Acta Psychologica 251 (2024) 104645

Contents lists available at ScienceDirect

Acta Psychologica

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

Assessing last month's stress levels with an automated facial behavior scan

Marnix Naber [a,*], Sterre I.M. Houben [b,c], Anne-Marie Brouwer [d,e]

[a] Experimental Psychology, Helmholtz Institute, Faculty of Social and Behavioral Sciences, Utrecht University, Utrecht, the Netherlands
[b] Centre for Man in Aviation, Royal Netherlands Air Force, Soesterberg, the Netherlands
[c] Defense Personnel and Organization Division, Ministry of Defense, Amsterdam, the Netherlands
[d] Human Performance, TNO Human Factors, Soesterberg, the Netherlands
[e] Artificial Intelligence, Radboud University, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Stress
Facial behavior
Computer vision
Emotion
Military

ABSTRACT

Stress is one of the most pressing problems in society as it severely reduces the physical and mental wellbeing of people. It is therefore of great importance to accurately monitor stress levels, especially in work environments. However, contemporary stress assessments, such as questionnaires and physiological measurements, have practical limitations, mostly related to their subjective or contact-based nature. To assess stress objectively and conveniently, we developed an automated model that detects biomarkers in webcam-recorded facial behavior indicative of heightened stress levels, using computer vision, artificial intelligence, and machine learning techniques. Heart-rate induced skin pulsations and facial muscle activity were extracted from videos of 264 participants that performed an online mental capacity test under considerable time pressure. The model could successfully use these facial biomarkers to explain a significant proportion of individual differences in scores on a self-perceived stress scale. Next, we used the model to objectively score stress levels of 63 military candidates (pre-hiring) and 69 military personnel (post-hiring) that also performed the mental capacity test. Results showed that military personnel expressed facial behavior indicative of significantly higher stress levels than military candidates. This suggests that joining the military heightens overall stress levels. With this study we take the first steps towards a non-contact, automated, and objective measure of stress that is easily applicable in a variety of health and work contexts.

1. Introduction

Stress is a mental and physiological condition raised by a combination of an overflow of environmental demands and an individual's inability to cope with the unpredictability and uncontrollability of stressful situations (Koolhaas et al., 2011). Being in a prolonged state of stress may cause a multitude of mental and physical problems (Chrousos, 2009; Yaribeygi, Panahi, Sahraei, Johnston, & Sahebkar, 2017), including impairments in brain function leading to cognitive and pathological disorders (Sandi, 2013), and impairments in the immune (Reiche, Vissoci, Vargas, & Morimoto, 2004), cardiovascular (Rozanski, Blumenthal, & Kaplan, 1999), and gastrointestinal system (Konturek, Brzozowski, & Konturek, 2011). Humanity is facing major stress-related problems in work settings (Ganster & Rosen, 2013) and beyond (Compas, 1987), putting a significant financial burden on society (Hassard, Teoh, Visockaite, Dewe, & Cox, 2018). Stress-related burnouts are highly prevalent worldwide (Low et al., 2019; Schaufeli & Enzmann, 1998) and these numbers are still rising in a number of occupational fields (e.g., Arigoni, Bovier, & Sappino, 2010; Ge et al., 2023).

Accurate measurements are key to the field of stress diagnostics. A large variety of diagnostic stress instruments exists (Abbas, Farah, & Apkinar-Sposito, 2013; Gormally & Romero, 2020; Sharma & Gedeon, 2012), though the most often applied stress detection methods can be divided in three types of measurements: (1) subjective self-evaluation questionnaires such as the Perceived Stress Scale (PSS; (Cohen, Kamarck, & Mermelstein, 1994)) and Connor-Davidson Resilience Scale (CDRS; (Connor & Davidson, 2003)), (2) physiological measurements such as heart rate variability (Järvelin-Pasanen, Sinikallio, & Tarvainen, 2018; Kim, Cheon, Bai, Lee, & Koo, 2018; Schiweck, Piette, Berckmans, Claes, & Vrieze, 2019), and (3) neuroendocrine measurements such as blood or saliva analyses on cortisol and other compounds (Biondi & Picardi, 1999; Carrasco, de Kar, & Louis, 2003). Each of these methods have specific benefits, but also shortcomings. The most persistent problems of stress measurements are caused by either their reliance on subjective feedback, time-consuming nature, impractical (contact-based) setups, or the sensitivity of measurement outcomes to contextual

* Corresponding author at: Room H0.24, Heidelberglaan 1, 3584CS Utrecht, the Netherlands.

https://doi.org/10.1016/j.actpsy.2024.104645
Received 11 June 2024; Received in revised form 24 November 2024; Accepted 4 December 2024
Available online 7 December 2024

factors. For example, performing a stress measurement may already be a stressor on itself. Also, asking questions about stress explicitly reveals the goal of the questionnaire or study, allowing subjects to provide answers contaminated by, for example, social desirability (Mezulis, Abramson, Hyde, & Hankin, 2004; Van de Mortel, 2008; Viswesvaran & Ones, 1999). Contact-based and invasive physiological measurements can be uncomfortable and analyses tend to be complex and slow. Here, we report on an investigation into the accuracy of a computer vision solution in detecting stress without being limited by these problems.

During the last two decades, several studies developed alternative stress detection methods using computer vision techniques. For example, Sharma and colleagues demonstrated that analysis of thermal patterns in videos of faces can separate stressed from non-stressed individuals (Sharma, Dhall, Gedeon, & Goecke, 2014). Other approaches investigated the relationship between stressors and facial behavior in humans, showing that the activation of anxiety-related mental states result in the expression of specific facial muscle changes (Bruin et al., 2024; Dinges et al., 2005; Gao, Yüce, & Thiran, 2014; Jabon, Bailenson, Pontikakis, Takayama, & Nass, 2010; Lerner, Dahl, Hariri, & Taylor, 2007; Liao, Zhang, Zhu, & Ji, 2005; Rimini-Doering, Manstetten, Altmueller, Ladstaetter, & Mahler, 2001). Several other studies specifically used this knowledge to detect stress in similar manners (Almeida & Rodrigues, 2021; Giannakakis et al., 2017; Metaxas, Venkataraman, & Vogler, 2004; Naidu, Sagar, Praveen, Kiran, & Khalandar, 2021). These pioneering studies investigated stress in lab-controlled settings where participants performed tasks, such as driving a car, flying an airplane, watching videos, or pressing buttons in response to conflicting stimuli, to evoke variations in stress levels. The observation that facial behavior changes as a function of the onset of stressors in such settings, predicts that stress levels could potentially be also determined through webcam recordings while people work behind a computer. However, face-based stress detectors have not yet brought to the test in such settings. Also, previous efforts either used inaccessible (closed-source) software, detected stress from facial emotions displayed in images rather than more information-rich dynamic action unit patterns, or did not provide enough information to replicate the results. Furthermore, previous studies were only able to dissociate stress levels within each participant rather than between participants that varied in stress levels.

Here, we aim to bring facial stress detection to the test by applying an open-source and state-of-the art facial action unit activity tracking (Baltrušaitis, Robinson, & Morency, 2016) to detect intermediate-term stress levels in individuals, and in groups that work under a lot of stress. Similar to recent research demonstrating the successful detection of a variety of mental states (Hoegen, Gratch, Parkinson, & Shore, 2019; Hoque, McDuff, & Picard, 2012; Kappen & Naber, 2021; Kuipers, Kappen, & Naber, 2023), we will analyze subtle, spontaneous, and dynamic facial behavior to provide objective insights into the stress states of military personnel. More specifically, in two separate experiments, we aim (1) to develop a model that links stressor-evoked facial behavior and heart rate changes to a person's last month's self-perceived stress level, and (2) to apply the same model to validate its sensitivity in measuring differences in stress levels between groups of pre- and post-hiring employees in the military, where people typically work in a stressor-rich environment (Bustamante-Sánchez et al., 2020; Campbell & Nobel, 2009; Langston, Gould, & Greenberg, 2007; Pflanz & Sonnek, 2002). We predict that people with heightened stress levels will show different patterns of facial behavior during a stressful task than people with lower stress levels, and that this behavior can be picked up by computer vision and reflects a person's overall stress level. In addition, we hypothesize that joining the military increases stress levels as post-hire military employees due to experiences with stressful (and potentially traumatic) events.

2. Methods

2.1. Participants for stress model production

A total of 264 Dutch and English participants (age: M = 33.9, SD = 11.4, range = 19–66; 124 women), recruited from the online human data crowdsourcing platform (Prolific, Oxford, UK), participated in this study. Participants provided written informed consent and the study was approved by the local ethics committee of Utrecht University (#19–079). In exchange for money (approximately €10) participants took part in an experiment featuring a job application training assessment of which the mental stress task was one aspect.

2.2. Military participants for model validation

A total of 63 military candidates (Age: M = 26.0, SD = 5.3, range = 20–46; 10 women; Job positions: 34 soldiers, 29 officers), and 42 non-combat military personnel and 27 special forces combat personnel participated in this study. Non-combat and combat personnel had been exposed to stressful events during training and/or deployment. The exact demographics of these groups were not collected due to strict privacy rules of the military, but the study examiners indicated that gender distributions did not deviate across groups, but that age was slightly higher in the military personnel groups (post-hiring) as compared with the candidate group (pre-hiring). Participants were recruited through the Dutch Ministry of Defense from the department of security. Participants provided written informed consent and the study was in line with the ethical guidelines of the Defense Commando Support Innovation Fund.

2.3. Procedure and apparatus for data collection for stress model production

Participants took part in a study by the Utrecht University on online, remote (e.g., from home with their own computer) job training assessments. Participants were instructed to find a well-lit room where they could not be disturbed by others during the assessment. The assessment was developed by a human recruitment software company (Neurolytics B.V., Utrecht, The Netherlands), using online software running on a Chrome browser (Google, Mountain View, California, USA). At the start of the assessment, participants allowed their browser to record webcam footage. Javascript-based software (RecordRTC) recorded videos at a frame rate between 15 Hz and 40 Hz and a resolution of 800 by 600 pixels. Participants were made aware of the recordings, but were not told that the videos were analyzed by computer vision software. Next, participants followed several instructions to improve the camera recordings, such as to position themselves facing a window or bright light source, centrally in front of the camera at the right distance (approximately 50–75 cm), and without wearing any hats or other objects covering their face. Some subjects may have kept their glasses on despite these instructions, potentially adding some noise to the eye-brow measurements.

One part of the assessment consisted of a mental capacity test, serving as a stressor known for increasing arousal levels most as compared with other typical stress tests (Bruin et al., 2024). The test consisted of four blocks of 10 multiple-choice questions, with each block asking different type of questions (see Table 1). One of four answer options could be chosen per question. To increase acute stress levels, a timer on top of the page counted down from 150 s per block, with the font color becoming conspicuously red during the last 10 s. As such, participants not only had to answer the questions correctly but also as quickly as possible, finishing as many questions as possible in time. The test only served as a stressor; how well and fast the participants answered the questions was not considered further in the current study. After the mental capacity test, we aimed to determine how stressed participants felt during the last month (i.e., an intermediate-term

Table 1
Question types per block of the mental capacity test.

Block number	Question type	Example multiple-choice question
1	General knowledge	The 10th month of the year is: [November; September; August; Oktober]
2	Math - logical reasoning	A box with chocolates contains chocolates with caramel or nuts. The box contains 3 times more caramel chocolates than nut chocolates. If the box contains 20 chocolates in total, how many caramel and nut chocolates are there? [9 caramel, 11 nut; 15 caramel, 5 nut; 18 caramel, 2 nut; 12 caramel; 8 nut]
3	Math - calculus	42 * 0 * 6 =? [7; 6; 252; 0]
4	Verbal reasoning	Which word is the odd one out? [River; Ocean; Swimming pool; Lake]

assessment) by having them conduct the 10-item Perceived Stress Scale (PSS) questionnaire (Cohen, Kamarck, & Mermelstein, 1983). The PSS contains questions like "In the last month, how often have you been upset because of something that happened unexpectedly?" with each five answer options for self-evaluation (0 = never, 1 = almost never, 2 = sometimes, 3 = fairly often, 4 = very often). To calculate a total stress score, the scores per questions were simply accumulated (0 = not stressed at all, 40 = severely stressed) and then rescaled to the score range 0–10. The entire assessment lasted approximately 30 min, of which 10 min were reserved for the mental capacity test. It is important to note that the online nature of the test did not allow any contact-based physiological assessments for comparison. For this, we refer to a study on the link between facial behavior and physiology-based stress using similar tests (Bruin et al., 2024). Also, see Kuipers et al. (2023) for results of a similar study on the link between facial behavior and short-term (acute) rather than intermediate-term state-anxiety.

2.4. Procedure and apparatus for military data collection for model validation

The pre- and post-hiring military participants took the same assessment as the participants for the model. Post-hiring militants were mass-invited by the head of their department to join a larger assessment day at the Dutch Ministry of Defense with the goal to participate in a study on stress resilience in selecting military personnel. Laptops and build-in cameras were distributed across several testing rooms. Participants were instructed to take place at an unoccupied room, where they could start the online assessment. The personnel were explicitly told that the outcomes of the stress measurement would be anonymized and would thus not affect their career. Pre-hiring militants were invited as part of the selection process, and were also explicitly told that participation in the online test was not mandatory and that the outcomes would not affect the hiring decision. The candidates conducted multiple online tests of which the mental stress assessment was just a small part. All participants received information about the video-based stress levels afterwards through a personalized report.

2.5. Face-behavior-based stress model

The automated computer vision model that rated stress levels based on facial behavior was programmed in Python (Van Rossum, 1995). We followed a modelling procedure that is identical to one recently described in a study that aimed to determine states of nervousness during job interviews based on facial behavior (Kuipers et al., 2023). In short, we implemented several programming steps using distinct software packages. First, the software package OpenFace (Baltrušaitis et al., 2016) extracted time series (10 min recordings, 30 frames per second) of activity levels of 17 facial action units (e.g., distance between lips; relative raising of inner eye browes; for details, see https://github.com/TadasBaltrusaitis/OpenFace/wiki/Action-Units). OpenFace accomplishes this through a sequence of computer vision steps using multiple AI and machine learning components, including deep neural networks and support vector machines: face detection, 68 facial landmark detection (see Fig. 1a), three-dimensional head orientation detection, action unit detection, and calculation of activity traces (see Fig. 1b) of relative and time-normalized changes in position and local skin textures of each landmark. We additionally combined activity traces of multiple action units to create traces representing activity in primary emotional expressions (e.g., happy; for details, see (Kappen & Naber, 2021)). Examples of features extracted from these activity traces were the average activity in original (ori_ave) and high-pass filtered traces (hp_ave), the trend in activity (slope of a fitted linear regression), and the frequency (rate), average amplitude (amp), and average area (area) of peak increases in activity. Second, we extracted heart rate by detecting subtle skin pulsations in the face videos using an opensource remote photoplethysmography software package (van der Kooij & Naber, 2019).

The conversion of time-dependent activity traces to time-independent features, as described above, led to an improved low feature-to-sample number ratio, though still meriting careful modelling procedures to prevent overfitting. The last analysis step included the application of a supervised machine learning approach to link facial features to self-evaluated stress levels. We selected only features that showed low multicollinearity with other features (i.e., with a variance inflation factor <10). Next, we applied a repeated (200 iterations) grid-search cross-validation (70 % training sample set; 30 % test sample set) algorithm (Krstajic, Buturovic, Leahy, & Thomas, 2014) for configuration and parameter tuning of a linear Ridge regression model to assess the strength of the relationship between the stress scores and each feature. Per iteration and per regularization parameter (i.e. lambda; log scale: −1 to +3), we trained a model and calculated the mean absolute error (MAE) between the modelled stress scores and self-assessed stress scores. The model with a combination of the smallest MAE of the test set and smallest difference in MAE between the train and test set was selected for the final model (log lambda = 1.6). For this model, we selected only features with significant links to the stress scores (i.e., 29 features in total).

3. Results

We first report on the accuracy of the statistical model that used facial behavior measurements, recorded during a stressful mental task, to determine last month's self-perceived stress levels of participants. To restate the main hypothesis, people in a baseline state of heightened stress, are more likely to display stressful behavior evoked by a strong stressor, as expressed by the facial muscles. Overall, participants scored moderately on stress according to the PSS questionnaire results (M = 5.4; SD = 1.9; range: 0–10). The model showed significant links between 29 facial behavior features and the stress scores (Fig. 2a). The most relevant features indicating higher levels of stress included strong (based on the amplitudes of activity peaks) fearful expressions and blinks, overall (average activity) tightened lips, eye lids, and pulled lip corners, increasingly more (slope of linear trend in activity) raised outer eye brows towards the end of the stressful task, frequent (rate of activity peaks) episodes of arousing expressions, and long lasting (duration of activity peaks) episodes of expressions of interest and surprise. Vice versa, participants with lower levels of stress mostly showed frequent wrinkling of the nose, dropping of the lip corners and jaw, and raising of the inner eye brows, high heart rates, long and/or strong (duration & area of activity peaks) dimpling of the cheeks and stretching of the lips, increasingly more relaxing state of the jaw and happy expressions towards the end, and an overall opened mouth (lips part).

The linear regression model, that combined all features to determine stress scores, could explain 25 % of the variance across individuals in stress scores (Fig. 2b; Pearson's correlation: $r = 0.51$, $p < 0.001$; Mean absolute error: MAE = 1.25), producing an identical mean though

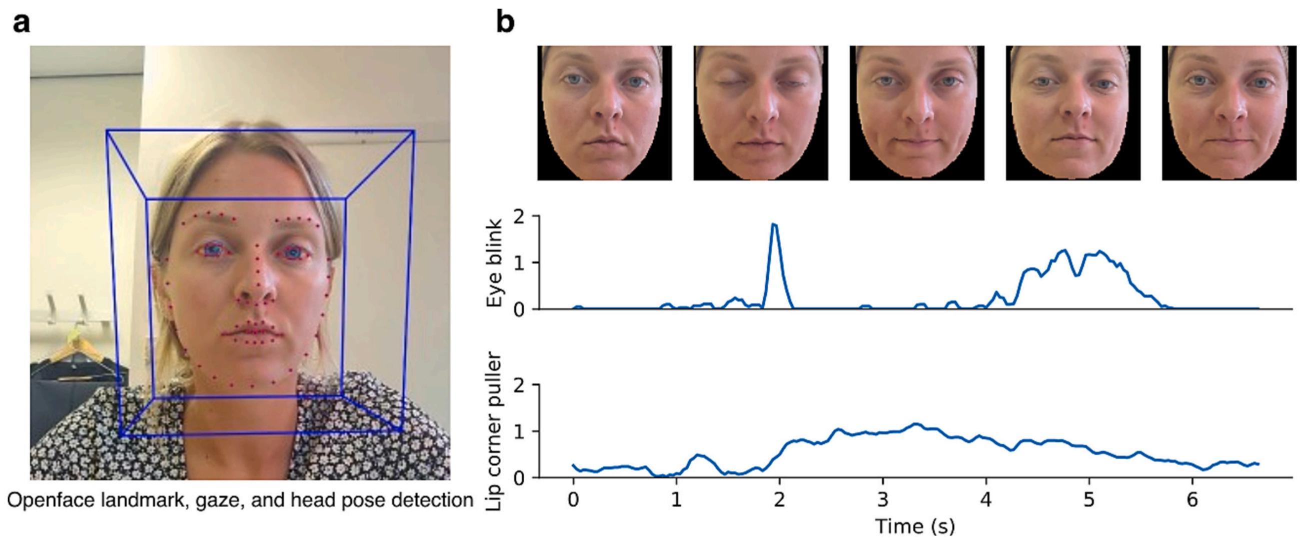

Fig. 1. OpenFace output. a, Image of OpenFace output on landmark, gaze direction, and head posture detection in the face of an actor. b, Examples of time series of action unit 45 (eye blink) and 12 (lip corner puller) with accompanying face images.

narrower distribution of scores as the PSS questionnaire (M = 5.4; SD = 0.8; range: 2.6–8.3). It is important to note that the production model may slightly overestimate true performance (estimation: 5–10 %), as we did not include a test on a validation dataset. Nonetheless, this result confirms the hypothesis that the mental test as a stressor evoked behavior that reflect the participant's stress levels. We here further demonstrate that computer vision can accurately track this behavior to create a statistical model that uses this behavior as input to calculate stress scores as output.

Next, to validate the model's sensitivity in detecting stress states, we assessed whether the model can dissociate between populations that experience different levels of stress. It is important to stress again that we assessed military personnel as it is known that this population typically experiences many and very stressful events during their jobs. We compared this (post-hire) population to a population of pre-hire candidates that applied for a job at the army. The model's stress scores per group (Fig. 2c) indicated that the group with pre-hire military candidates scored lower in stress (M = 4.6; SD = 2.1; range: 1.1–8.5) than post-hire military personnel (M = 6.2; SD = 2.0; range: 0–8.5; $t(105) = 3.79$, $p < 0.001$, Cohen's $d = 0.76$, AUROC = 0.72) and special forces (M = 5.9; SD = 1.9; range: 2.1–8.5; $t(90) = 2.61$, $p = 0.009$, Cohen's $d = 0.63$, AUROC = 0.67), while military personnel and special forces did not differ in stress scores ($t(69) = 0.75$, $p = 0.453$, Cohen's $d = 0.15$, AUROC = 0.55). These results indicate that post-hire military personnel experience more stress than pre-hire candidates, and that the face-behavior-based stress model has the sensitivity to dissociate groups with differences in stress states.

4. Discussion

Traditional methods of stress diagnostics have limitations that hamper implementations in daily life and work settings. In the current approach a webcam recorded dynamic facial behavior as potential markers of stress, while participants performed a challenging task under time pressure in a web-browser. A linear regression model, trained to link changes in activity of a variety of facial behaviors, explained a significant portion of variance in self-perceived stress levels across individuals. The test is (1) convenient for testers due to the full automation, (2) non-contact and convenient for participants, allowing them to perform the test at home in a safe environment, (3) objective as the results rely on an objective, computerized evaluation process (despite that it is trained on subjective data), and (4) inexpensive because it only requires a consumer-level computer with camera and an internet connection.

The face-behavior-based stress measurement showed a reasonable correlation with the self-perceived stress measurement, despite the known limitations of such self-assessments (Habersaat, Abdellaoui, & Wolf, 2021). It is important to note that, in contrast to questionnaires that assess stress, the current application requires a stressful mental task that lasts multiple minutes. We hope future research will investigate how long such stressors need to last for a computer vision model to pick up on enough stress-evoked behaviors to determine an individual's stress level at a validity and reliability that is comparable to those of existing questionnaires. Nevertheless, in addition to previously demonstrated measurements of acute forms of state-anxiety (Kuipers et al., 2023), the current model indicates that a person's less recent (i.e., last month) stress level reflects back in specific facial behavior evoked by a stressor. In line with a number of recent publications on the employment of novel computer vision techniques to link complex facial behaviors to mental states (Bruin et al., 2024; Giannakakis et al., 2017; Hoegen et al., 2019; Hoque et al., 2012; Kappen & Naber, 2021; Kuipers et al., 2023), this study demonstrates the usefulness of such techniques in a novel context and, more importantly, in detecting a more stable rather than acute mental state.

An analysis of the observed behavior during heightened stress levels can be interpreted as if facial changes serve (1) a reduction in chemosensory interaction (lip and eye lid tightening) and (2) a reduction in energy-neutral, relaxed expressions (opening of the mouth, and jaw and lip corner drops), and (3) a communication of distress through a mixture of more arousal, fear, surprise, and interest expressions and less happy expressions. This functional interpretation makes sense in the context of adaptive and beneficial behavior for stressful situations, which are typically uncontrollable and unpredictable (Koolhaas et al., 2011). Such environments require conservative interaction schemes with minimal chemosensory (and physical) contact, effort (energy), and warnings to others about being not at ease and potential threats.

Multiple facial features can indicate stress (Giannakakis et al., 2017). The here reported stress-evoked facial changes, including the changes to the lips, jaw, and eye brows, play a similar role in expressing nervousness as a form of state-anxiety during job interviews (Kuipers et al., 2023). Changes in the mouth and eyebrow regions also highlight stressful episodes while people fly an airplane (Dinges et al., 2005), drive a car (Jabon et al., 2010), or perform computer tasks (Liao et al., 2005). The same features and several of the other relevant features, namely fearful and happy expressions, have also been reported as crucial markers of stress evoked during (i) challenging arithmetic tasks (Lerner

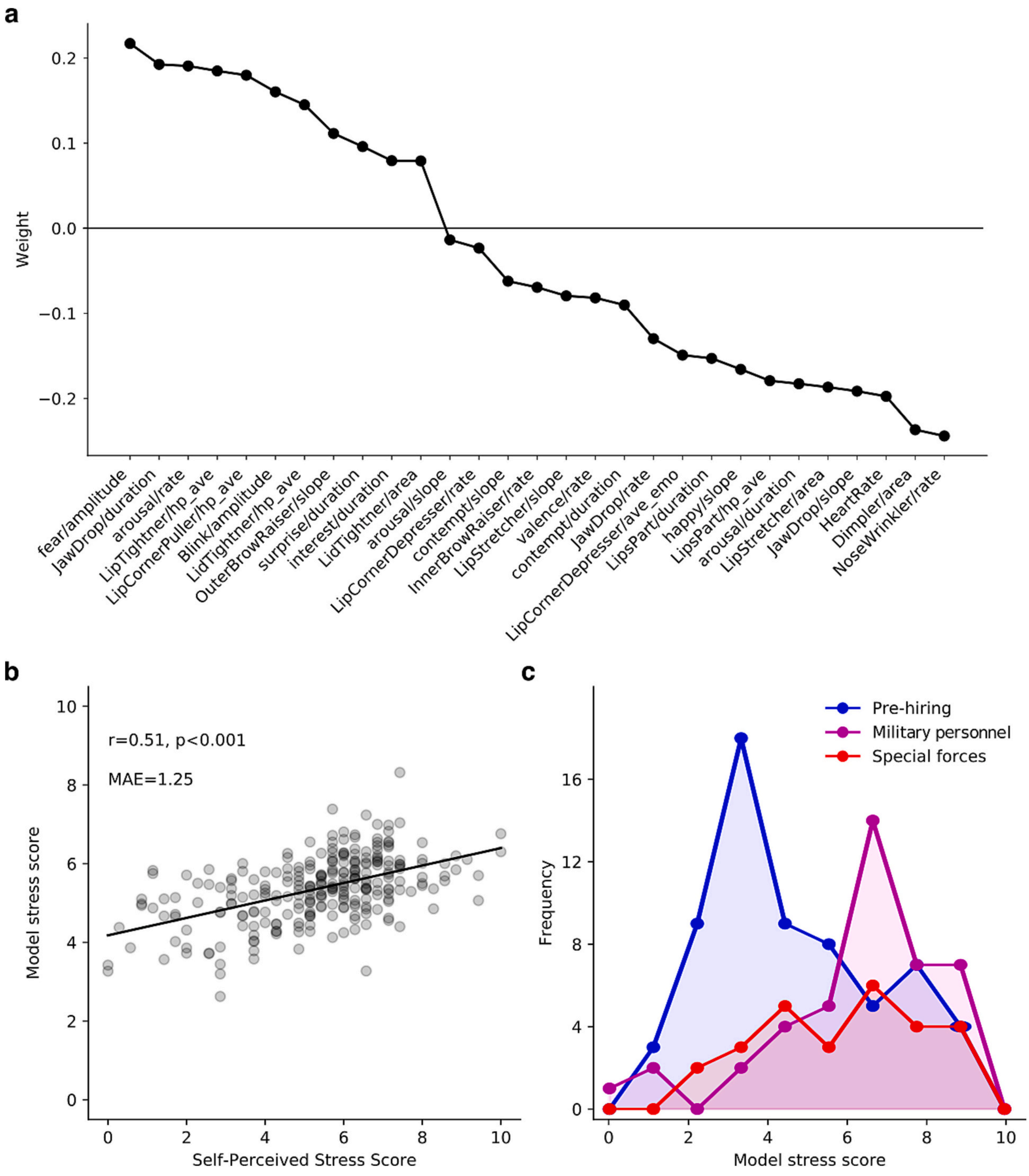

Fig. 2. Facial modelling results. a, Feature weights (betas) of Ridge model that predicts stress levels based on facial features (x-axis). The dots indicate to what degree and with which sign each facial feature links to stress levels. The x-labels describe each facial action unit (e.g., lipCornerPuller) together with the type of features extracted from the activity traces as shown in Fig. 1b (ori_ave = average of original activity trace; hp_ave = average of high-pass filtered activity trace; rate = frequency of activity peaks; area = average area under activity peaks; slope = general inclining or declining trend of activity trace; duration = average duration of activity peaks; amplitude = average height of activity peaks). b, Scatter plot of modelled (face-behavior-based) and self-perceived stress (PSS) scores. c, Histogram of modelled scores of military candidates (blue; pre-hire), and military personnel (magenta; post-hire) and special forces (red; post-hire).

et al., 2007), (ii) frustrating episodes while trying to deal with erroneous online forms (Hoque et al., 2012), and (iii) expressive periods of frustration and emotion regulation in a social cooperation versus competition task (Hoegen et al., 2019). Another study, using baseline conditions and within-subject comparisons, also discovered several facial stress markers (Bruin et al., 2024). Nevertheless, the current work provides an alternative approach that led to new insights. Rather than predicting temporary and acute increases in task-evoked stress within individuals, as already accomplished in previous studies, this study shows how differences in stress levels across individuals, self-assessed during a longer time period (i.e., a month), can be determined by fully automated analyses of temporal patterns of spontaneous facial behavior evoked by a task stressor, without the need for a baseline condition. We can conclude that being in a heightened state of stress for at least a month facilitates stress behavior when confronted with a stressor.

A rather unexpected result was the negative association between heart rate and stress levels. Increased mental stress levels typically relate to increased heart rates (e.g., Burns, Sun, Fobil, & Neitzel, 2016; Lazarus, Speisman, & Mordkoff, 1963; Taelman, Vandeput, Spaepen, & Van Huffel, 2009; Toet, Bijlsma, & Brouwer, 2017; Vrijkotte, Doornen, Lorenz, Geus, & Eco, 2000). However, unpleasant and arousing stimuli – the challenging questions in our test can be labelled as such – can also lower heart rate (Bradley & Lang, 2000; Brouwer, Van Wouwe, Mühl, van Erp, & Toet, 2013). Thus, results on the association between heart rate and arousing, negative stimuli have been mixed. The direction of the association could be task- or stimulus-specific (Kreibig, 2010), or a consequence of a yet unknown though complex interaction.

The model's findings led to a subsequent experiment in which we tested whether the model could dissociate between a group of military personnel (post-hiring) and candidates that applied for a military job (pre-hiring). Non-combat and combat (special forces) personnel in the military showed more stressful behavior than military candidates during the test. This suggests that the facial-behavior-based stress detection model has the sensitivity to dissociate between groups that likely experience different levels and types of stress. Unfortunately, we could not assess how these groups differed on demographics, such as age, gender, and education. Stress levels can, for example, decrease as a function of age (Klein et al., 2016). However, it was estimated that the average age was higher in the post- than pre-hiring group, which should have resulted in an opposite effect as reported here. Future studies will hopefully be allowed to incorporate detailed demographic information about military personnel to scrutinize such links more clearly.

Our findings and model open up new avenues to measure effects of prolonged – perhaps even chronic – periods of stress. The nature of the current approach allows to monitor stress levels within a relatively short time period, while subjects sit behind a computer or operate a mobile phone. Future work could also focus on training models that predict health and work outcomes based on stress-related behavior evoked in varying circumstances. Potential applications of such models could, for example, include the continuous monitoring of stress at (office) work to prevent burnout, the use of occasional stress tests in the military and other risky work environments to prevent stress-related mental problems such as PTSD, and efficient stress resilience assessments during hiring to improve person-job fit.

How individuals react in terms of facial behaviors and physiological responses to acute stressors could potentially link to personality traits such as temperament (Soliemanifar, Soleymanifar, & Afrisham, 2018). For example, individuals scoring high in temperamental reactivity may also react strongly to stressors and experience high levels of stress (Henderson & Wachs, 2007; Williams, Smith, Gunn, & Uchino, 2011). Although more research would be needed to confirm this, we deem it not unlikely that personality traits such as temperament may moderate facial and physiological stress responses. But because temperament is known as a highly stable trait after childhood (Martin, Lease, & Slobodskaya, 2020), such personality traits cannot explain the differences in stress between pre-hire and post-hire militants, which are caused by short-term, environmental factors rather than long-term, trait-related factors.

One's self-assessment of stress levels during the last month is likely determined by the combination of one's general stress level and sensitivity to react to stress (trait stress), and the occurrence of stressful live events in the last few weeks (delayed state stress). We are curious whether future models perform comparable to the current model when measuring more general rather than last month's stress levels. Another yet to be studied aspect is whether the presence of a stressor is a requirement for the model to detect stress levels. In other words, is it necessary to evoke behavior with a demanding, frustrating, unpredictable, or uncontrollable task, in order to gain insights in a person's stress level? Or does being in a prolonged stress state automatically lead to signs of stress in facial behavior, even when not confronted with stressors? Although several studies consistently found similar types of stress-related facial behaviors to predict stress states, future studies may also want to investigate whether models generically detect stress independent of the type of context. Another interesting line of research is the exploration of how the duration of being in a heightened stress state, validated by both subjective (questionnaires) as objective (physiological and endocrine) measures, relate to the frequency, intensity, and likelihood of showing stressful facial behavior in several contexts. For instance, a recent study showed that students that evaluated their perceived stress state during last month as high, also scored high on trait-anxiety but not state-anxiety (Liu, Qiao, & Lu, 2021). Although out of the scope of the current study, we are looking forward to a more thorough investigation into the role of such personality characteristics, stress resilience, and stress-evoked facial behaviors.

CRediT authorship contribution statement

Marnix Naber: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sterre I.M. Houben:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization. **Anne-Marie Brouwer:** Writing – original draft, Validation, Resources, Funding acquisition.

Declaration of competing interest

Authors AB and SH declare no competing interests. Author MN declares to have received shares in Neurolytics B.V. in exchange for consultancy on the development of the here reported models in the context of a government funded, joint valorization project in collaboration with Utrecht University.

Acknowledgments

We thank Mithras Kuipers and Sebastian Restrepo for help with collecting the data and building the stress model, Felix Hermsen and Belen Hein for help designing the study on military candidates and personnel, and psychological selection department of the Dutch ministry of defense, DOSCO innovation fund, Object Security Brigade of the Royal Military Police, and Commando Corps of the Royal Netherlands Army for help with setting up the study and recruiting candidates and personnel for participation. This study was financially supported by the NWO take-off valorization grant (number 17777) and the DOSCO Innovation Fund of the Dutch Ministry of Defense.

Data availability

The data that support the findings of this study are available upon reasonable request. Restrictions may apply to the availability of the behavioral data that was used to train the stress model because it is intellectual property of Neurolytics BV. These data are however available from the authors upon reasonable request and with permission of Neurolytics BV. Enquiries can be sent to m.naber@uu.nl.

References

Abbas, S. G., Farah, A., & Apkinar-Sposito, C. (2013). Measuring the immeasurable! An overview of stress & strain measuring instruments. *Mediterranean journal of. Social Sciences, 4*(10).

Almeida, J., & Rodrigues, F. (2021). Facial expression recognition system for stress detection with deep learning. *Paper presented at the ICEIS, 1*.

Arigoni, F., Bovier, P. A., & Sappino, A.-P. (2010). Trend in burnout among Swiss doctors. *Swiss Medical Weekly, 140*(3132), w13070.

Baltrušaitis, Tadas, Robinson, Peter, & Morency, Louis-Philippe. (2016). *Openface: an open source facial behavior analysis toolkit.* Paper presented at the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV).

Biondi, M., & Picardi, A. (1999). Psychological stress and neuroendocrine function in humans: The last two decades of research. *Psychotherapy and Psychosomatics, 68*(3), 114–150.

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology, 37*(2), 204–215.

Brouwer, A.-M., Van Wouwe, N., Mühl, C., van Erp, J., & Toet, A. (2013). Perceiving blocks of emotional pictures and sounds: Effects on physiological variables. *Frontiers in Human Neuroscience, 7*, 295.

Bruin, J., Stuldreher, I. V., Perone, P., Hogenelst, K., Naber, M., Kamphuis, W., & Brouwer, A.-M. (2024). Detection of arousal and valence from facial expressions and physiological responses evoked by different types of stressors. *Frontiers in Neuroergonomics, 5*, Article 1338243.

Burns, K. N., Sun, K., Fobil, J. N., & Neitzel, R. L. (2016). Heart rate, stress, and occupational noise exposure among electronic waste recycling workers. *International Journal of Environmental Research and Public Health, 13*(1), 140.

Bustamante-Sánchez, A., Tornero-Aguilera, J. F., Fernández-Elías, V. E., Hormeño-Holgado, A. J., Dalamitros, A. A., & Clemente-Suárez, V. J. (2020). Effect of stress on autonomic and cardiovascular systems in military population: A systematic review. *Cardiology Research and Practice, 2020*.

Campbell, D. J., & Nobel, O. B.-Y. (2009). Occupational stressors in military service: A review and framework. *Military Psychology, 21*(sup2), S47–S67.

Carrasco, G. A., de Kar, V., & Louis, D. (2003). Neuroendocrine pharmacology of stress. *European Journal of Pharmacology, 463*(1–3), 235–272.

Chrousos, G. P. (2009). Stress and disorders of the stress system. *Nature Reviews Endocrinology, 5*(7), 374–381.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 385–396.

Cohen, S., Kamarck, T., & Mermelstein, R. (1994). Perceived stress scale. *Measuring stress: A guide for health and social scientists, 10*(2), 1–2.

Compas, B. E. (1987). Stress and life events during childhood and adolescence. *Clinical Psychology Review, 7*(3), 275–302.

Connor, K. M., & Davidson, J. R. T. (2003). Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depression and Anxiety, 18*(2), 76–82.

Dinges, D. F., Rider, R. L., Dorrian, J., McGlinchey, E. L., Rogers, N. L., Cizman, Z., & Metaxas, D. N. (2005). Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, Space, and Environmental Medicine, 76*(6), B172–B182.

Ganster, D. C., & Rosen, C. C. (2013). Work stress and employee health: A multidisciplinary review. *Journal of Management, 39*(5), 1085–1122.

Gao, Hua, Yüce, Anil, & Thiran, Jean-Philippe. (2014). *Detecting emotional stress from facial expressions for driving safety*. Paper presented at the 2014 IEEE International Conference on Image Processing (ICIP).

Ge, M.-W., Hu, F.-H., Jia, Y.-J., Tang, W., Zhang, W.-Q., & Chen, H.-L. (2023). Global prevalence of nursing burnout syndrome and temporal trends for the last 10 years: A meta-analysis of 94 studies covering over 30 countries. *Journal of Clinical Nursing, 32*(17–18), 5836–5854.

Giannakakis, G., Pediaditis, M., Manousos, D., Kazantzaki, E., Chiarugi, F., Simos, P. G., & Tsiknakis, M. (2017). Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control, 31*, 89–101.

Gormally, B. M. G., & Romero, L. M. (2020). What are you actually measuring? A review of techniques that integrate the stress response on distinct time-scales. *Functional Ecology, 34*(10), 2030–2044.

Habersaat, S., Abdellaoui, S. H., & Wolf, J. M. (2021). Social desirability, stress and health in police officers: Preliminary results. *Policing: An. International Journal, 44*(2), 213–229.

Hassard, J., Teoh, K. R. H., Visockaite, G., Dewe, P., & Cox, T. (2018). The cost of work-related stress to society: A systematic review. *Journal of Occupational Health Psychology, 23*(1), 1.

Henderson, H. A., & Wachs, T. D. (2007). Temperament theory and the study of cognition–emotion interactions across development. *Developmental Review, 27*(3), 396–427.

Hoegen, Rens, Gratch, Jonathan, Parkinson, Brian, & Shore, Danielle. (2019). *Signals of emotion regulation in a social dilemma: Detection from face and context.* Paper presented at the 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII).

Hoque, M. E., McDuff, D. J., & Picard, R. W. (2012). Exploring temporal patterns in classifying frustrated and delighted smiles. *IEEE Transactions on Affective Computing, 3*(3), 323–334.

Jabon, M., Bailenson, J., Pontikakis, E., Takayama, L., & Nass, C. (2010). Facial expression analysis for predicting unsafe driving behavior. *IEEE Pervasive Computing, 10*(4), 84–95.

Järvelin-Pasanen, S., Sinikallio, S., & Tarvainen, M. P. (2018). Heart rate variability and occupational stress—Systematic review. *Industrial Health, 56*(6), 500–511.

Kappen, M., & Naber, M. (2021). Objective and bias-free measures of candidate motivation during job applications. *Scientific Reports, 11*(1), 21254.

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation, 15*(3), 235.

Klein, E. M., Brähler, E., Dreier, M., Reinecke, L., Müller, K. W., Schmutzer, G., & Beutel, M. E. (2016). The German version of the perceived stress scale–psychometric characteristics in a representative German community sample. *BMC Psychiatry, 16*, 1–10.

Konturek, P. C., Brzozowski, T., & Konturek, S. J. (2011). Stress and the gut: Pathophysiology, clinical consequences, diagnostic approach and treatment options. *Journal of Physiology and Pharmacology, 62*(6), 591–599.

Koolhaas, J. M., Bartolomucci, A., Buwalda, B., de Boer, S. F., Flügge, G., Korte, S. M., & Palanza, P. (2011). Stress revisited: A critical evaluation of the stress concept. *Neuroscience & Biobehavioral Reviews, 35*(5), 1291–1301.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394–421.

Krstajic, D., Buturovic, L. J., Leahy, D. E., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics, 6*(1), 1–15.

Kuipers, M., Kappen, M., & Naber, M. (2023). How nervous am I? How computer vision succeeds and humans fail in interpreting state anxiety from dynamic facial behaviour. *Cognition and Emotion, 37*(6), 1105–1115.

Langston, V., Gould, M., & Greenberg, N. (2007). Culture: What is its effect on stress in the military? *Military Medicine, 172*(9), 931–935.

Lazarus, R. S., Speisman, J. C., & Mordkoff, A. M. (1963). The relationship between autonomic indicators of psychological stress: Heart rate and skin conductance. *Psychosomatic Medicine, 25*(1), 19–30.

Lerner, J. S., Dahl, R. E., Hariri, A. R., & Taylor, S. E. (2007). Facial expressions of emotion reveal neuroendocrine and cardiovascular stress responses. *Biological Psychiatry, 61*(2), 253–260.

Liao, Wenhui, Zhang, Weihong, Zhu, Zhiwei, & Ji, Qiang. (2005). *A real-time human stress monitoring system using dynamic bayesian network.* Paper presented at the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)-workshops.

Liu, B., Qiao, K., & Lu, Y. (2021). The relationship between perceived stress, state-trait anxiety, and sleep quality among university graduates in China during the COVID-19 pandemic. *Frontiers in Psychology, 12*, Article 664780.

Low, Z. X., Yeo, K. A., Sharma, V. K., Leung, G. K., McIntyre, R. S., Guerrero, A., & Nguyen, L. H. (2019). Prevalence of burnout in medical and surgical residents: A meta-analysis. *International Journal of Environmental Research and Public Health, 16*(9), 1479.

Martin, R. P., Lease, A. M., & Slobodskaya, H. R. (2020). *Stability of temperament traits and profiles* (pp. 151–158). Temperament and Children: Profiles of Individual Differences.

Metaxas, D., Venkataraman, S., & Vogler, C. (2004). Image-based stress recognition using a model-based dynamic face tracking system. In *Paper presented at the International Conference on Computational Science*.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin, 130*(5), 711.

Naidu, P. R., Sagar, S. P., Praveen, K., Kiran, K., & Khalandar, K. (2021). *Stress recognition using facial landmarks and CNN (Alexnet)*. Conference Series: Paper presented at the Journal of Physics.

Pflanz, S., & Sonnek, S. (2002). Work stress in the military: Prevalence, causes, and relationship to emotional health. *Military Medicine, 167*(11), 877–882.

Reiche, E. M., Vissoci, N., Vargas, S. O., & Morimoto, H. K. (2004). Stress, depression, the immune system, and cancer. *The Lancet Oncology, 5*(10), 617–625.

Rimini-Doering, M., Manstetten, D., Altmueller, T., Ladstaetter, U., & Mahler, M. (2001). *Monitoring driver drowsiness and stress in a driving simulator* (Paper presented at the Driving assessment conference).

Rozanski, A., Blumenthal, J. A., & Kaplan, J. (1999). Impact of psychological factors on the pathogenesis of cardiovascular disease and implications for therapy. *Circulation, 99*(16), 2192–2217.

Sandi, C. (2013). Stress and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(3), 245–261.

Schaufeli, Wilmar, & Enzmann, Dirk. (1998). The burnout companion to study and practice: A critical analysis: CRC press.

Schiweck, C., Piette, D., Berckmans, D., Claes, S., & Vrieze, E. (2019). Heart rate and high frequency heart rate variability during stress as biomarker for clinical depression. *A systematic review. Psychological medicine, 49*(2), 200–211.

Sharma, N., Dhall, A., Gedeon, T., & Goecke, R. (2014). Thermal spatio-temporal data for stress recognition. *EURASIP Journal on Image and Video Processing, 2014*, 1–12.

Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine, 108*(3), 1287–1301.

Soliemanifar, O., Soleymanifar, A., & Afrisham, R. (2018). Relationship between personality and biological reactivity to stress: A review. *Psychiatry Investigation, 15*(12), 1100.

Taelman, Joachim, Vandeput, Steven, Spaepen, Arthur, & Van Huffel, Sabine. (2009). *Influence of mental stress on heart rate and heart rate variability*. Paper presented at the 4th European Conference of the International Federation for Medical and Biological Engineering: ECIFMBE 2008 23–27 November 2008 Antwerp, Belgium.

Toet, A., Bijlsma, M., & Brouwer, A.-M. (2017). Stress response and facial trustworthiness judgments in civilians and military. *SAGE Open, 7*(3), 2158244017725386.

Van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *The Australian Journal of Advanced Nursing, 25*(4), 40–48.

van der Kooij, K. M., & Naber, M. (2019). An open-source remote heart rate imaging method with practical apparatus and algorithms. *Behavior Research Methods, 51*(5), 2106–2119.

Van Rossum, G. (1995). *Python reference manual*.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*(2), 197–210.

Vrijkotte, T. G. M., Doornen, V., Lorenz, J. P., Geus, D., & Eco, J. C. (2000). Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension, 35*(4), 880–886.

Williams, P. G., Smith, T. W., Gunn, H. E., & Uchino, B. N. (2011). *Personality and stress: Individual differences in exposure, reactivity, recovery, and restoration* (pp. 231–245). The handbook of stress science: Biology, psychology, and health.

Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T. P., & Sahebkar, A. (2017). The impact of stress on body function: A review. *EXCLI Journal, 16*, 1057.