The short Sing-a-Song Stress Test: A practical and valid test of autonomic responses induced by social-evaluative stress

D.J. van der Mee,⁎, Q. Duivestein, M.J. Gevonden, J.H.D.M. Westerink, E.J.C. de Geus

Department of Biological Psychology, VU University Amsterdam, Amsterdam, the Netherlands
Eindhoven University of Technology, Eindhoven, the Netherlands
Philips Research, Eindhoven, the Netherlands

ARTICLE INFO

Keywords:
Sing-a-Song Stress Test
Social-evaluative stress
Affect
Parasympathetic nervous system
Sympathetic nervous system

ABSTRACT

The Sing-a-Song Stress Test (SSST) was recently developed as an alternative to the Trier Social Stress Test (TSST) to investigate autonomic nervous system responses to social-evaluative stress. In the SSST, participants are suddenly cued to sing a song in the presence of confederates. However, the SSST is still quite long (~15 min) and the requirement for confederates makes it labor-intensive. The current study tested whether a shorter (~6.5 min), single-experimenter, version of the SSST can still reliably elicit subjective and physiological stress reactivity.

Our sample consisted of 87 healthy young adult participants (age range: 18–35 years). During the short SSST and a speeded reaction time task, in which aversive loud tones were to be avoided (TA), we measured heart period (HP), sympathetic nervous system (SNS) activity using pre-ejection-period (PEP), skin conductance level (SCL), and non-specific skin conductance responses (ns.SCR), and parasympathetic nervous system (PNS) activity using respiratory-sinus-arrhythmia (RSA) and the root-mean-square of successive differences (RMSSD).

The short SSST induced significant decreases in positive affect and increases in negative affect. MANOVA on the clusters of SNS and PNS variables showed that the short SSST elicited significant HP (~118.46 ms), PEP (~7.76 ms), SCL (~4.85 μS), ns.SCR (~8.42 peaks/min) and RMSSD (~14.67) reactivity. Affective, SNS, and PNS reactivity to the new SSST social-evaluative stress task were of comparable magnitude to that evoked by the TA mental stressor.

We conclude that the short SSST is a valid and cost-effective task for large scaled studies to induce social-evaluative stress to a sufficient degree to evoke measurable changes in PNS and SNS activity and affective state.

1. Introduction

The importance of having valid tests for different kind of stressors originates from the idea that different stressors lead to different stress responses within individuals, known as the response specificity theory (Bosch et al., 2009; Dickerson and Kemeny, 2004; Skoluda et al., 2015). This theory is grounded in the belief that different physiological response patterns have evolved to effectively cope with the variety of different stressors (Weiner, 1992). Such stressor can be broadly, but not exclusively, divided in two categories: those focusing mainly on the mental effort or challenge-appraisal component and those focusing on the social self-as described by the Social Self Preservation Theory (Dickerson et al., 2004) referred to as social-evaluative stressors. Many tasks have been developed to study the different aspects of stress induced by “mental effort” type, like the Paced Auditory Serial Addition Test (Gronwall, 1977; Tombaugh, 2006; Tanosoto et al., 2012), Stroop test (Stroop, 1935; Renaud and Blondin, 1997; van Lien et al., 2013) or aversive speeded reaction time tasks (Geus et al., 1993). However, for eliciting “social-evaluative” stress there is currently only one single default paradigm: the Trier Social Stress Test (TSST), where participants prepare to give a speech in front of a panel of judges (Kirschbaum et al., 1993).

While the TSST reliably evokes significant social-evaluative stress, it was primarily designed to induce prolonged stress to measure the slow responding adrenocortical axis reactivity. To investigate autonomic nervous system (ANS) reactivity to social-evaluative stress (e.g. through cardiac and electrodermal activity (EDA) responses), which are much faster, a shorter test that is also easier to implement would be preferred.

To this purpose, the Sing-a-Song Stress Test (SSST) was developed by Brouwer and Hogervorst (2014). In this test, participants are
suddenly cued to sing a song in the presence of confederates, which elicits social-evaluative stress comparable to TSST. However, the SSST is still quite long (~15 min) and the use of multiple confederates makes it labor-intensive. This makes it impractical and costly for studies in large samples. Therefore, we developed a shorter (~6.5 min), single-experimenter version of the SSST. The following adaptations were made to the original SSST (Brouwer and Hogervorst, 2014). First, the read-only conditions were reduced from nine to three. Second, we added a ‘practice’ condition in which they received an instruction to say the word vacuum out loud twice in short succession after an anticipatory countdown period. This condition was added to counteract the risk of ‘too early singing’ which was found in a substantial number of participants in the original SSST (Brouwer and Hogervorst, 2014). Lastly, no physical presence of a confederate was involved, but instead the crucial instruction to prepare to sing a song explicitly mentioned that the audio and video recordings would be shared with an audience of music professionals interested in variation in musical ability. This short version of the SSST (short SSST) can easily be implemented in large-scale epidemiological studies on the effects of stress on health outcomes.

To investigate if the short SSST can be used as an index of social-evaluative stress, positive and negative affect were measured before and after the short SSST to gauge the subjective stress response. To study the effect of social-evaluative stress on ANS activity we studied both sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) activity, in line with Brouwer and Hogervorst (2014), but expanding the set of ANS measures. SNS activity was measured as the pre-ejection period (PEP; Matyas and King, 1976), skin conductance level (SCL; Boucsein, 2012) and non-specific skin conductance responses (ns.SCRs; Boucsein, 2012). PNS activity was measured as respiratory sinus arrhythmia (RSA, Grossman et al., 1990) and root mean square of successive differences (RMSSD, Goedhart et al., 2007). Lastly, heart period (HP), a mixture of both SNS and PNS activity, was measured. To validate the stress component we compared the subjective and ANS stress reactivity of participantsto this new social-evaluative stressor to that of an often employed mental stressor, the tone avoidance (TA) speeded reaction time task.

We expect the short SSST to decrease positive affect and increase negative affect. Concerning ANS reactivity we expect increased SNS activity and decreased PNS activity reflected in an increase in SCL and ns.SCRs and a decrease in HP, PEP, RSA and RMSSD. The effects sizes are expected to be at least as big as those generated by the TA mental stress task.

2. Participants

A total of 113 participants participated in the study (age range 18–57). Exclusion criteria were a body-mass index above 30, heart disease, high blood pressure, high cholesterol, diabetes, and thyroid or liver disease, as these can all influence the functioning of the ANS. Additional exclusion criteria were the use of antidepressants or any other medication that has been shown to influence the ANS. If applicable, female participants were measured in the first two weeks after the last day of their menstrual cycle.

Due to the lack of applicants over the age of 35 (N = 5) we decided to exclude these participants in the current study. Of the remaining 108 participants, 11 were excluded because they were obese (N = 3) or had high blood pressure (N = 11) (some participants met multiple exclusion criteria). An additional 10 participants were excluded because they did not sing a song.

Participants who were students at the VU University of Amsterdam received research credits, while the other non-student participants were compensated with a €50 gift voucher. All participants provided informed consent before the start of the experiment. The study was approved by the VUmc medical ethical committee (NL62442.029.17).

3. Materials

3.1. The short Sing-a-Song Test

Participants were told that they had to sit as still as possible in front of a computer while they were shown several messages, followed by a counter from 60 to 0 s. They were informed that some of these messages only needed to be read whereas others might contain an instruction that they had to follow when the counter had reached 0. A detailed description of the experiment is given in Fig. 1.

For the three read-only trials, participants were instructed on-screen to quietly read the presented messages while sitting as still as possible. It was important to select phrases that did not elicit any stress or emotions. Therefore, three phrases in big black letters from the Dutch Wikipedia site about vacuums were shown on a monitor with a white background (translated example: “A vacuum is a device that sucks dust and other small particles”), similar to the original SSST (Brouwer and Hogervorst, 2014). The neutral Wikipedia phrases and read instruction were shown for 12 s, followed by a counter counting down from 60 s to

![Fig. 1. Short Sing-a-Song stress test experimental set-up.](image-url)

The task consisted of three read conditions, a speak condition and a sing condition. Each condition was followed by a countdown from 60 to 0 s. The first two messages were neutral text with no instruction. The third message contained the instruction to say the word “vacuum” twice when the timer reached 0. This was followed by another message with neutral text. Lastly an instruction to sing a song when the timer reached 0 was shown. In our analyses we focused on the anticipatory stress during the sing-a-song countdown, which is provided with a bold outline.
0 s. The instruction to read out aloud twice the word “vacuum cleaner” after countdown also lasted 12 s, followed by a counter from 60 s to 0 s and a 5 s period in which the word “vacuum cleaner” was shown on the screen. In the final sing-a-song trial, an instruction was provided for 12 s telling them to pick a song of their own choice and prepare to sing it aloud after the counter reached zero. It was also stated that their performance would be recorded and investigated by musical professionals. The short SSST ended with the instruction to sing a song that lasted for 20 s.

3.2. Tone avoidance test

The tone avoidance test is a stress-inducing task of the “active coping” type. During the tone avoidance task subjects have to react to a stimulus (an “X”) that flares up irregularly in one of the corners of a computer screen. Subjects have to respond to this stimulus, within a 550 ms timeframe, by pressing the button opposite to this corner on their response panel using one hand only.

Participants started with 50 points. During the task, incorrect or too slow responses were punished with a red bar, a loud noise burst (1000 Hz, 85 dB) and a loss of 1 point. Correct responses were rewarded by a green bar (Benschop and Schedlowski, 1999). When participants responded correctly for five consecutive times or more a point was added. Participants were told that they had to sit as still as possible during the test, only moving the hand they use for button pressing.

3.3. Affect questionnaire

Positive affect scores were obtained before and after each test by asking the participants to rate on a scale of 1 (not at all) to 7 (very) whether they felt relaxed, cheerful, enthusiastic, and content. Negative affect was obtained from items rating whether they felt insecure, lonely, anxious, irritated, and down (Myin-Germey et al., 2001). Positive and negative affect were then defined as the mean score of the individual items.

3.4. Physiology

To measure electrocardiography (ECG), impedance cardiography (ICG), and EDA, the VU-Ambulatory Monitoring System (VU-AMS) (Vrije Universiteit, Amsterdam, the Netherlands) was used. To record ECG and ICG signals, five adhesive 55 mm Kendall H98SG hydrogel ECG electrodes (Medtronic, Heerlen, Netherlands) were placed on the subject’s torso (Fig. 2). A recording frequency of 1000 Hz was used.

EDA was recorded on the participant's non-dominant hand. No preparations were performed on the skin to preserve its electrical properties (Dawson et al., 2000). A Biopac Systems ELS07 EDA isotonic gel electrode (Biopac Systems Inc., Goleta, US) was placed on the thenar eminence of the non-dominant hand (Dawson et al., 2000). A 55 mm Kendall H98SG hydrogel ECG electrode (Medtronic, Eindhoven, Netherlands) was placed on the inside of the non-dominant forearm approximately 15 cm below the hand electrode. Before applying this electrode, dead skin cells were removed by lightly scrubbing the skin with sandpaper. A recording frequency of 10 Hz was used.

4. Methods

4.1. Procedure

This study was part of a larger study that focuses on the validation of a wristwatch based technology, developed by Philips (Eindhoven, Netherlands), to measure EDA in a laboratory and ambulatory settings (see Appendix 1 for a complete overview of the study). The larger experiment across two days was presented to the participants as a general study on the detection of stress through measurement of ANS activity using wearable technology.

When entering the lab, participants were informed that their voice, facial expressions and posture would be recorded by video. Participants were shown the control room that the experimenter would sit in. It had a one-way mirror overlooking the experimental room where they would undergo the various tests. The control room contained multiple monitors and speakers that generate high-quality video footage and voice recordings from the camera and microphone placed in the experimental room. The participants were made aware of this intense monitoring throughout the experiment. They were not told upfront that the tests would involve singing nor that the recordings of their performance would be shared with an unseen audience. Throughout, no actual footage or sound was recorded and the deliberate deception about being recorded as well as its purpose was explained in the debriefing at the end of the experiment.

At the start of the experiment on day 1, resting blood pressure and body-mass index (BMI) were measured followed by a structured interview regarding the subject’s demographics, medication use, perceived physical and mental health and lifestyle behaviors, to confirm that...
participants met inclusion criteria for the study. Next, the system for continuous monitoring of SNS and PNS activity was attached.

The experimental stress manipulations on day 2 consisted of a baseline measure, in which participants were instructed to sit as still as possible for 3 min, followed by the 4 min TA task and the ~6.5 min short SSST. In between the TA task and the short SSST, participants had a two-minute recovery period. Immediately after the baseline and after both stressors participants were instructed to fill in a short 9-item questionnaire to measure their negative and positive affect (Myin-Germeyns et al., 2001).

4.2. ECG and ICG derived PEP and heart period variability measures

ECG and ICG analysis were performed using the VU-DAMS (Vrije Universiteit, Amsterdam, Netherlands) software (version 4.0). The software detects and scores all R peaks and automatically detects the start of inspiration and expiration for each breath. Possible heart period (HP) artifacts were marked by the software and visual inspection was used to remove or correct artifacts (i.e. wrongly scored R peaks). RSA was obtained by peak-valley estimation as described elsewhere (Nederend et al., 2018) combining the HP time series with the respiration signal that was extracted from the lower frequency changes in thorax impedance. RSA values were set to be zero for breaths with an invalid RSA. RMSSD was calculated by taking the root mean square of successive differences in heart period. Quality of RSA and RMSSD was checked by inspecting the respiration and heart rate signal manually, removing noisy data when necessary. For each condition, an ensemble average impedance cardiomgram of all corresponding complexes of adequate quality was calculated using the VU-DAMS software. Given its sensitivity for movement artifacts, the ICG signal was filtered using a 60 Hz low pass filter. Each impedance cardiomgram was inspected visually, and the B, C and X points were scored automatically and manually corrected when necessary (Nederend et al., 2017). PEP was obtained by calculating the time between the start of ventricular depolarization in the ECG (Q onset) and the time the aortic valve opens in the impedance cardiomgram (B point). PEP has been shown to be a reliable non-intrusive way to measure SNS activity (Sherwood et al., 1990).

4.3. EDA derived SCL and ns.SCR measures

All EDA signals were cleaned with a simple automated artifact rejection algorithm (i.e. sudden drastic drops or increases in μS, flattening of the signal) in MATLAB (2016a). SCL and ns.SCRs per condition were obtained using the EDA master toolkit (Joffily, 2012) in MATLAB (2016a). The SC signal was filtered using a low-pass 0.5 Hz Butterworth filter (Taylor et al., 2015). SCL was calculated as the average over the artifact-free, filtered signals. A ns.SCR was identified when the peak amplitude exceeded 0.01 μS but was not larger than 2.5 μS and the rise time was between 0.1 and 5 ms. Overlapping responses (a ns.SCR that occurs during the rise time of a preceding ns.SCR) were counted to detect stacking of responses. The total number of identified responses was divided by the artifact free time to obtain ns.SCR frequency per minute.

4.4. Statistical analysis

Data were analyzed using SPSS (ver. 25.0, 2017). For the analyses, the mean of all ANS measures during the 3 min baseline, 4 min TA task and 60 s short SSST sing anticipation was used. All ANS variables were checked for normal distribution and outliers. If a variable was not normally distributed, it was log-transformed. A value was considered an outlier if it deviated from the mean with more than three standard deviations. All outlier values were removed. Concerning to SCL, values over 35 μS were deemed implausible and therefore censored at 35 μS.

We expected all neutral anticipatory conditions to be different from the SSST anticipation but had no reason to expect differences between the neutral anticipatory conditions and the baseline or between the neutral anticipatory conditions themselves. This was borne out by preliminary comparisons. Therefore, our analyses were simplified to physiological reactivity of the short SSST by focusing on the contrast between the sitting baseline condition and a) the Sing anticipation condition and b) the TA task.

To investigate the effect of the two stressors on ANS activity, a repeated measures MANOVA on the clusters of SNS and PNS variables was performed with type of stressor (short SSST vs TA) and condition (baseline vs stress exposure) as the repeated measures. The multivariate cluster of SNS variables included HP, PEP, SCL and ns.SCR. The multivariate cluster of PNS measures included HP, RMSSD, and RSA. Significant main effects of the repeated measures MANOVA were followed by post-hoc testing on each stressor and each ANS measure separately. To obtain the effect size of each ANS measure cohen’s d was calculated. We notice that HP reactivity was used twice in this approach. Although HP reactivity does truly reflect both SNS and PNS reactivity, we did risk the results being dominated by HP effects. To examine whether this was the case, we repeated the analyses without HP in both clusters. As this did not noticeably alter the pattern of results, we report only on the MANOVA on clusters with HP left in for brevity.

All analyses were performed with age and sex as covariates. Age was transformed into a binary variable with 0 for participants under the age of 25 and 1 for participants of 25 years and older. Since respiration rate (RR) has been associated with RSA (De Geus et al., 1995; Grossman et al., 1991) and RMSSD (Schipke et al., 1991) RR was added to the analyses of these variables. If the assumption of sphericity was violated the Greenhouse-Geisser results were reported. The large amount of tests performed (N = 26) required a correction of our experiment-wise p-value to reduce type I errors. Since the variables in the PNS and SNS clusters are highly interrelated, a Bonferroni correction would be overly conservative. Instead, we used Matrix Spectral Decomposition (maSpD) to estimate the equivalent number of independent variables in the full correlation matrix of all SNS and PNS variables tested and we adjusted the p-value accordingly (QIMR Genetic Epidemiology Laboratory, Dale’s homepage https://gump.qimr.edu.au/general/daleN/matSpD/). This led to a p-value threshold of .002 for a result to be declared significant.

5. Results

Descriptive statistics of the study population (N = 87) can be found in Table 1. RSA and RMSSD were not normally distributed and therefore log-transformed. Regarding data quality: two outlier values (> 3SD) were removed for log-transformed RSA, eight for log-transformed RMSSD and four for PEP. For two participants the data quality of the EDA recording was considered too low for reliable peak detection.

5.1. Affect

To test whether the short SSST affected subjective reporting of positive and negative affect and whether this effect was similar to the TA task, Wilcoxon signed rank tests were performed. Fig. 3 shows that participants felt both significantly less positive (short SSST: N = 87, Z = −3.65, p < .001; TA: N = 87, Z = −5.54, p < .001) and more negative (SSST: N = 87, Z = −4.69 p < .001; TA: N = 87, Z = −6.44, p < .001) after these tests. There was no significant difference in affect scores between the two tests (N = 87, Z = −2.29, p = .022).

5.2. ANS reactivity

Table 2 shows the means and standard deviations of the ANS
measures and RR during the different conditions.

There was a significant difference in SNS activity between the three conditions (Greenhouse-Geisser: F(1.86, 140) = 24.86, p < .001). Contrasts analyses indicated a significant difference between baseline and SSST (F(1,70) = 69.08, p < .001), baseline and TA task (F(1,70) = 1582.84, p < .001) and SSST and TA task (F(1,70) = 1657.38, p < .001). The difference in SNS activity between baseline and the short SSST was driven by all individual SNS measures, with medium to large effect sizes (Table 3).

There was also a significant difference in PNS activity between the three conditions (Greenhouse-Geisser: F(1.80, 162) = 24.70, p < .001). Contrasts analyses indicated a significant difference between baseline and SSST (F(1,81) = 1775.77, p < .001), baseline and TA task (F(1,81) = 3094.80, p < .001) and SSST and TA task (F(1,81) = 2279.60, p < .001). The difference in PNS activity between baseline and the short SSST was driven by HP with a large effect size, with a trend for RMSSD with a medium effect size (Table 3).

The observed difference between the two stress tasks was entirely driven by the larger HP reactivity to the short SSST compared to the TA task, the individual SNS and PNS variables all showed comparable reactivity.

5.3. Response stereotypy across short SSST and TA tasks

To assess response stereotypy, Pearson correlations were computed on the reactivity scores (stress test – baseline) for the TA and short SSST tasks across all 6 variables. There was a significant positive correlation (p < .001) between the short SSST and TA reactivity for all SNS and PNS variables (Fig. 4) showing autonomic stress reactivity to be a stable individual characteristic across the mental and social-evaluative domains.

6. Discussion

The “Sing-a-Song Stress Test” (SSST) is shown to be a valid shorter alternative for the longer and labor-intensive Trier Social Stress Test (TSST) to evoke social-evaluative stress (Brouwer and Hogervorst, 2014). The current study shows that a shorter and more practical version of the SSST still effectively induces social-evaluative stress reflected by both affective responses and physiological reactivity.

In the current study several improvements have been made to the original SSST. First, the short SSST contains fewer trials, thus decreasing the overall duration from ~15 (SSST) to ~6.5 minutes. Second, confederates are no longer required. Third, by adding a training condition (read aloud), we were able to eliminate the problem of participants starting to sing too early; as none of the participants started singing before they were instructed to do so. Last, the short SSST was validated using a more diverse range of ANS measures providing broader insight into the ANS reactivity caused by this stressor.

In accordance with our expectations, the short SSST significantly decreased positive affect, increased negative affect, and shifted the ANS to a state of increased SNS activity and decreased PNS activity, with medium to large effect sizes. With regard to the cardiac ANS measures,
the short SSST led to higher HP and RMSSD reaction and effect size (Brindley et al., 2014). Consistent with the findings of previous studies investigating cardiac ANS reactivity to a wide array of other stress tasks in both direction and magnitude for all cardiac ANS measures. These findings are consistent with that of previous studies investigating cardiac ANS reactivity to a wide array of other stress tasks in both direction and effect size (Bosch et al., 2014). Consistent with the findings of Bosch et al. (2009), the short SSST led to higher HP and RMSSD reactivity compared to the TA task. However such an effect was not observed for PEP.

With regard to our EDA measures, the reactivity to the short SSST and TA test were also of similar direction and magnitude. The SCL results, however, showed larger differences between the short and the original SSST (SCLshort SSST: 4.1 μS vs. SCLoriginal SSST: 10.9 μS). This may be partly explained by the strong sensitivity of absolute SCL levels to the type and placement of the electrodes as well as the room temperature. Therefore we measured ns.SCRs as an alternative read-out of skin SNS. Measuring ns.SCRs has two advantages. First, it is less sensitive to temperature and type and placement of the electrodes. Second, during debriefing we informed the participants that none of their singing was actually recorded. It would be interesting to investigate this in future studies.

Thenature of their in compliance is unknown. It might be that they just did not feel like singing or that they were too stressed to even start singing. It would be interesting to investigate this in future studies. Our HP results (converted to HR for comparison) are consistent with those of the original longer SSST (Brouwer and Hogervorst (2014) but of slightly smaller magnitude (HRshort SSST: 11.8 bpm vs. HRoriginal SSST 15.3 bpm). When compared to an often employed mental stress test, a speeded reaction time task in which incorrect and slow responses are punished by aversive loud tones, the short SSST evoked reactivity of similar direction and magnitude for all cardiac ANS measures. These findings are consistent with that of previous studies investigating cardiac ANS reactivity to a wide array of other stress tasks in both direction and effect size (Brindley et al., 2014). Consistent with the findings of Bosch et al. (2009), the short SSST led to higher HP and RMSSD reactivity compared to the TA task. However such an effect was not observed for PEP.

Using a variety of ANS and affective measures our results show that the short SSST is a potent stress-inducing task. This is further supported by the substantial correlation of our ANS measures between the short SSST and the TA task, suggesting that the short SSST captures the general trait of being a low or high ‘stress-reactor’ rather well. This suggests that social-evaluative stress can be effectively induced even without the need for confederates. However, we do note several limiting factors to the use of the short SSST. First, several participants refused to sing entirely, causing their data to be unusable for this study. The nature of their in compliance is unknown. It might be that they just did not feel like singing or that they were too stressed to even start singing. It would be interesting to investigate this in future studies.

Second, during debriefing we informed the participants that none of their singing was actually recorded. We noticed that some of the participants indicated that they already suspected this because they had not signed formal informed consents that their performance would be shared, but that, even so, they were not entirely sure. Unfortunately we did not document this, therefore we could not investigate a possible effect on task outcome. Third, it is unlikely that the short SSST can be used repeatedly within the same subject to the same effect. The task, like the original TSST and the SSST, requires a form of deception that demands full debriefing from an ethical point of view, which may greatly reduce its impact on repeated exposure. For the same reason, a direct comparison of the original longer SSST with the new short SSST in the same participants was not feasible. Last, the observed ANS effects were a little smaller than those found by Brouwer and Hogervorst. Though this could be due to differences in study population, we cannot exclude that the difference might be due to the lack of physically present confederates.

In conclusion, the short SSST is a more time-efficient and less labor-intensive stressor that may be more easily available for research purposes. The short SSST is a potent stress-inducing task that may serve as an alternative to the original TSST and SSST, particularly in settings where time is limited. Further research is needed to investigate the potential of the short SSST as a stressor in different populations and settings.
Fig. 4. Response stereotypy.
A scatter plot with regression line of short SSST reactivity with tone avoidance reaction time task (TA) reactivity is shown for each ANS measure. PEP ($r = 0.51, p < .001$), SCL ($r = 0.63, p < .001$), ns.SCR ($r = 0.41, p < .001$), HP ($r = 0.56, p < .001$), RSA ($r = 0.88, p < .001$) and RMSSD ($r = 0.71, p < .001$).
intensive alternative to the SSST and TSST. It induces social-evaluative stress to a sufficient degree and evokes measurable changes in affective state, PNS and SNS activity. We believe that this test can be successfully used in large scale studies on the causes and consequences of individual differences in autonomic responding to stress.

Acknowledgments

The authors declare no conflicts of interest.

This study was partially funded by Philips Research, Eindhoven, the Netherlands.

Appendix 1. Study overview

The full study protocol entails a larger series of ANS recordings on two consecutive days.

Day 1.

On the first day of the experiments participants visit the lab for a structured interview (regarding the subject's demographics, medication use, perceived physical and mental health and lifestyle behaviors), anthropometric measurements (length, weight, fat percentage, muscle percentage and blood pressure) and device application. They are instructed to fill in a diary (consisting of questions on affect, current activity and environment) which will be presented to them on an electronic device hourly between 09:00 h and 23:00 h. The electronic diary is practiced once to make sure all questions are understood. Participants continue with their everyday routine to return to the lab on the next day (~23 h later).

Day 2.

The second day of the experiment consists of an extensive laboratory protocol of ~2 h presented in the figure below (the R in R-Sitting, R-Standing and R-treadmill stand for recovery). After each task an affect measurement was taken (indicated by a red line and labels A1 to A9). After the laboratory protocol all devices are removed and subjects are debriefed. The questions on affect are exactly the same as the ones used in the electronic diary.

Timeline of experiment

| Posture | Mental stress | Daily activities | Mental stress | Physical stress |
|---------|---------------|----------------|---------------|----------------|
| A1      | A2            | A3             | A4            | A5             |
| A6      | A7            | A8             | A9            | Part that was used in the current analyses |

References

Benschop, R.J., Schedlowski, M., 1999. Chapter 15: acute physiological stress. In: Schedlowski, M., Tewes, U. (Eds.), Psychoneuroimmunology: An Interdisciplinary Introduction. Springer Science + Business Media, LLC, New York, USA, pp. 293–306.

Bosch, J.A., De Geus, E.J., Carroll, D., Goedhart, A.D., Anane, L.A., van Zanten, J.J.V., ... Edwards, K.M., 2009. A general enhancement of autonomic and cortisol responses during social evaluative threat. Psychosomatic medicine 71 (8), 877.

Boucsein, W., 2012. Electrodermal Activity. Springer Science & Business Media.

Brouwer, A.-M., Hogervorst, M.A., 2014. A new paradigm to induce mental stress: the Sing-a-Song Stress Test (SSST). Front. Neurosci. 8, 224.

Dawson, E., Schell, A., Filion, D.L., 2000. No title. In: Dawson, E., Schell, A., Filion, D.L. (Eds.), Handbook of Psychophysiology, 2nd edn. Cambridge University Press, Cambridge, pp. 203–223.

Dickerson, S.S., Kemeny, M.E., 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol. Bull. 130 (3), 355–391.

Dickerson, S.S., Gruenewald, T.L., Kemeny, M.E., 2004. When the social self is threatened: shame, physiology, and health. J. Pers. 72 (6), 1191–1216.

Erdmann, G., Janke, W., Bispring, R., 1984. Wirkungen und Vergleich der Wirkungenvon... 

Erdmann, G., Janke, W., Bispring, R., 1984. Wirkungen und Vergleich der Wirkungenvon... 

Goedhart, A.D., van der Sluijs, S., Houtveen, J.H., Willemsen, G., de Geus, E.J.C., 2007. Comparison of time and frequency domain measures of RSA in ambulatory recordings. Psychophysiology 27 (6), 205–215.

Gronwall, D.M.A., 1977. Paced auditory serial-addition test: a measure of recovery from concussion. Percept. Mot. Skills 44 (2), 367–373.

Dickerson, S.S., Gruenewald, T.L., Kemeny, M.E., 2004. When the social self is threatened: shame, physiology, and health. J. Pers. 72 (6), 1191–1216.

Grossman, P., Beek, J.V., Wientjes, C., 1990. A comparison of three quantification methods for estimation of respiratory sinus arrhythmia. Psychophysiology 27 (6), 702–714.

Grossman, P., Karmaker, P., Wieling, W., 1991. Prediction of tonic parasympathetic cardiac control using respiratory sinus arrhythmia: the need for respiratory control. Psychophysiology 28 (2), 201–216.

Joffily, M., 2012. EDA master toolbox. Available at. https://github.com/mateusjoffily/EDA/

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. Neuropeychobiology 28, 76–81.

van Lien, R., Schutte, N.M., Meijer, J.H., de Geus, E.J., 2013. Estimated pre-ejection period (PEP) based on the detection of the R wave and GZ/gh-min peaks does not adequately reflect the actual PEP across a wide range of laboratory and ambulatory conditions. Int. J. Psychophysiol. 87, 60–69.

Matyas, T.A., King, M.G., 1976. Stable T wave effects during improvement of heart rate control with biofeedback. Physiology & behavior 16 (1), 15–20.

Myin-Germeys, I., van Os, J., Schwartz, J.E., Stone, A.A., Delespaul, P.A., 2001. Emotional reactivity to daily life stress in psychosis. Arch. Gen. Psychiatry 58 (12), 1137–1144.

Nederend, I., ten Harkel, A.D., Blom, N.A., Berntson, G.G., de Geus, E.J., 2017. Impedence cardiography in healthy children and children with congenital heart disease: improving stroke volume assessment. Int. J. Psychophysiol. 120, 136–147.

Nederend, I., de Geus, E.J., Kroft, L.J., Westenberg, J.J., Blom, N.A., ten Harkel, A.D., 2018. Cardiac autonomic nervous system activity and cardiac function in children after coarctation repair. Ann. Thorac. Surg. 105 (6), 1803–1808.

Novak, D., Mihelj, M., Munihi, M., 2011. Psychophysiological responses to different levels...
of cognitive and physical workload in haptic interaction. Robotica 29 (3), 367–374.

Renaud, P., Blondin, J.-P., 1997. The stress of Stroop performance: physiological and emotional responses to color-word interference, test pacing, and pacing speed. Int. J. Psychophysiol. (2), 87–97.

Schipke, J.D., Arnold, G., Pelzer, L., 1991. Effect of respiration rate on short-term heart rate variability. J. Clin. Basic Cardiol. 2 (1), 92–95.

Sherwood, A., Allen, M.T., Fahrenberg, J., Kelsey, R.M., L ovallo, W.R., Van Doornen, L.J., 1990. Methodological guidelines for impedance cardiography. Psychophysiology 27 (1), 1–23.

Skoluda, N., Strahler, J., Schlotz, W., Niederberger, L., Marques, S., Fischer, S., ... Nater, U.M., 2015. Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity. Psychoneuroendocrinology 51, 227–236.

Stroop, J.R., 1935. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18, 643–662.

Tanoso t o, T., et al., 2012. A Paced Auditory Serial Addition Test evokes stress and differential effects on masseter-muscle activity and haemodynamics. Eur. J. Oral Sci. 120, 363–367.

Taylor, S., et al., 2015. Automatic identification of artifacts in electrodermal activity data. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2015, 1934–1937.

Tombaugh, T.N., 2006. A comprehensive review of the paced auditory serial addition test (PASAT). Arch. Clin. Neuropsychol. 21 (1), 53–76.

Weiner, H., 1992. Perturbing the Organism: The Biology of Stressful Experience. University of Chicago press.