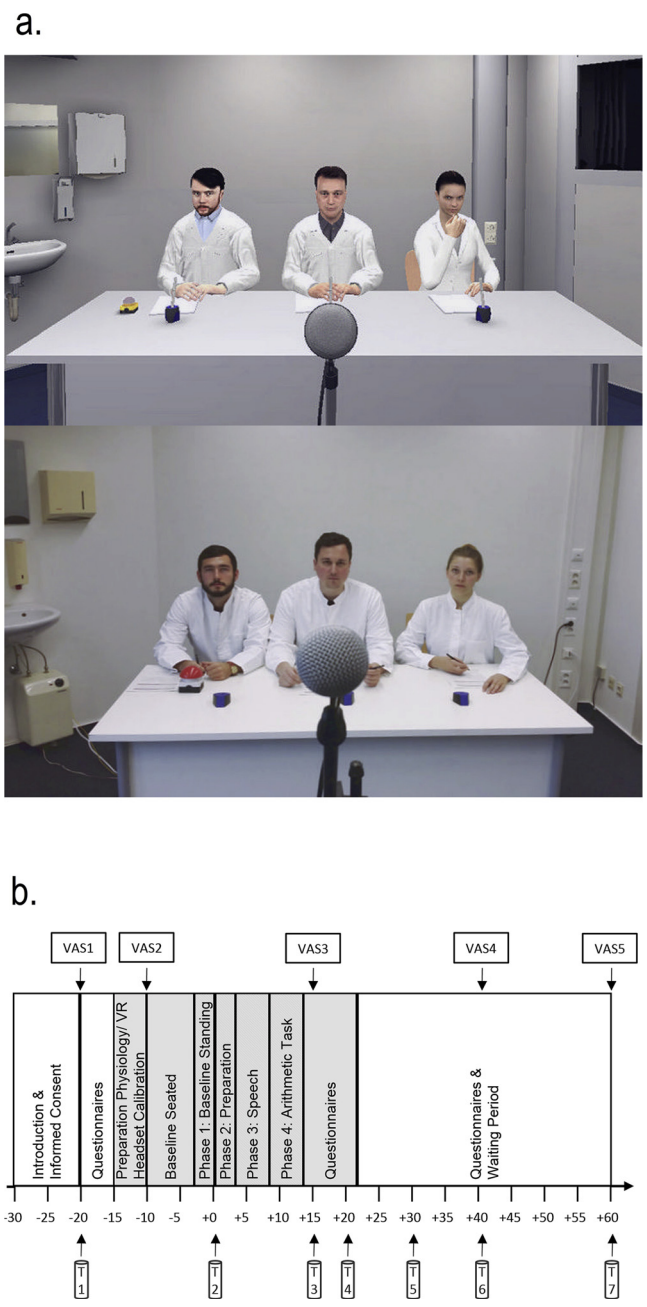

Fig. 1. (a) Picture of the VR (upper panel) and the in vivo (lower panel) judges in the stress conditions and (b) experimental procedure depicting experimental phases and time of assessment of subjective stress ratings (VAS) and saliva samples (S). Procedures in the preparation room are depicted in white; procedures in the VR laboratory have been marked in grey. Hatched patterns represent crucial phases of the (placebo) TSST procedure.

changes were made to facilitate the implementation into a virtual environment. These changes included a shorter preparation time of 3 min in front of the panel after they had been introduced to the task instead of preparing their speech for 10 min in another room alone. In addition, due to the virtual environment they were not able to take notes with paper and pencil. These changes were made to both TSST conditions.

The entire procedure was controlled by the experimenter behind a one-way mirror. Furthermore, prompts to maintain eye contact were automatically triggered in the TSST-VR after five seconds without eye-contact (i.e., not looking into a predefined area surrounding the judges' heads). After the task, the judges left the room (and the screen turned black in VR). The experimenter subsequently reentered the VR laboratory and participants in the VR conditions took off the HMD. All participants remained in the lab until 60 min after TSST started and provided five questionnaires with VAS and seven saliva samples in total. Following the last sample, participants were debriefed and compensated. Further details about the experimental procedure can be found in Fig. 1 or in the supplementary methods.

2.5. Statistics

Mixed repeated-measures ANOVAs were conducted to test for effects of stress condition (TSST vs. Control), experimental environment (VR vs. in vivo) and time over the course of the experiment (as a repeated-measures factor) on subjective and physiological measures. In cases where Mauchly's test indicated a violation of the assumption of sphericity, we used Greenhouse-Geisser correction and calculated ε-and corrected p-values. All analyses were conducted with SPSS for Windows (Version 24). Significance level was set at $p < .05$. Effect sizes are reported as η_p^2 with 95% Confidence Intervals. All pairwise comparisons were Bonferroni corrected.

3. Results

3.1. Free salivary cortisol

The cortisol responder rates to the different experimental conditions give a first indication of the success of the stress manipulation (see Table 1). To further evaluate whether the manipulation of stress was successful, we conducted two separate chi-square tests including the factor Strain and cortisol response—using the conservative criterion of a baseline-to-peak increase of 2.5 nmol/l—for the VR (χ^2 (1) = 3.81, $p = .051$) and in vivo condition (χ^2 (1) = 17.53, $p < .001$). In line with our hypothesis, the odds suggest that it is more likely to show a cortisol reaction to stress than to the control conditions in VR (3.3 times) and in vivo (40 times).

To follow up on these analyses, we conducted a 2 (Strain [stress, control]) x 2 (Reality [VR, in vivo]) x 7 (Time) repeated measures ANOVA (see Fig. 2, upper panels) which revealed significant main effects for Time ($F(6, 498) = 29.04$, $\varepsilon = .51$, $p < .001$, $\eta_p^2 = .26$, 95% CI [.19; .31]), Strain ($F(1, 83 = 13.36$, $\varepsilon = .51$, $p < .001$, $\eta_p^2 = .14$, 95% CI [.03; .28]), and Reality ($F(1, 83 = 4.10$, $\varepsilon = .51$, $p = .046$, $\eta_p^2 = .05$, 95% CI [.00; .16]). In line with our prediction, these main effects were qualified by the significant two-way interaction between the factors Time and Strain ($F(6, 498) = 13.94$, $\varepsilon = .51$, $p < .001$, $\eta_p^2 = .14$, 95% CI [.08; .19]), but also by the three-way interaction between the factors Time, Strain, and Reality ($F(6, 498) = 3.86$, $\varepsilon = .51$, $p = .01$, $\eta_p^2 = .04$, 95% CI [.01; .07]). Pairwise comparisons revealed that stress conditions differed from their corresponding control conditions (VR: +15 to + 60 min. post stress induction, all $ps < .016$; in vivo: +20 to + 60 min post stress induction, all $ps < .011$). Comparing in vivo and VR stress conditions using pairwise comparisons revealed significant effects at + 30 min and + 40 min (all $ps < .004$). All in all, and although cortisol rose and declined earlier in the VR than in the in vivo stress condition, the results indicate that the virtual TSST can activate the HPA-axis in a similar pattern.

Similar results were obtained when computing Area under the

Table 1
Cortisol Responder rates by conditions for a liberal and a conservative response criterion (1.5 vs. 2.5 nmol/l baseline-to-peak increase in free salivary cortisol).

	VR		In vivo	
	Stress	Control	Stress	Control
Response criterion	n (%)	n (%)	n (%)	n (%)
1.5 nmol/l increase	18 (75%)	9 (42.9%)	21 (100%)	10 (47.6%)
2.5 nmol/l increase	15 (62.5%)	7 (33.3%)	20 (95.2%)	7 (33.3%)

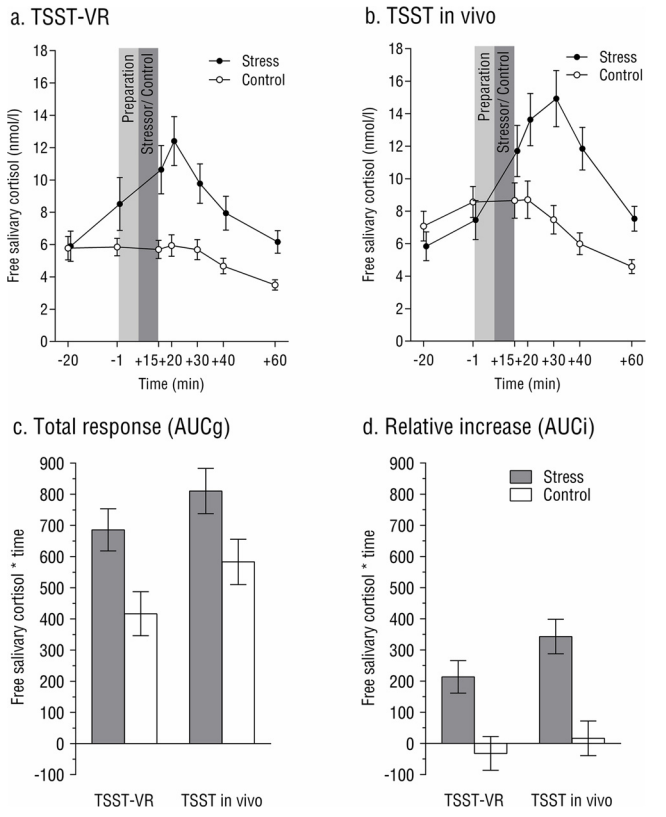

Fig. 2. Concentration of free salivary cortisol in response to the stress and control condition sampled at seven time points over the course of the experiment and as a function of the experimental conditions. (a) VR vs. (b) In vivo condition. (c) Area under the Curve with respect to ground (AUCg). (d) Area under the Curve with respect to increase (AUCi). Error bars denote standard errors.

Curve values using the formulas by (Pruessner et al., 2003) – see Fig. 2 lower panels. Conducting a 2 (Strain [stress, control]) x 2 (Reality [VR, in vivo]) ANOVA with area under the curve with respect to ground (AUCg) as the dependent variable, we found a significant main effect for Strain ($F(1, 83) = 11.70$, $p = .001$, $\eta_p^2 = .12$, 95% CI [.12; .26]) while the factor Reality missed significance ($F(1, 83) = 3.93$, $p = .051$, $\eta_p^2 = .05$). There was no significant Strain*Reality interaction ($F(1,83) = .06$, $p = .809$, $\eta_p^2 > .01$). Similar results were found for area under the curve with respect to increase (AUC_i). Again, a significant main effect for Strain was found ($F(1, 83) = 27.86$, $p < .001$, $\eta^2 = .25$, 95% CI [.10; .39]) while the main effect for Reality did not reach statistical significance ($F(1, 84) = 2.84$, $p = .096$, $\eta_p^2 = .03$). No significant Strain*Reality interaction ($F(1,83) = .46$, $p = .501$, $\eta_p^2 = .01$) was found. In sum, no differences between VR and in vivo stress conditions were found when using area under the curve values as an indicator of cortisol output in response to the experimental manipulation.

3.2. Salivary alpha amylase

As with cortisol, we conducted a 2 (Strain [stress, control]) x 2 (Reality [VR, in vivo]) x 7 (Time) repeated measures ANOVA (see Fig. 3) testing the effects on salivary alpha amylase. This ANOVA yielded a significant main effect of the factor Time ($F(6, 498) = 10.57$, $\varepsilon = .71$, $p < .001$, $\eta_p^2 = .11$, 95% CI [.06; .16]). The main effects of the factors Strain and Reality did not reach statistical significance ($F(1, 83) = 1.57$, $\varepsilon = .71$, $p = .21$, $\eta_p^2 = .02$ and $F(1, 83) = .03$, $\varepsilon = .71$, $p = .87$, $\eta_p^2 < .01$, respectively).

Furthermore, the two-way interaction between the factors Time and Strain ($F(6, 498) = 2.89$, $\varepsilon = .71$, $p = .020$, $\eta_p^2 = .03$, 95% CI [.00; .06]) reached statistical significance while the two-way interaction between Time and Reality ($F(6, 498) = .35$, $\varepsilon = .71$, $p = .857$, $\eta_p^2 < .01$) and the three-way interaction between Time, Strain, and Reality ($F(6, 498) = .08$, $\varepsilon = .71$, $p = .992$, $\eta_p^2 < .01$) did not. As predicted, further examination of the significant two-way interaction via pairwise comparisons revealed that the stress conditions differed significantly from the control conditions at +15 ($p = .037$) but not at any other time points (all $ps > .105$). Since the three-way interaction did not reach significance, we did not obtain evidence suggesting differences in the efficacy of in vivo and VR stressors concerning SAM-activation as measured by sAA concentration.

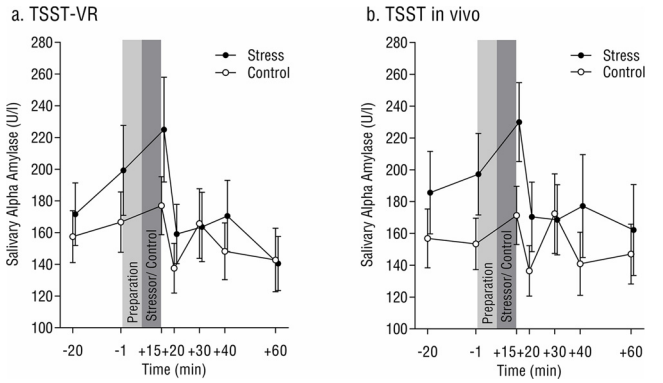

Fig. 3. Concentration of salivary alpha amylase (a) VR vs. (b) in vivo. Error bars denote standard errors.

3.3. Heart rate

To analyze another indicator of the SAM, we conducted a 2 (Strain) x 2 (Reality) x 4 (Time) repeated measures ANOVA using HR as a dependent variable (see Fig. 4). The four time points refer to the different phases of the TSST: Baseline measurements while standing, TSST preparation phase, TSST Interview, and TSST arithmetic task. This ANOVA yielded a significant main effect of the factor Time ($F(3, 246) = 112.40$, $\varepsilon = .72$, $p < .001$, $\eta_p^2 = .58$, 95% CI [.50; .63]) and Reality ($F(1, 82) = 4.39$, $\varepsilon = .72$, $p = .039$, $\eta_p^2 = .05$, 95% CI [.00; .17]). The main effect of the Factors Strain did not reach statistical significance ($F(1, 82) = .20$, $\varepsilon = .72$, $p = .656$, $\eta_p^2 > .01$).

Furthermore, the two-way interactions between the factors Time and Strain ($F(3, 246) = 8.47$, $\varepsilon = .72$, $p < .001$, $\eta_p^2 = .09$, 95% CI [.03; .16]) and Time and Reality ($F(3, 246) = 3.03$, $\varepsilon = .72$, $p = .047$, $\eta_p^2 = .04$, 95% CI [.00; .08]) reached statistical significance. These interaction effects were qualified by the significant three-way interaction

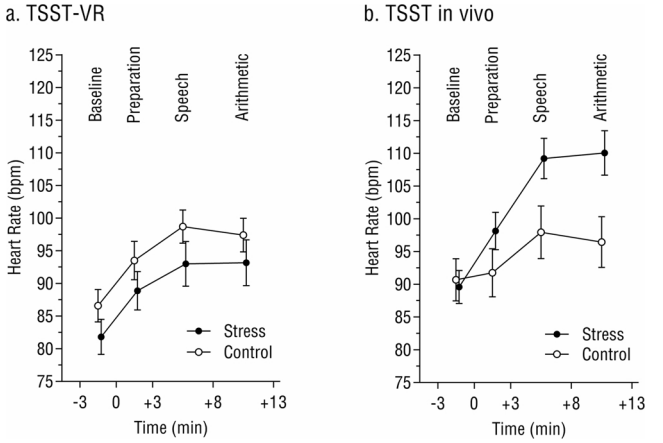

Fig. 4. Average heart rate during the four phases. (a) VR vs. (b) in vivo condition. Error bars denote standard errors.

between Time, Strain, and Reality ($F(3, 246) = 8.57$, $\varepsilon = .72$, $p < .001$, $\eta_p^2 = .10$, 95% CI [.03; .16]). Pairwise comparisons revealed that in the in vivo conditions, stress differed significantly from control during the TSST interview and arithmetic task ($p = .021$ and $p = .006$). In the VR conditions, no significant effects emerged at any of the TSST phases (all $ps > .222$). Comparing the respective VR and in vivo conditions, pairwise comparisons revealed significant differences between the stress conditions in all stages of the experimental procedure (all $ps < .046$); these differences in HR indicate that the in vivo stressor was more efficient in activating the SAM.

3.4. Skin conductance level

Analogous to the HR analysis, we conducted a 2 (Strain) x 2 (Reality) x 4 (Time) repeated measures ANOVA with SCL as the dependent variable. This ANOVA yielded only a significant main effect of the factor Time ($F(3, 237) = 10.41$, $\varepsilon = .69$, $p < .001$, $\eta_p^2 = .12$, 95% CI [.04; .19]), indicating a significant rise in SCL independent of conditions. None of the other the main effects nor the interactions reached statistical significance (all $Fs < 1.39$ and $ps > .251$).

3.5. Subjective measures

In accordance with previous research, we focused on the question "How stressed are you at the moment?" to analyze participant's subjective stress ratings (Shiban et al., 2016). We conducted a 2 (Strain [stress, control]) x 2 (Reality [VR, in vivo]) x 5 (Time) repeated measures ANOVA (see Fig. 5) that demonstrated a significant main effect of the factor Time ($F(4, 332) = 37.36$, $\varepsilon = .81$, $p < .001$, $\eta_p^2 = .31$, 95% CI [.23; .38]). Additionally, the two-way interaction between the factors Time and Strain ($F(4, 332) = 6.27$, $\varepsilon = .81$, $p < .001$, $\eta_p^2 = .07$, 95% CI [.02; .12]) reached significance. To examine this interaction, pairwise comparisons were used revealing the predicted data pattern; the stress conditions deviated from the control conditions at VAS3 (directly after the TSST, $p = .002$) but not at any other point in time (all $ps > .155$). No other effects or interactions were significant (all $Fs < 2.09$ and $ps > .152$). We therefore conclude that the VR and in vivo stress conditions are equally efficient in inducing subjective stress.

4. Discussion

The present study assessed whether a refined version of the TSST in VR poses a viable alternative to traditional face-to-face in vivo stress induction methods in the laboratory. Overall, the results suggest that subjective and physiological reactions to the VR and the in vivo version of the TSST were largely comparable. To the best of our knowledge, this is the first study to examine the effects of the TSST-VR and the TSST in vivo in a completely controlled experimental design with control groups in vivo and in VR. This orthogonal design enabled us to compare both versions of the TSST with their respective control group and thus assess the effect of the social stress induction independently. Taken together, we found similar patterns of results on most of our dependent subjective and physiological variables. While participants showed an increase of stress levels on almost all of our psychobiological stress markers, no such rise was observed in our control groups.

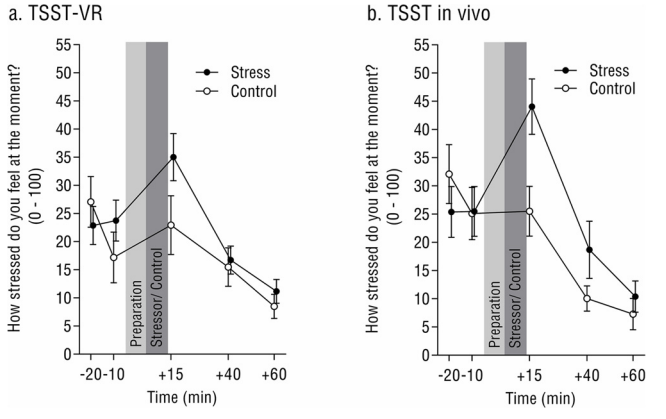

Fig. 5. Subjective ratings of stress (a) VR vs. (b) in vivo. Error bars denote standard errors.

Focusing on salivary cortisol as a major endocrine stress marker, the results are indeed consistent with the assumption that a robust and reliable stimulation of the HPA-axis with a social-evaluative stressor is possible in VR. With a 62.5% responder rate using the 2.5 nmol/l criterion and an on-average twofold increase, our results are comparable with the average cortisol reaction usually obtained with in vivo versions of the TSST (Goodman et al., 2017). The cortisol response to the present, improved version of the TSST-VR was more pronounced and more robust compared to previous studies that used variations of the TSST in VR (e.g., Kelly et al., 2007; Ruiz et al., 2010; Shiban et al., 2016). As stated above, this might be mainly due to rapid technological progress, especially regarding advanced graphics, which can be considered a main prerequisite for an increased immersion and the subjective feeling of presence in the virtual adaptation of the TSST. It is our understanding that the current technological status at the time of experimentation plays an essential role in this field of research. In the past, researchers often had to resort to VR headsets that are described as rather clunky and uncomfortable, (e.g. Kelly et al., 2007 refer to their headset as a helmet with a small viewing screen) and although technical aspects like weight, resolution, or viewing angle are often not reported, it can be assumed that these devices might have hindered the participants from experiencing the degree of presence that modern HMDs achieve. This assumption is largely supported by the fact that studies using a CAVE system to realize their VR conditions (an immersive stereoscopic room in which images are projected onto the walls and participants wear specifically designed glasses instead of HMDs) tend to report large stress effects on physiological markers (Jönsson et al., 2010; Fich et al., 2014). In a direct comparison of both modalities of presentation, Juan and Pérez (2009) found that exposure therapy provoked more anxiety and a higher sense of presence in acrophobic patients when it was conducted in a CAVE system than with an HMD. Since that time, however, technical progression and the introduction of virtual reality headsets to a wider audience via the medium of video games provided researchers with light, relatively comfortable HMDs with high resolution displays and effective motion tracking mechanisms. It would thus be quite informative to experiment with both modes of presentation with state of the art technology in order to elucidate whether there are still significant differences in effectiveness.

In addition, we specifically aimed at maximizing comparability between the different conditions by carefully emulating the in vivo surroundings in the virtual environment and increasing interactivity by introducing automated eye-tracking-based verbal feedback when the participants did not maintain eye-contact with the agents. Both factors—the sophistication of the graphical presentation and the high level of perceived interactivity—might have promoted immersion and presence and thus contributed to the comparability of psychobiological stress responses in the stress conditions (Diemer et al., 2015).

Beyond examining cortisol responses, the comparability of psychological stress reactions in the virtual condition and in vivo can be shown by the rise in sAA—a valid index of sympathetic activation (Nater and Rohleder, 2009) —and the increase in subjective stress ratings. On these measures, the TSST-VR elicited stress responses that were equally high in the VR and the in vivo setting. This supports the conclusion that stress induction paradigms in a virtual environment, such as the TSST-VR, can be potent reflections of a stressful situation in reality.

Moreover, the orthogonal experimental design permits us to infer that the observed stress reactions in the TSST-VR were indeed elicited by the stressful characteristics of the task itself and not by the fact that

it was performed in an unfamiliar artificial environment. As assessed via self-report at the end of the experiment, almost all of our participants reported little or no previous experience with immersive virtual reality technology. It is therefore conceivable that the novelty of being immersed in a virtual environment that is entirely under the control of the experimenter might be sufficient to make participants feel uncomfortable and thus induce stress. This alternative explanation for the observed stress effects in previous studies (Jönsson et al., 2010; Fich et al., 2014) cannot be discarded without the implementation of a control group in VR. In the present study, the comparison between the TSST-VR and the control condition in VR showed differential patterns of reactions. The fact that we found stress effects on our dependent measures only in the stress condition suggests that the immersion in a foreign virtual environment is not sufficient to elicit a stress response, at least in terms of HPA-axis activation.

It should be noted, however, that cortisol and HR responses to the in vivo stress condition were still slightly more pronounced than to the VR stress condition, whereas the SCL response was not affected differently in the conditions. Possible explanations for these findings may be that the overall rise in heart rate in the virtual control condition might be attributed to increased activation caused by the speaking task and anticipatory arousal due to the unfamiliar virtual environment, and the low reactivity in SCL to constraints of the measurement (e.g., very sweaty palms). It may, however, also be the case that virtual adaptations of the real world—although potent reflections of many aspects of real situation—are still limited by technological restraints which lead to slightly attenuated psychobiological reactions to these environments. Furthermore and more specifically, the TSST might be especially difficult to replicate in a virtual environment because of its conceptualization as a stressor that uses a performance situation in the presence of unapproachable human judges to generate social evaluative stress. These necessary characteristics—evaluation and negative feedback by human experts and uncontrollability of the situation (Dickerson and Kemeny, 2004)—should make the translation into virtual reality difficult, since participants will still be able to envision that they are not actually performing in front of real human beings but programmed entities. Nevertheless, previous studies and the present findings suggest that a majority of participants still adhere to social conventions (Garau et al., 2005) and experience social evaluative stress in the presence of virtual agents, as indicated by the subjective and endocrine reactions (Jönsson et al., 2010; Kothgassner et al., 2016; Montero-López et al., 2016; Shiban et al., 2016).

As mentioned in the methods section, the implementation of the TSST into virtual reality and the parallelization of the experimental conditions required some alterations to the original study protocol (Kirschbaum et al., 1993) mainly in the preparation phase. A recent meta-analysis on protocol variations of the TSST has, however, shown that the stress induction effect is quite robust against a variety of changes that have been made to the paradigm over the years of its application (Goodman et al., 2017). The substantial stress effects that we report in both TSST conditions in our study seem to further support the idea that the strict adherence to the original protocol might not be a necessary precondition for successful stress induction as long as the main stressful features, social threat and uncontrollability are realized (Dickerson and Kemeny, 2004). It might be an interesting question for future studies, whether the TSST protocol can be generally simplified without reducing its stressfulness.

Some potential limitations of the present experiment should be noted. As in many fundamental studies on endocrine stress reactivity, we started by examining an exclusively male sample of participants. Besides the fact that men and women differ in their endocrine profiles and reactivity to social stress (Kudielka and Kirschbaum, 2005; Kelly et al., 2008), some studies show a differential effect of gender on the perception of virtual environments (Munafo et al., 2017), especially regarding Sense of Presence (Felnhofer et al., 2012). Although the widely assumed concept that men and women differ in their affinity to video gaming and virtual environments in general is slowly being disproved (Rehbein et al., 2016), video games still occupy a larger role in men's free time than in women's (Borgonovi, 2016). Secondly, we assessed the endocrine stress responses by using salivary measures of cortisol and alpha amylase. Although these measures have been proven valid indicators of HPA-axis and catecholaminergic stress reactivity (Hellhammer et al., 2009; Nater and Rohleder, 2009), direct measures of ACTH, cortisol and catecholamines in plasma would have possibly been more sensitive in the assessment of subtle differences between the VR and in vivo version of the TSST. Lastly, it should be noted that although the participant was alone in the room during all VR procedures, the experimenter was in the adjacent room behind a one-way mirror and supervised the experimental sessions and controlled the agents' reactions to the participants' performance. Moreover, the necessity of taking a saliva sample right before the start of the task required the experimenter to re-enter the room and hand the participant the Salivette. In the VR groups, this was done while the participants were wearing the headset so that they consequently saw neither the experimenter nor their own hands while chewing the cotton swab. We therefore cannot rule out that the participants were, to some extent, aware of the experimenter's presence. Thus, the feeling of being socially evaluated might not have been exclusively conveyed by the virtual agents, but to some degree also by the experimenter. Future studies should evaluate the influence of the experimenter's presence on immersion and presence in the virtual reality.

5. Conclusion

Taken together, the present study demonstrates that social evaluative stress can be successfully induced in a virtual environment resulting in stress responses on several physiological measures associated with the HPA axis and the SAM system. By using a refined VR version of the TSST, we could show that situations realized in VR have the potential to realistically simulate complex social interactions and evoke comparable subjective and physiological reactions. Due to its computer-generated nature, the TSST-VR has several key advantages: First, it is entirely standardized with no variation between testing sessions. Secondly, it is very economic insofar as it reduces the necessary amount of personnel from at least three to one and makes training judges obsolete. Lastly, it facilitates the variation of parameters of interest. In sum, the present study demonstrates that a technologically sophisticated version of the TSST-VR that maximizes interactivity and presence might be a valuable alternative to the traditional in vivo stress induction for experiments in psychoneuroendocrinology.

Author contributions

GD, EW and PZ designed the study and drew up the study protocol. PZ, BB and GH performed the experiments. GH supervised the procedures in the VR laboratory and adjusted the VR software to the requirements of the experiment. BB, GD and PZ performed the statistical analyses and wrote the first draft of the manuscript. All authors contributed to writing and have approved the final manuscript.

Declaration of conflicting interests

The authors declare no conflicts of interest with respect to the authorship or the publication of this article.

Acknowledgements

This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), [grant number INST 246/124-1 FUGG awarded to E.W.] and the University Research Priority Program "Psychobiology of Stress" funded by the State Rhineland-Palatinate. We thank Konstantin Hiesinger for his help in programming

the virtual environments. We are also grateful to Lisa Born, Michael Kalnitskyy, Leonie Matter, Simon Manderfeld, Nina Mendes-Pereira, William Standard, and Julia Strojny for helping to conduct the experiments. Furthermore, we wish to thank Dr. Carolyn Wu for language editing.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.psyneuen.2018.11.010.

References

Baños, R.M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., Rey, B., 2004. Immersion and emotion: their impact on the sense of presence. Cyberpsychol. Behav. 7, 734–741.

Benedek, M., Kaernbach, C., 2010. A continuous measure of phasic electrodermal activity. J. Neurosci. Methods 190, 80–91.

Borgonovi, F., 2016. Video gaming and gender differences in digital and printed reading performance among 15-year-olds students in 26 countries. J. Adolesc. 48, 45–61. https://doi.org/10.1016/j.adolescence.2016.01.004.

Dawson, M.E., Schell, A.M., Filion, D.L., 2016. The electrodermal system. In: Berntson, G.G., Cacioppo, J.T., Tassinary, L.G. (Eds.), Handbook of Psychophysiology, Cambridge Handbooks in Psychology. Cambridge University Press, Cambridge pp. 217–243.

Dickerson, S.S., Kemeny, M.E., 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol. Bull. 130, 355–391.

Diemer, J., Alpers, G.W., Peperkorn, H.M., Shiban, Y., Mühlberger, A., 2015. The impact of perception and presence on emotional reactions: a review of research in virtual reality. Front. Psychol. 6.

Felnhofer, A., Kothgassner, O.D., Beutl, L., Hlavacs, H., Kryspin-Exner, I., 2012. Is virtual reality made for men only? Exploring gender differences in the sense of presence. Proc. Int. Soc. Presence Res. 103–112.

Fich, L.B., Jönsson, P., Kirkegaard, P.H., Wallergård, M., Garde, A.H., Hansen, Å., 2014. Can architectural design alter the physiological reaction to psychosocial stress? A virtual TSST experiment. Physiol. Behav. 135, 91–97.

Garau, M., Slater, M., Pertaub, D.-P., Razzaque, S., 2005. The responses of people to virtual humans in an immersive virtual environment. Presence Teleoper. Virtual Environ. 14, 104–116.

Goodman, W.K., Janson, J., Wolf, J.M., 2017. Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. Psychoneuroendocrinology 80, 26–35.

Hellhammer, D.H., Wüst, S., Kudielka, B.M., 2009. Salivary cortisol as a biomarker in stress research. Psychoneuroendocrinology 34, 163–171. https://doi.org/10.1016/j.psyneuen.2008.10.026.

Het, S., Rohleder, N., Schoofs, D., Kirschbaum, C., Wolf, O.T., 2009. Neuroendocrine and psychometric evaluation of a placebo version of the 'Trier Social Stress Test'. Psychoneuroendocrinology 34, 1075–1086.

Jönsson, P., Wallergård, M., Österberg, K., Hansen, Å.M., Johansson, G., Karlson, B., 2010. Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the Trier Social Stress Test: a pilot study. Psychoneuroendocrinology 35, 1397–1403.

Juan, M.C., Pérez, D., 2009. Comparison of the levels of presence and anxiety in an acrophobic environment viewed via HMD or CAVE. Presence Teleoperators Virtual Environ. 18, 232–248. https://doi.org/10.1162/pres.18.3.232.

Kaufmann, T., Sütterlin, S., Schulz, S.M., Vögele, C., 2011. ARTiiFACT: a tool for heart rate artifact processing and heart rate variability analysis. Behav. Res. Methods 43, 1161–1170.

Kelly, M.M., Tyrka, A.R., Anderson, G.M., Price, L.H., Carpenter, L.L., 2008. Sex differences in emotional and physiological responses to the Trier Social Stress Test. J. Behav. Ther. Exp. Psychiatry 39, 87–98. https://doi.org/10.1016/j.jbtep.2007.02.003.

Kelly, O., Matheson, K., Martinez, A., Merali, Z., Anisman, H., 2007. Psychosocial stress evoked by a virtual audience: relation to neuroendocrine activity. Cyberpsychol. Behav. 10, 655–662.

Kirschbaum, C., Pirke, K.-M., Hellhammer, D.H., 1993. The 'Trier Social Stress Test' – a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 28, 76–81.

Kothgassner, O.D., Felnhofer, A., Hlavacs, H., Beutl, L., Palme, R., Kryspin-Exner, I., Glenk, L.M., 2016. Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment. Comput. Hum. Behav. 62, 124–135.

Kotlyar, M., Donahue, C., Thuras, P., Kushner, M.G., O'Gorman, N., Smith, E.A., Adson, D.E., 2008. Physiological response to a speech stressor presented in a virtual reality environment. Psychophysiology 45, 1034–1037.

Kudielka, B.M., Kirschbaum, C., 2005. Sex differences in HPA axis responses to stress: a review. Biol. Psychol. 69, 113–132. https://doi.org/10.1016/j.biopsycho.2004.11.009.

Kudielka, B.M., Wüst, S., 2010. Human models in acute and chronic stress: assessing determinants of individual hypothalamus–pituitary–adrenal axis activity and reactivity. Stress 13, 1–14.

Miller, R., Plessow, F., Kirschbaum, C., Stalder, T., 2013. Classification criteria for distinguishing cortisol responders from nonresponders to psychosocial stress: evaluation of salivary cortisol pulse detection in panel designs. Psychosom. Med. 75, 832–840.

Montero-López, E., Santos-Ruiz, A., García-Ríos, M.C., Rodríguez-Blázquez, R., Pérez-García, M., Peralta-Ramírez, M.I., 2016. A virtual reality approach to the Trier Social Stress Test: contrasting two distinct protocols. Behav. Res. Methods 48, 223–232.

Munafo, J., Diedrick, M., Stoffregen, T.A., 2017. The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. Exp. Brain Res. 235, 889–901. https://doi.org/10.1007/s00221-016-4846-7.

Nater, U.M., Rohleder, N., 2009. Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: current state of research. Psychoneuroendocrinology 34, 486–496.

Niu, S.-F., Chung, M.-H., Chen, C.-H., Hegney, D., O'Brien, A., Chou, K.-R., 2011. The effect of shift rotation on employee cortisol profile, sleep quality, fatigue, and attention level: a systematic review. J. Nurs. Res. 19, 68–81.

Pruessner, J.C., Kirschbaum, C., Meinlschmid, G., Hellhammer, D.H., 2003. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. Psychoneuroendocrinology 28, 916–931.

Rehbein, F., Staudt, A., Hanslmaier, M., Kliem, S., 2016. Video game playing in the general adult population of Germany: can higher gaming time of males be explained by gender specific genre preferences? Comput. Hum. Behav. 55, 729–735. https://doi.org/10.1016/j.chb.2015.10.016.

Ruiz, A.S., Peralta-Ramirez, M.I., Garcia-Rios, M.C., Muñoz, M.A., Navarrete-Navarrete, N., Blazquez-Ortiz, A., 2010. Adaptation of the trier social stress test to virtual reality: psycho-physiological and neuroendocrine modulation. J. Cyber Ther. Rehabil. 3 (4).

Schwabe, L., Höffken, O., Tegenthoff, M., Wolf, O.T., 2013. Stress-induced enhancement of response inhibition depends on mineralocorticoid receptor activation. Psychoneuroendocrinology 38, 2319–2326.

Shiban, Y., Diemer, J., Brandl, S., Zack, R., Mühlberger, A., Wüst, S., 2016. Trier Social Stress Test in vivo and in virtual reality: dissociation of response domains. Int. J. Psychophysiol. 110, 47–55.

Smeets, T., Cornelisse, S., Quaedflieg, C.W.E.M., Meyer, T., Jelicic, M., Merckelbach, H., 2012. Introducing the Maastricht Acute Stress Test (MAST): a quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. Psychoneuroendocrinology 37, 1998–2008.

von Dawans, B., Fischbacher, U., Kirschbaum, C., Fehr, E., Heinrichs, M., 2012. The social dimension of stress reactivity: acute stress increases prosocial behavior in humans. Psychol. Sci. 23, 651–660.

Wallergård, M., Jönsson, P., Johansson, G., Karlson, B., 2011. A virtual reality version of the trier social stress test: a pilot study. Presence Teleoper. Virtual Environ. 20, 325–336.