Sensing Stress

Stress detection from physiological variables
in controlled and uncontrolled conditions

Jacqueline Wijsman

Sensing Stress

Stress detection from physiological variables in controlled and uncontrolled conditions

J.L.P. Wijsman

Members of the Graduation Committee

prof.dr. P.M.G. Apers (chairman and secretary)	University of Twente
prof.dr.ir. H.J. Hermens (promotor)	University of Twente
prof.dr. M.M.R. Vollenbroek-Hutten	University of Twente
prof.dr.ir. P.H. Veltink	University of Twente
prof.dr. D.K.J. Heylen	University of Twente
prof.dr.ir. P.P.C.C. Verbeek	University of Twente
prof.dr.ir. C.A.A. Van Hoof	KU Leuven
prof.dr. J.C.N. de Geus	VU University Amsterdam

SENSING STRESS

STRESS DETECTION FROM PHYSIOLOGICAL VARIABLES IN
CONTROLLED AND UNCONTROLLED CONDITIONS

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof.dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op donderdag 4 december 2014 om 16.45 uur

door

Jacqueline Louise Petronella Wijsman
geboren op 7 januari 1986
te Rotterdam

Dit proefschrift is goedgekeurd door de promotor:

prof.dr.ir. H.J. Hermens

Contents

1

Introduction

On the homepage of the academic database Scopus, the default example search term is *"heart attack" AND stress*. When this search is executed, 421 documents are found that mention both heart attack and stress in title, abstract or keywords (August 2014). Since the 1970s, the number of documents published per year has rapidly increased and more than half of the documents were published in the past 10 years. This increase is an indication of the topicality of the relation between heart attacks and stress.

1.1 Stress

More than a half of all European workers considered work-related stress to be common in their workplace [47]. Stress is the second most frequently reported work-related health problem in Europe. The total cost of mental health disorders in Europe is estimated to be 240 billion €/year, of which 43% consists of direct costs such as medical treatment and 57% consists of lost productivity, including sick leave [46].

The general concern about work stress is a recent phenomenon. From the second half of the 20th century, research on work stress received increasing amounts of attention. In the late 20th century, it was recognized that work stress is a threat to employees' health and productivity. During this period, work became more service-oriented and knowledge-based. This required knowledge about performance under stressful conditions. Hence, experts began to recognize the importance of psychological well-being of employees, for example through recognition and opportunities for personal development. The intention of these experts was to improve organizational well-being and productivity by improving individual well-being [164].

Gradually, the effects of stress on health became evident and researchers realized

that prolonged stress could cause illness. By the late 1990s, psychological health of the employees had become a major concern. Numerous work stress models were developed and many organizations launched intervention programs [164].

There is no unique definition of stress, but three approaches to the definition and study of stress were identified [33]:

1. The *engineering approach* sees stress as being caused by the environment. The amount of work load and demand are objectively measurable factors that cause stress.

2. The *physiological approach* sees stress as the physiological response to certain stimuli. This is a general and non-specific response to any stressor.

3. The *psychological approach* focuses on the interaction between a person and his environment when studying stress.

The engineering approach and the physiological approach form the earlier approaches to defining and measuring stress. These approaches have suffered from much criticism, though. A major point of critique towards the engineering approach was its emphasis on environmental factors. These factors do not always have the same effect on, for example, performance or comfort. For the physiological approach, it is mainly the non-specificity that was criticized. It was proven that the stress response is not identical among various stressors, but varies for different types of stressors. Moreover, both approaches ignore individual differences.

Current stress research is mainly based on the psychological approach, which focuses on the relations between a person and his environment and the cognitive and emotional processes that are involved in these relations. There are two interaction theories that evolved from the psychological approach that explain the interaction between a person and his environment [33]. One interaction theory is the person-environment fit: if a person and his environment (e.g. work) do not fit, this situation can lead to stress. Another interaction theory is the demand-control model: if the environment is highly demanding and the person working in the environment has low control this can also lead to stress.

A distinction can be made between short-term and chronic (long-term) stress. Short-term stress follows from incidental stressful conditions such as a short period of high demand to finish a crucial report. Chronic stress is caused by a long-term imbalance between person and environment and/or demand and control.

1.2   Effects of Chronic Stress

Chronic stress induces negative psychological and social effects, as well as physiological effects. For example, stress causes adrenaline and cortisol levels to rise. Chronically elevated levels of these hormones negatively influence cardiovascular health because of their effects on blood pressure and cholesterol levels. The hormone imbalance also suppresses the immune system.

Mental and physical health problems that are usually associated with chronic (work) stress are, for example [33, 99]:

- Bronchitis
- Coronary heart disease or other cardiovascular diseases
- Headaches and migraine
- Musculoskeletal problems
- Diabetes
- Obesity
- Anxiety
- Burnout
- Depression
- Difficulty in concentrating
- Drug and alcohol abuse

These issues do not only have a personal impact, but they also have an impact on the organization and society due to poor performance at work, increased accident and injury rates and increased absenteeism [99]. Incidental, short-term, stress (e.g. to meet a deadline) is not harmful and often even beneficial for performance (except for a very traumatic stress experience that can result in post-traumatic stress disorder). Prolonged, chronic, stress however, results in various mental and physical symptoms and needs to be prevented. Therefore stress should be detected, monitored and managed by employers and employees. When frequent short-term stress is detected, job content can be changed to reduce workload, for example, and the transition to chronic stress can be prevented.

1.3   Measuring Stress

Timely detection of stress is essential and to achieve this, a personal stress management system would be useful. Such a system could measure stress and give feedback on a personal level. Effective feedback empowers the creation of awareness of stress level and the application of a personal treatment, if necessary.

An overview of the main processes that are involved in the development and effects of stress is shown in Figure 1.1. Based on the effects, one could think of methods to detect stress. Several options have been proposed for measuring stress to facilitate detection in an early stage. These options can be categorized in three approaches [125]:

1. Questionnaires

2. Biochemical measures

3. Physiological measures
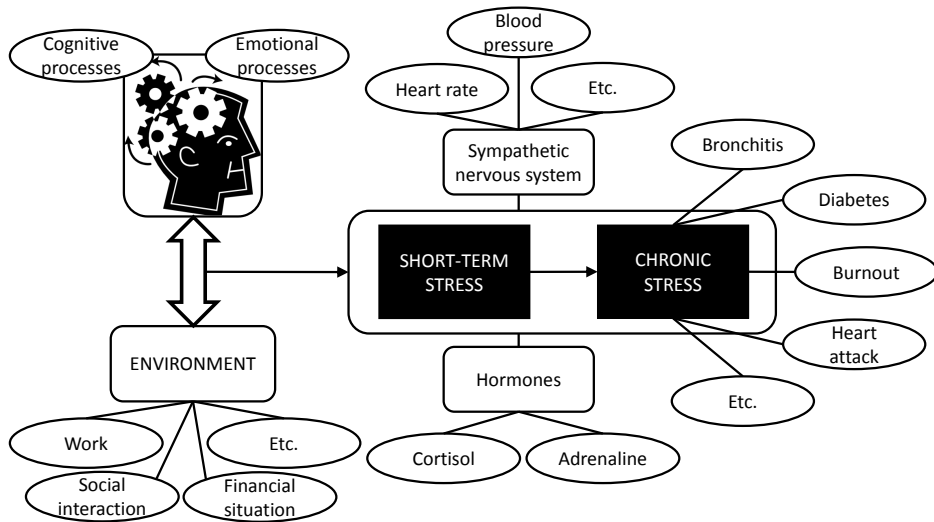
Figure 1.1: Overview of the main processes involved in the development and the effects of stress.

1.3.1 Questionnaires

Stress assessment questionnaires follow different approaches. One approach is to assess life events (both acute and chronic) that may have impact on stress experience. However, not everyone reacts in the same way to certain life events. Therefore, another approach is to assess perceived stress. In this way only the stress that is experienced by the subject is assessed. Stress levels can be assessed through a checklist or a personal interview. A checklist is cheaper and faster than a personal interview, but an interview assesses the life events and their timing more precisely [166].

Commonly used questionnaires to assess stress are the Life Stress Inventory, the Perceived Stress Scale and the Social Readjustment Rating Scale. Several questionnaires exist to assess post-traumatic stress disorder, anxiety and depression [125].

A drawback is that most of these tests are time-consuming for doctors, assistants and patients [125]. Also, questionnaires can be unreliable because they might not be filled in sincerely and may suffer from recall bias [40]. Hence, there is the need for more objective measures to assess stress.

1.3.2 Biochemical Measures

The main biochemical measures for stress are the levels of the hormones adrenaline and cortisol in various body fluids: blood, saliva, urine. Stress increases secretion

of these hormones and this increase is measurable in these body fluids. When using hormone levels to assess stress, one must be aware that the hormones are not released in a constant fashion and fluctuate during the day. Ideally, one would place an indwelling catheter to monitor the hormone concentrations over a 24-hour period. Obviously, this is often not possible and the assessment needs to be based on just a few measurements [125].

Adrenaline releases glucose to prepare the body for an appropriate response to a stressor. The adrenaline concentration in blood increases due to stress, but also due to general engagement in tasks, because adrenaline is a hormone that mobilizes energy when performance is needed. Results of studies investigating the relation between stress and adrenaline levels show increased adrenaline concentrations during work, but not necessarily during work *stress* [40]. This phenomenon is a drawback of measuring adrenaline levels to asses stress. Another drawback is that blood sampling is required for adrenaline measurement and it is often unwanted to perform such an invasive measurement.

Cortisol releases glucose and lipid acids and acts as a catalyst for adrenaline. Advantage is that the cortisol level can be determined from saliva, which does not require invasive techniques in order to take a sample. Cortisol is released in stressful conditions, but also shows a diurnal rhythm. Cortisol level peaks during awakening and decreases gradually during the day. This influence of time is substantially larger than the influence of stress on cortisol level. Therefore one should carefully control the sampling times when assessing cortisol levels. The results of studies investigating the relation between stress and cortisol levels are not conclusive and sometimes point in opposite directions [64]. Cortisol level increases following acute stress, but the effects of chronic stress are unclear. Chronic stress might even cause decreased cortisol levels [40].

1.3.3   Physiological Measures

The physiological effects of stress are expressed in both the central nervous system (CNS) and the peripheral nervous system (PNS). The brain and spinal cord are in the CNS and are connected to the other body parts by the PNS. This thesis is focused on the expression of stress through the PNS. The assessment of CNS activity through, for example, brain waves is out of scope.

Stress experiences cause rapid activation of the autonomic nervous system (ANS) in the PNS. The sympathetic part of the ANS becomes predominant over the parasympathetic part to prepare the body for action to eliminate the stressor. Hence, measuring the sympathetic activation gives insight in stress experience of a person. A number of noninvasively measurable physiological variables can be useful for this purpose. These variables are based on the following physiological measures:

- Heart rate
- Blood pressure

- Respiration
- Skin conductance
- Muscle activation
- Skin temperature

It was shown already in numerous studies that these physiological variables and mental stress are related (this will be elaborated upon in Chapter 2). However, it is not possible to model the relation between the physiological features and the various aspects of mental stress exactly, for a number of reasons. First, the physiological reaction to stress varies substantially among persons. On group level, differences were shown between subject groups with different levels of stress. However, individual reactions and physiological stress sensitivity vary tremendously among persons. Second, a physiological reaction is an indirect expression of mental stress. The translation from mental stress in the brain to peripheral physiological variables is influenced by numerous factors that influence the relationship between the two. That is, the subjective stress experience in the brain is translated by neuronal and hormonal links that produce the peripheral physiological signals that we measure. Based on the basic physiological stress reaction, classification between stress and rest conditions has been successfully applied numerous times. Yet, the indirect translation and differences among persons could be an explanation why correlations between (continuous) subjective stress levels and (continuous) physiological feature values have not been found often [40].

Still it is useful to measure physiological variables for stress assessment. The physiological reaction is the mechanism that causes stress-related physical complaints, and not the subjectively-experienced mental stress. Therefore the physiological measurements enable quantification of the impact of stress on the body [40].

1.4   Taking the Next Step

1.4.1   Limitations of Current Research

As discussed in the previous section, there have been various attempts at detecting mental stress. None of them showed results convincing enough to become a standard method for stress measurement. Probably the physiological measures are most promising for a practical solution to measure stress ambulatory. Questionnaires require user input and biochemical measures require invasive sampling. Physiological measures can be recorded with unobtrusive wearable devices.

Physiological measures have been most useful so far in comparing the effects of stress in large groups. A typical study would measure baseline physiological values in a group of subjects, apply a stressor, and measure the physiological response (e.g. [105]). Another approach that is commonly taken is to measure physiological variables in two different groups of subjects: one subject group experiencing (chronic) stress, the other group not experiencing this stress (e.g. [156, 165]). The outcome is the difference between the measured physiological values of the

two groups. In this way the two groups are compared and some insights can be obtained about the effects of stress.

Whereas most of the current studies were assessing stress reponses on group level, it should also be possible to assess the stress reaction within one person and for example compare the response to various conditions. In this way, personal stress responses can be determined and modeled. Such stress models would be useful for the development of a personal stress management system. Users of such a system can get feedback on their personal stress level and become aware of the situations that may cause stress. Through this awareness, changes in behavior can be triggered that lower stress level and chronic stress can be prevented.

Most studies published in literature were conducted in controlled, laboratory environments. However, a personal stress management system will need to work in real-world environments. There is no conclusive evidence that the stressors applied in controlled environments induce the same reactions as real-world stress [73].

Furthermore, the equipment that is used in laboratory environments is often not portable. Numerous ambulatory measurement systems have been developed and these devices show much potential for future developments in stress monitoring. Commercial heart rate monitors are widely available for sport exercise purposes. Most systems work with a chest belt, but some of the newest models work with a sensor on the wrist (e.g. Philips' Mio sensors). Some sensors have also been developed that can measure skin conductance ambulatory (e.g. Shimmer's GSR+ unit). A combination of these types of devices into one device would be a large step forward in the development of stress management systems. This illustrates the need for reliable, small and unobtrusive wearable physiological sensors to enable stress detection in real-world environments.

1.4.2 Goal

Considering the limitations of current research, the goal of the research described in this thesis is:

to assess the feasibility of constructing personal models for the relation between mental stress and physiological variables, for use in ambulatory stress management systems.

The steps that were taken to achieve this goal are discussed next.

1.4.3 Thesis Overview

First, the literature on the topic of stress detection from physiological variables was reviewed. This review provides insight in what was done on the topic so far and which approaches seem promising. The outcomes can be found in **Chapter 2**.

Next, a stress experiment was designed and conducted. Subjects were exposed to three different stressors in a controlled environment. During the stress exposure, physiological variables were recorded. Some insight was acquired from this study into the relevant, ambulatory recordable, physiological variables for stress detection. The procedure of the experiment and the results from the trapezius muscle surface EMG recordings are described in **Chapter 3**. The results of the analysis on a larger set of physiological signals (EMG, ECG, respiration, skin conductance) collected in the same study are discussed in **Chapter 4**.

To take the next step, physiological variables were recorded in a less controlled environment. The second experiment was conducted during a simulation training for medical students in which they had to respond to an emergency situation. The results show the effect of context conditions, especially physical activity, on the physiological variables. A third study was also conducted during simulation training with students. In this training the students had to perform simulated laporoscopic surgery. These two studies are described in **Chapter 5**.

The final study was designed and conducted in a truly natural setting. Physiological variables as well as reference measures were recorded during five working days while subjects, who were researchers, continued their normal daily work activities. The results reveal important clues on the feasibility of personal stress detection. This study is described in **Chapter 6**.

Finally, all insights from these studies are summarized and discussed and suggestions for future work are given in **Chapter 7**.

How Mental Stress Influences Physiological Variables: a Selective Review

Mental stress is a growing problem in society. To prevent stress-related health problems, it is crucial to assess mental stress levels and to take action when stress becomes too high and/or chronic. Noninvasively measurable physiological parameters are useful to measure stress in an unobtrusive way. In this review, we discuss how mental stress is associated with four noninvasively measurable physiological parameters: heart rhythm, blood pressure, respiration and skin conductance. We selected a representative set of 15 studies for each of the four parameters and analyzed the outcomes of these studies. Numerous features are extracted from these parameters. We identified five features that positively correlate to mental stress: heart rate, systolic and diastolic blood pressure, skin conductance level and number of skin conductance responses. We hypothesize that these features can be used in the development of a stress management system that can give adequate feedback to users about their stress levels.

2.1   Introduction

Stress is a major problem in society. In the US, one-third of employees report their job as stressful. The number of work days lost due to anxiety, stress and neurotic disorders is four times higher than the number of lost days due to other nonfatal injuries and illnesses [122]. In the European Union, more than 40 million individuals are affected by work-related stress. Stress is one of the most commonly

---

reported causes of occupational illness by workers and costs approximately 20 billion euros per year in lost productivity and medical expenses [68].

Walter Cannon introduced the term 'homeostasis': the "coordinated physiological processes which maintain most of the steady states in the organism". Cannon demonstrated that there are different responses to psychological and physical factors that disrupt homeostasis. Hans Selye was the first to introduce the term 'stress' in scientific and medical literature. A *stressor* is a stimulus that disrupts homeostasis. Stressors could be divided in four types: (1) physical stressors, (2) psychological stressors, (3) social stressors and (4) stressors that challenge cardiovascular and metabolic homeostasis. Stressors can be either acute or chronic. In contrast to Cannon, Selye defined stress as the nonspecific response of the body to any demand. Hence he emphasizes that the same reaction would follow any type of stressor [128]. Currently, many definitions exist of the concept 'stress'. We will focus on mental (psychological and/or social) stressors in this chapter and use the following definition of stress: an imbalance between stimuli and psychosocial resources [153].

The two major physiological systems that respond to stress are the hypothalamic-pituitary-adrenal (HPA) axis and the autonomic nervous system (ANS), especially the sympathetic part of the ANS [112]. Activation of these systems results in, among other things, increased attention, cardiac output and catabolism, accelerated respiration and redirection of blood flow to brain, heart and muscles [162]. Chronic activation of the stress response systems can cause various health problems, such as hypertension, stroke, diabetes, obesity, autoimmune and inflammatory disorders and depression [128, 162].

There are various ways to detect the physiological stress responses. One option is to measure the levels of certain hormones typically involved in the stress reaction, such as cortisol and epinephrine. These hormone levels are usually determined in blood, saliva or urine. Another option is to measure the ANS responses, such as heart rhythm and respiration. An advantage of the latter option is the possibility to measure the parameters noninvasively (from the outside of the body).

Considering the increasing number of people affected by (work-related) stress, long-term monitoring of stress becomes a concrete option in the modern world. Long-term monitoring will enable people to be aware of and potentially manage their stress level, and forms a promising method for preventing long-term health problems caused by chronic stress. Such monitoring must be done noninvasively and as unobtrusively as possible. We envision a wearable sensor system (body area network) consisting of small, low-power, wireless sensors that measure various physiological signals. An automated stress detection algorithm could run on such a system and provide (real-time) feedback to the user.

To design such a system, it is necessary to know which noninvasively measurable physiological parameters and features are relevant for mental stress detection. Over the years, the relationship between various types of stressors and the corresponding physiological responses has been intensively studied. Some review papers

have been published in this field, covering for example the neuroendocrine effects of stress [117], the HPA axis response [51], the effects on the brain and immune system [112], the cardiovascular system [45], the changes of the acute stress response with various chronic psychosocial factors [28] and negative health effects of chronic stress [71]. A number of papers has been published on the effect of *emotions* on autonomic nervous system activity, see for example [86]. In her paper, Kreibig discusses the relevant literature on the response of the autonomic nervous system (cardiovascular, respiratory and electrodermal responses) on various emotions.

To the best of our knowledge, there is no review that discusses the associations of general mental stress to multiple noninvasively measurable physiological parameters. To fill this gap, we conducted a literature study on this topic and report the results in the current article.

The research question for our study is: *Which physiological features, that can be measured noninvasively, are the most suitable to detect mental stress?*

We chose to select four common physiological parameters and discuss the performance of these parameters in stress detection research. The parameters can all be measured noninvasively by body area network sensors for use in a wearable automatic stress detection system towards real-time and continuous stress monitoring. We discuss the procedure for selecting the parameters and relevant literature in Section 2.2. The relevant results from this literature on the relation between stress and physiology are reported in Section 2.3. These findings are discussed further in Section 2.4 and conclusions are drawn in Section 2.5.

2.2 Methods

2.2.1 Considerations Regarding Study Selection

We based our search approach on two aspects of our research question: (1) mental stress and (2) physiological parameters. We listed the most relevant keywords for both aspects (stress keywords: *arousal, stress, psychosocial load, mental demand, mental load, distress, strain, effort-reward imbalance*; physiological keywords: *physiol*, electrocardiogra*, ECG, heart rate, interbeat interval, electromyogra*, EMG, respiration, electrodermal activity, skin conduct*, galvanic skin response, electroencephalogra*, EEG, voice, impedance cardiogra*, ICG, blood pressure, temperature*) and performed an initial search in Scopus [42]. Results had to include at least one stress keyword and at least one physiological keyword. The search yielded a considerably large number of 1.8 million results. Next to the studies we intended to find, the results also included for example studies on mechanical stress, studies on animals and studies on a number of diseases. We attempted to exclude these non-relevant results by excluding subject areas such as agricultural and materials sciences, by excluding keywords such as 'animals' and by excluding source titles that clearly involved non-relevant topics such as materials and me-

chanics. However, we did not manage to limit the number of results to reasonable proportions. The lowest number of results we could achieve was still over 15,000.

In our search, we did not set restrictions on, for example, study design, selection of subjects or type of stressor, and therefore we found significant variety on these aspects of the studies included. We could have limited our scope to a specific study design, subject group and stress elicitation method to rule out this phenomenon. However, we would have excluded a significant number of relevant studies by setting strict inclusion criteria. We consider the variety of studies as a strength of our review. By including diverse studies, we show the relevance of considering physiological changes caused by stress in broad applications.

Due to the relatively high heterogeneity among search results it was complicated to compare the results of the studies in a quantitative way. Therefore we chose to include a broad range of studies and to describe the similarities and differences among the selected studies in a more qualitative way. Next to this, based on the high number of search results that were retrieved, we concluded that the number of relevant studies that fit our research question was too high to include them all in a systematic review. Therefore we included a limited number of relevant studies for every parameter considered. One way to identify the relevant studies that have high impact in a research field is to consider the number of times a study was cited by other authors. A limitation of this selection method, however, lies in the fact that recent studies did not have the chance to gather many citations yet. Therefore we took a hybrid approach for selecting the studies to include in our review. For every physiological parameter we considered, we selected the ten most cited studies and five relevant studies that were published in the period 2003–2013. We identified the relevant studies by using the relevance ranking of Scopus, which is based on how many times the keywords appear in the record and on the location of the search terms in the record (title/abstract/body).

2.2.2 Search Strategy

We executed our literature search in Scopus during the period December 2012 to January 2013. Included studies needed to mention at least one of the stress keywords mentioned in Section 2.2.1 in the title. Furthermore, studies needed to mention the physiological aspect in title, abstract or keywords. We used subsets of the physiological keywords mentioned in Section 2.2.1, depending on the specific physiological parameter under consideration.

We limited the results to document type 'article' and included only articles written in English. Initially, we retrieved many non-relevant studies. To overcome this issue, we limited our search further by excluding keywords, subject areas and source titles related to non-relevant fields.

We chose to narrow our scope by selecting only the four most common physiological parameters by ranking them on the number of results per parameter. The parameters that were initially included were noninvasively measurable parameters that are

commonly used in mental stress measurement: blood pressure, electroencephalography (EEG), electromyography (EMG), heart rhythm, impedance cardiography (ICG), respiration, skin conductance and voice. The four parameters with the highest number of search results were included in our review: heart rhythm, blood pressure, respiration and skin conductance.

We set three inclusion criteria to select studies for our review:

1. The study should concern mental (i.e. psychological and/or social) stress. Studies that involved physical stress, such as exercise, were not included.

2. The subject group had to include healthy individuals. We think that the physiological effects of stress on healthy individuals should be studied first. Once these effects are known, the effects of stress on unhealthy individuals can be studied. Healthy control groups of studies involving patient groups were included.

3. As our research question says, we want to identify the specific physiological features that are influenced by mental stress. Therefore we only included studies that reported on changes due to stress in individual features. This criterion implies that studies that used a set of features for classification, for example, were not included if they did not report on changes in individual features.

For every included physiological parameter, we sorted the search results by the number of citations. We selected the top ten most cited studies that matched our inclusion criteria. Then, we limited the search to the period 2003–2013 and sorted the results by relevance. We selected the five most relevant studies that matched our inclusion criteria from this list.

2.3 Results

2.3.1 Heart Rhythm

It is worth noting that four of the ten most cited papers that comply with our criteria were published during the past ten years and have already achieved citation numbers high enough to be included in the ten most cited papers in the field. An overview of the studies selected is shown in Table 2.1. An overview of the five papers from the past ten years that received a high-relevance ranking based on the search criteria is shown in Table 2.2. Only heart-rhythm-related features are mentioned.

Study Design

We found considerable differences among the designs of the studies analyzed. Eleven times there was one subject group of which all members went through

Table 2.1: Ten most cited heart-rhythm-related studies, sorted from most cited to least cited

Reference (citations)	Stress elicitation method	Subjects	Study design	Features analyzed	Results
[146] (168)	TSST	38 men, 27 women	Three repetitions of TSST. Only subjects that showed cortisol response during TSST 1 were considered 'responders' and invited for TSST 2 and 3.	HR	= HR reactivity to TSST 1 between responders and nonresponders. HR response in all three TSST sessions. From TSST 1 to TSST 3, ↓ HR responses by approximately 18%.
[165] (158)	2 workdays and 1 non workday	109 middle aged male white-collar workers	Groups divided by high and low effort-reward imbalance and high and low overcommitment.	HR, RMSSD	→ work and leisure HR, ↓ RMSSD for high imbalance.
[63] (150)	Driving in different environments	9 subjects	One subject group. Comparison of three different driving conditions. Stress metric based on coded video recordings.	Normalized mean and variance of HR, LF/HF	97.4% classification among three different stress states. HRV and HR 2nd and 3rd best correlation with stress metric.
[156] (147)	Color-word interference and mirror tracing	12 subjects exposed to stress and 8 control subjects	Stress group and control group.	HR	↑ HR in stress group. = HR in control group.
[5] (140)	Nonverbal math task, mirror tracing task, Stroop Color-word interference task	20 men, 22 women	Group of women and men. Three different stress tasks.	HR	↑ HR. Women showed greater HR responses than men during mirror tracing and Stroop tasks, but no gender differences during math.
[121] (134)	TSST	24 men	One subject group. TSST and control condition on separate days with randomized sequence.	HR	↑ HR.
[105] (126)	Mental arithmetic, Stroop color word test	62 women	One subject group. Two different mental stress tasks.	HR	↑ HR.
[10] (116)	Mental arithmetic	12 subjects	One subject group. Mental stress aloud and silent.	RR interval, SDNN, LF, HF: absolute and normalized	↓ RR interval and ↓ HF, ↑ LF.
[27] (119)	Work stress	2769 subjects	Cumulative work stress, assessed at two times 1–5 years apart. Association with HRV 12–15 years later.	Log SDNN, log LF, log HF	Greater reports of work stress were associated with ↓ log SDNN, ↓ log LF and ↓ log HF.
[104] (105)	University examination	30 medical students	One subject group. Assessment shortly before university examination. Control condition 3 months later, during holiday.	LF, HF: absolute and normalized	↓ RR interval, ↓ HF, ↑ LF, ↑ LF/HF on stress day.

→ increase/higher, ↓ decrease/lower, = no change/no difference. HF: high frequency heart rate variability (0.15–0.4 Hz), HR: heart rate, HRV: heart rate variability, LF: low frequency heart rate variability (0.04–0.15 Hz), RMSSD: root mean square of successive differences, SDNN: standard deviation of interbeat intervals, TSST: Trier Social Stress Test.

Table 2.2: Five recent and relevant heart-rhythm-related studies

Reference (citations)	Stress elicitation method	Subjects	Study design	Features analyzed	Results
[59] (91)	Asked to give a speech on awakening the next morning	59 subjects	Stress group and control group, measured during sleep.	HF, LF/HF	↓ HF during NREM and REM sleep, ↑ LF/HF during NREM sleep in stress group compared to control group.
[36] (3)	Lab: aversive images; Real-world acute: First-time skydive	56 subjects, 33 artifact-free data sets analyzed	Two 5-min lab recordings separated by one ambulatory 24-hour recording. Recording during skydive afterwards.	LF, HF, LF/HF	LF/HF correlations between: neutral/image laboratory period vs. jump/second 30 minutes recovery; wake portions of 24-hour HRV vs. 15 minutes before jump/recovery; sleep vs. second 30 minutes recovery.
[141] (1)	Stroop test	10 female and 14 male students	One subject group, one stress task.	HR, RR interval, SDNN, CV, root mean square of standard deviation, PNN50, HRV index, TINN, stress index, VLF, (normalized) LF, (normalized) HF, LF/HF	↑ Mean HR, ↑ LF/HF, ↑ normalized LF, ↑ SDNN, ↑ CV, ↑ HRV index, ↑ TINN, ↓ mean RR, ↓ RMSSD, ↓ PNN50, ↓ HF, ↓ normalized HF, ↓ stress index, ↓ VLF, ↓ LF during Stroop test.
[147] (20)	Speech task	22 men, 28 women	Subjects slept at research center. Stress test the next morning.	Mean and SD of RR intervals, HF, LF, LF/HF, correlation dimension D2	↓ Correlation dimension D2, ↑ HR mean ↑ HR SD, ↓ LF power, ↑ HF power during stress.
[74] (0)	Ambulance alarms	20 ambulance professionals	7 days, 24 hours HR registration. Comparison between various phases of duty.	HR	↑ HR unrelated to physical effort during emergency alarm and response.

↑ increase/higher, ↓ decrease/lower. CV: coefficient of variance, HF: high frequency heart rate variability (0.15–0.4 Hz), HR: heart rate, HRV: heart rate variability, LF: low frequency heart rate variability (0.04–0.15 Hz), PNN50: percentage of heart beat intervals with difference in successive intervals greater than 50 ms, SDNN: standard deviation of interbeat intervals, TINN: triangular interpolation of heart beat interval histogram, VLF: heart rate variability in the very low frequencies

the same procedure. Two studies included a stress group and a control group. Schommer et al. [146] split their subject group into a "high responder" and a "low responder" group and repeated the same stressor three times in the high responder group. Vrijkotte et al. [165] split their subject group based on high and low effort-reward imbalance at work and high and low overcommitment at work. Despite these differences the studies all aimed to answer similar research questions regarding the influence of mental stress on heart rhythm and in that sense they are considered comparable.

Stress Elicitation

Nine studies reported on the effects of artificially induced laboratory stress: arithmetic tasks [5, 10, 105], mirror tracing tasks [5, 156], Stroop Color-word interference tasks [5, 105, 141, 156] and speech tasks [59, 121, 146, 147]. Three studies involved work stress [27, 74, 165], one driving stress [63] and one study involved examination stress [104]. Dikecligil and Mujica-Parodi [36] combined a laboratory stressor of negative images with a first time skydive as real-world acute stressor. Hall et al. [59] and Karlsson et al. [74] studied ECG during sleep. Still, despite differences in design of the studies, the various results were similar for comparable occasions when stress was measured. For example, Dikecligil and Mujica-Parodi, who compared laboratory induced and real-world acute stress, found that the heart rate variability reaction during real-world stress can be predicted from the reaction to laboratory stress.

Features

As stated in Section 2.1, stress activates the sympathetic part of the ANS. This sympathetic activation causes the heart rate (HR) to rise. As a result, the heart pumps the blood through the body faster and oxygen is delivered faster to the organs and skeletal muscles that are involved in eliminating the stressor.

HR, or the equivalent RR interval, is one of the most simple and straightforward measures of heart activity. HR is defined as the number of heartbeats per minute; RR interval is the time between consecutive heartbeats. All studies that reported the effect of stress on HR (11 studies) found an increase due to stress [5, 10, 74, 104, 105, 121, 141, 146, 147, 156, 165], indicating that increased HR is a reliable stress indicator.

Karlsson et al. [74] found the largest increase in HR. The HR reactivity to an ambulance alarm was 49 beats per minute (bpm). After this short-term initial HR peak, HR stayed elevated during the entire ambulance response, but this was a more moderate increase of 22.5 bpm. Schommer et al. [146] also found a large increase in HR. During the first Trier Social Stress Test (TSST), the HR increased with 30.0 bpm for the high responder group, and with 23.3 bpm for the low responder group. Moreover, the other study that involved a TSST by Nater et al. [121] also showed an increase of HR of about 30 bpm. As the TSST involves a

speech in front of an audience, it is likely that the subjects were standing up during the stress phase of this protocol, which might have contributed to the relatively large HR increases.

The other laboratory-based studies found more moderate HR increases of up to 13 bpm. An exception was the study of Bernardi et al. [10] who found a difference in HR of 17.7 bpm between rest and a silent mental arithmetic stress task. Although all their tests were carried out supine, the subject was performing the task "while writing the results on a blackboard held in front of the subject". This procedure probably involved some physical activity to hold the arms up to write, which might have contributed to the relatively large HR increase. When HR during reading aloud was taken as baseline and the same stress task was done aloud, HR increased with 5.3 bpm from baseline to stress. These values comply more with the ranges found in other studies.

Regarding the real-world stress studies, Vrijkotte et al. [165] found about 5 bpm higher HR at work for the high effort-reward imbalance group than for the low effort-reward imbalance group. Lucini et al. [104], on the other hand, found a difference of almost 16 bpm between a university examination day and a control day. This difference might have been caused by the fact that Vrijkotte's control group was working and, therefore, active although not stressed, while Lucini's subjects were completely at rest during the measurements.

Next to heart rate, there are various measures of heart rate variability (HRV), which indicate the variability in the heart rhythm, that change during stressful situations. HRV can be assessed in time domain by calculating, for example, the root mean square of the successive differences between heartbeats (RMSSD), like Vrijkotte et al. did [165]. They used RMSSD as an indicator of vagal tone and found a lower RMSSD value for subjects that indicated a high effort-reward imbalance at work.

HRV can also be assessed in frequency domain. HRV is typically evaluated as the power in a low frequency (LF, 0.04–0.15 Hz) and high frequency band (HF, 0.15–0.4 Hz). The LF band reflects both sympathetic and parasympathetic activity, whereas the HF band corresponds only to parasympathetic activity. The ratio of the power in these two frequency bands (LF/HF) therefore informs about the sympathetic modulation [59]. It is expected that LF and LF/HF increase, and HF decreases during exposure to a stressor. This was indeed found by four authors [10, 59, 104, 141].

Contradictory HRV results were found by Chandola et al. [27] (lower interbeat variance, LF and HF) and Schubert et al. [147] (increased standard deviation of interbeat intervals, LF and HF). Possibly the HF result of Schubert et al. was caused by the stress task, which involved speech. Speech influences respiration and respiration influences HF [59]. Chandola et al.'s decreased interbeat variance and HF were expected. However, it remains unclear why LF also decreased.

Finally, there are various other heart-rhythm-related features that were found to react to stress. However, these are not widely used (yet) and need further

investigation. Salahuddin et al. [141] analyzed a number of ECG derived features for "ultra short term analysis of heart rate variability". Coefficient of variance, HRV index and triangular interpolation of heartbeat interval histogram (TINN) increased during stress. Root mean square of standard deviation, percentage of heartbeat intervals with difference in successive intervals greater than 50 ms, stress index and very low frequency HRV decreased during stress. Schubert et al. [147] investigated the changes of the HR complexity measure correlation dimension D2 due to stress. They found a decreased D2 during short-term stress, which indicates a reduction in HR complexity and may represent a lower adaptability and fitness of the cardiac pacemaker. Furthermore, they found a significant negative correlation between chronic stress and D2.

Summary

It is clear that HR increases due to stress. However, some studies found large responses that might be partly caused by physical activity instead of mental stress only. HRV results are mostly consistent (decrease in time-domain features, decrease of HF, increase of LF and LF/HF), but some contradictory results were found.

2.3.2 Blood Pressure

The ten most cited papers reporting on the relation between mental stress and blood pressure (BP) are listed in Table 2.3. The five relevant recent papers discussing mental stress and BP are listed in Table 2.4. Even when studies involved various physiological measures, only details related to blood pressure are listed in the tables.

Study Design

As for the heart-rhythm studies, the BP studies show considerate differences among study designs. In nine studies, all subjects went through the same procedure, two studies included a stress group and a control group. Four times the subject group was divided into subgroups based on BP or job strain. Two studies included a long follow-up period of several years [145, 81].

Stress Elicitation

Nine studies elicited stress in a controlled environment. The stressors included a color-word interference test [5, 70, 105, 156], mirror tracing [5, 110, 156], mathematical tasks [5, 10, 26, 25, 70, 105], anticipation of a speech performance [127] and a video game [110]. Six studies reported on the influence of work stress on BP [81, 91, 143, 144, 145, 165].

Table 2.3: Ten most cited blood-pressure-related studies, sorted from most cited to least cited

Reference (citations)	Stress elicitation method	Subjects	Study design	Features analyzed	Results
[143] (222)	Working day	215 men	Divide subject group in hypertension and control group. Assess relation between job strain level and case-control status.	DBP	Job strain is a predictor of case-control status. Exposure to job strain increases likelihood of being classified as having hypertension.
[145] (178)	Two days of ambulatory monitoring; T1=0, T2=3 years	195 men	24-hour ABP on two occasions 3 years apart. Four groups identified: (1) no job strain at either assessment, (2) job strain at both times, and two crossover groups (3) and (4).	SBP, DBP	High job strain showed ↑ SBP and ↑ DBP at work and home and ↑ SBP during sleep. Chronic high job strain group (2) showed highest ABP levels, group (1) lowest, groups (3) and (4) were in between.
[144] (165)	Working day	264 men	24-hour ABP, job strain assessed by questionnaire.	SBP, DBP	High job strain showed ↑ SBP and ↑ DBP at work and home and ↑ SBP during sleep.
[127] (161)	Anticipation of a speech performance	27 men, 41 women	Baseline BP measure and BP measure after 10-minute speech preparation period (stress group) or 10-minute listening to emotionally neutral text (control group).	SBP, DBP	↑ SBP in stress group, ↓ SBP and ↓ DBP in control group.
[165] (158)	2 workdays and 1 non workday	109 middle aged male white-collar workers	Groups divided by high and low effort-reward imbalance and high and low overcommitment.	SBP, DBP	↑ work SBP and ↑ home SBP for high imbalance.
[156] (147)	Color-word interference and mirror tracing	12 subjects exposed to stress and 8 control subjects	Stress group and control group.	SBP, DBP	↑ SBP and ↑ DBP in stress group. = SBP and = DBP in control group.
[5] (140)	Nonverbal math task, mirror tracing task, Stroop Color-word interference task	20 men, 22 women	Group of women and men. Three different stress tasks.	SBP, DBP	↑ SBP, ↑ DBP. Men showed greater DBP responses than women during math and Stroop tasks, but no gender differences during mirror tracing.
[105] (127)	Mental arithmetic, Stroop color-word test	62 women	One subject group. Two different mental stress tasks.	SBP, DBP	↑ SBP, ↑ DBP.
[10] (117)	Mental arithmetic	12 subjects	One subject group. Mental stress aloud and silent.	SBP, DBP	↑ SBP, ↑ DBP.
[70] (112)	Color-word test and mental arithmetic	10 men	One subject group. Two different mental stress tasks, analyzed as if they were one.	SBP, DBP	↑ SBP, ↑ DBP.

↑ increase/higher, ↓ decrease/lower, = no change/no difference. ABP: ambulatory blood pressure, BP: blood pressure, DBP: diastolic blood pressure, SBP: systolic blood pressure

Table 2.4: Five recent and relevant blood-pressure-related studies

Reference (citations)	Stress elicitation method	Subjects	Study design	Features analyzed	Results
[26] (68)	Mental arithmetic (PASAT)	1458 subjects (792 women, 666 men)	BP measures at baseline and during task.	SBP, DBP	↑ SBP, ↑ DBP. No difference between sexes after adjustment for performance.
[25] (2)	Mental arithmetic (PASAT)	1196 subjects (645 women, 551 men)	BP measures at baseline and during task.	SBP, DBP	↑ SBP, ↑ DBP.
[110] (62)	Video game and mirror tracing task	2816 subjects	BP measures at baseline and during task.	SBP, DBP	↑ SBP, ↑ DBP.
[91] (40)	Normal working life	213 men	24-hour ABP. Division of subjects into job strain and no job strain group.	SBP, DBP	↑ SBP, ↑ DBP for men with job strain compared to men with no job strain.
[81] (4)	Economic crisis with low job control	218 men	Annual BP measurements before, during and after a 3-year economic crisis.	SBP, DBP	↑ SBP, ↑ DBP during crisis, returned to slightly above baseline after crisis and remained stable for next 5 years.

↑ increase/higher. ABP: ambulatory blood pressure, BP: blood pressure, DBP: diastolic blood pressure, PASAT: Paced auditory serial addition test, SBP: systolic blood pressure

Four of the studies on work stress monitored and analyzed ambulatory blood pressure for at least 24 hours. These studies revealed that work stress does not only influence BP at work, but also at home [91, 144, 145, 165]. This finding indicates that the physiological influence of work stress is maintained even beyond working hours.

Features

Two features are distinguished when measuring blood pressure: systolic blood pressure (SBP) and diastolic blood pressure (DBP). SBP is the peak blood pressure, reached when the left ventricle contracts and pumps blood in the aorta. DBP is the lowest blood pressure during a cardiac cycle, when the ventricle contraction is finished and the aortic semilunar valve closes. In stress situations, the sympathetic part of the ANS is activated. This causes the HR and stroke volume (SV) to rise, leading to a rise in cardiac output (HRxSV, the amount of blood pumped out of a ventricle in one minute). Next to this, sympathetic activation also leads to vasoconstriction (narrowing of blood vessels). The increase in cardiac output and the vasoconstriction both contribute to increases in SBP and DBP.

When looking at the absolute changes in BP induced by mental stress, it can be noticed that, in general, the controlled stressors induce larger responses than work stress. This was not the case for the study by Kjeldsen et al. [81], who

reported BP responses in their work-stress based study similar to responses to controlled stressors in other studies. However, it must be noted that the time between their measurements covered several years. BP tends to increase when people become older, so this increase might have contributed to the total BP increase, that was therefore probably not strictly caused by stress. Also Schnall et al. [145] reported larger BP differences between stress and no stress than the other work-stress studies. We could not identify where this difference might have come from.

Taking a closer look at the results of the laboratory studies, the large physiological response reported by Bernardi et al. [10] again attracts attention. Most of the laboratory-based studies found SBP increases in the range of 10–15 mmHg and DBP increases in the range of 5–11 mmHg. Bernardi et al. reported an increase in SBP of 25 mmHg and in DBP of 13.8 mmHg from rest to the stress task (silent mental arithmetic). Again, we believe that part of this large difference was caused by physical activity of the subject writing on a blackboard held in front of him while being in supine position. The differences between reading aloud and performing the mental stress task aloud were 7.5 mmHg for SBP and 7.0 mmHg for DBP. These values comply more with the ranges we found in other studies; the SBP increase is even a bit lower than in other laboratory stress studies. Also Jern et al. [70] (SBP +17.2 mmHg, DBP +16.2 mmHg) and Steptoe et al. [156] (SBP +16.4 mmHg, DBP +12.6 mmHg) reported relatively large BP increases. However, the subject groups of these studies consisted of only 10 and 12 subjects, respectively, which makes the results more uncertain than they would have been with a large subject group.

All work-related stress studies included large groups of over 100 subjects. Three of the laboratory-based studies included over 1000 subjects. The results from these large subject groups are reliable and confirm that work-related and laboratory stress actually increase BP levels. The existence of large studies, both in real-world working conditions and in the lab, could be resulting from the fact that BP can be measured in a very non-invasive and easy way. Fully automatic devices to measure BP easily from a cuff placed around the arm are widely commercially available. One drawback of BP monitoring lies in the fact that BP cannot easily be monitored continuously, and that monitoring is usually expensive. If cheap and noninvasive continuous BP measurement would be available, the dynamics of BP and the causes of BP changes could be studied in more detail.

Something else worth mentioning is the variety of types of stressors that induce BP increases. Short-term artificial stressors increase BP, but also long-term work-related stressors increase BP. The BP even remains elevated once a long-term stressor disappears. Kjeldsen et al. [81] found that it took three years following an economic crisis for BP to return to nearing the original baseline. Next to that, Landsbergis et al. [91] found that SBP at home of men who were employed for ≥ 25 years while being exposed to job strain for 50% of their work life was higher than SBP of men with no past exposure, independent of current exposure. This finding provides an indication that chronic stress might lead to long-term or even

permanent physiological changes.

Summary

Both SBP and DBP increase due to stress. These findings are supported by studies with large subject groups. Elevated BP can be maintained for years after a stressor has disappeared and therefore high BP does not have to be the result of a current stressor, but can also be a long-term effect of a stressor in the past.

2.3.3 Respiration

Table 2.5 lists the ten mostly cited papers reporting on the relation between mental stress and respiration. Table 2.6 lists the five relevant recent papers discussing mental stress and respiration. Even when studies involved various physiological measures, only details related to respiration are listed in the tables.

Study Design

The study designs of the selected studies were relatively standard for this kind of research. The majority of studies (12 studies) analyzed one subject group performing some type of stressor. Some authors made a comparison between two groups with high and low stress levels. Dishman et al. [37] compared subjects with high and low perceived emotional stress during the past week, Fenz and Epstein [49] compared novice and experienced parachutists, while Cacioppo et al. [24] compared a group of caregivers with a control group of noncaregivers.

Something that we should pay attention to when studying respiration-related features is the fact that in some studies the subjects were speaking during the measurements. As far as we are aware, six of the selected studies involved speaking [8, 24, 115, 136, 139, 155]. Speaking is especially important when the baseline or control condition did not require subjects to speak, but performing the stress task did require speaking. Obviously, speaking influences the breathing pattern so comparing between a silent rest condition and a stress condition during which subjects are speaking may lead to results that are not comparable and consequently a wrong interpretation. We should keep this in mind when evaluating the results of these six studies.

Stress Elicitation

Nine of the selected studies involved stress elicitation methods in a laboratory environment. The stressors varied from mental arithmetic [10, 137, 139], to divided attention and risk taking tasks [67], the TSST [136], IAPS slides [137], visual matrix problem solving [155], mirror drawing [155], speech [24, 155], response

Table 2.5: Ten most cited respiration-related studies, sorted from most cited to least cited

Reference (citations)	Stress elicitation method	Subjects	Study design	Features analyzed	Results
[63] (162)	Driving in different environments	9 subjects	One subject group. Comparison of three different driving conditions. Stress metric based on coded video recordings.	Normalized mean and variance of respiration, respiration spectral power density in 0–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4 Hz	97.4% classification among three different stress states. Respiration did not correlate well with stress metric.
[10] (120)	Mental arithmetic	12 subjects	One subject group. Mental stress aloud and silent.	Respiration variability in LF and HF, respiration frequency and relative change in minute ventilation	↓ Frequency of breathing, so ↓ HF and ↑ LF. Slight ↓ minute ventilation.
[104] (106)	University examination	30 medical students	One subject group. Assessment shortly before university examination. Control condition 3 months later, during holiday.	Respiration rate	= respiratory rate on stress day compared to control day.
[37] (101)	No specific method, just daily life	52 men, 40 women	Subjects filled out CPPS, rested supine 15 min, stood 5 min, then 1 min respiration measurement.	Respiration rate, minute ventilation	= respiration rate, = minute ventilation between low and high scores on PSS.
[67] (95)	Divided attention task and risk taking task	20 women	Baseline recording followed by two stress tasks and rest period.	Respiration rate	↑ respiration rate.
[137] (89)	Mental arithmetic and IAPS pictures depicting injuries, mutilation and corpses	8 men, 16 women	Order of stress elicitation methods and film clips inducing different emotions was counterbalanced.	Respiration rate, tidal volume, minute volume	Respiratory parameters were not significantly affected by the stress tasks.
[155] (70)	Visual matrix problem-solving, mirror drawing, speech	64 men, 68 women	All subjects performed all three stress tasks.	Respiration rate, tidal volume	↑ respiration rate, ↑ tidal volume.
[139] (64)	Mental arithmetic	7 men, 19 women	Mental arithmetic in between or before two physical stressors.	Respiration rate, minute volume	↑ respiration rate, ↑ minute volume.
[49] (40)	Parachute jump	10 novice and 10 experienced sport parachutists	Measure on control day and before boarding, throughout flight and shortly after landing.	Respiration rate	↑ respiration rate of novices compared to experienced, except during control day and shortly after landing.
[24] (34)	Caring for spouse with progressive dementia, math and speech	27 women caregivers, 37 women noncaregivers	Caregiver and control groups. Each subject performed two stress tasks.	Respiration rate	↑ respiration rate due to stress, but no differences between caregiver and control groups.

↑ increase/higher, ↓ decrease/lower, = no change/no difference. CPSS: Cohen's Perceived Stress Scale, HF: high frequency respiration variability (>0.15 Hz), IAPS: International Affective Picture System, LF: low frequency respiration variability (0.03–0.14 Hz)

Table 2.6: Five recent and relevant respiration-related studies

Reference (citations)	Stress elicitation method	Subjects	Study design	Features analyzed	Results
[136] (3)	TSST	7 men, 12 women	Quiet sitting measurements pre- and poststress, TSST in between.	Tidal volume (V_T), total respiratory cycle time (T_{TOT}), inspiratory time (T_I), expiratory time (T_E), respiratory timing (T_I/T_{TOT}), inspiratory flow (V_T/T_I), minute ventilation (V_E), inspiratory and expiratory volume instability ($V_{I\,RMSSD}$ and $V_{E\,RMSSD}$), ribcage inspiratory contribution to V_T (%RC$_I$), inspiratory and expiratory thoracoabdominal asynchrony	↓ T_{TOT}, ↓ T_E, ↓ T_I, ↓ $V_{I\,RMSSD}$, ↓ $V_{E\,RMSSD}$ from prestress to poststress quiet sitting. ↑ V_T, ↑ V_T/T_I, ↑ $V_{I\,RMSSD}$, ↑ $V_{E\,RMSSD}$, ↑ %RC$_I$, ↑ asynchrony during stress. ↓ T_I, ↓ T_E, ↓ T_I/T_{TOT} during stress.
[63] (162)	*See Table 2.5*				
[8] (10)	Mental arithmetic during driving, driving in noisy environment and in narrow alley	4 men	All subjects performed all stressors. Alternating driving with and without stress.	Respiration rate	↓ respiration rate in narrow alley. Mixed results during mental arithmetic and noisy environment.
[115] (20)	Auditory prompt–verbal response n-back task in driving simulator	59 men, 62 women	Reference driving; driving during three levels of difficulty of task.	Respiration rate	↑ respiration rate from baseline to difficulty level 1 and from level 1 to level 2. No significant increase from level 2 to level 3.
[124] (9)	Two-choice reaction-time test	9 men, 35 women	Baseline, 1 hour test, 30 min rest.	Respiration rate	↑ respiration rate due to stress, but ↓ respiration rate from beginning to end of task.

↑ increase/higher, ↓ decrease/lower. TSST: Trier Social Stress Test

[115] and reaction-time tasks [124]. Most of these tasks are commonly used in stress studies and we have seen many of them already in Sections 2.3.1 and 2.3.2.

There were six studies that involved real-world stress. Two of them involved driving [8, 63]. The other studies involved a university examination [104], parachute jump [49], caring for a spouse [24] and general perceived stress [37]. We see that the effects of both short-term and chronic real-world stressors were studied. The two studies evaluating chronic stress did not find any respiration differences caused by the stress. The studies evaluating short-term real-world stress found mixed results.

Features

Experiencing a stressor activates the sympathetic part of the ANS and this prepares the body for action to eliminate the stressor. One of the effects of this activation is the need for more oxygen. This increased oxygen demand is answered by rapid and deep breathing.

The most common way to measure respiration is by placing one or two belts around the upper body and measure the expansion. Sometimes the belts are calibrated against a spirometer for respiratory volume. Baek et al. [8] extended this idea and used the seat belt in a car to measure chest expansion. Another option is to place a thermistor at nose and mouth [124] or to use a flow meter [37, 137].

The most popular respiration feature is the respiration rate. It can be measured relatively easily with a belt around the chest that detects the chest expansions and contractions caused by breathing, and by calculating the number of breaths per minute from the signal. Seven of the twelve studies that investigated respiration rate found an increase due to stress [24, 49, 67, 115, 124, 139, 155]. However, there were also some studies that reported no effect of stress on respiration rate [37, 104, 137]. Baek et al. [8] and Bernardi et al. [10] reported a decrease in respiration rate. Baek et al. found a decrease while driving in a narrow alley, Bernardi et al. found a decrease during silent and aloud mental stress. For the aloud mental stress case, the speaking could have influenced the breathing rate. For the other cases, we suggest that muscle activity in the upper body might have interfered with chest expansion and consequently decreased the breathing rate. While driving in a narrow alley, subjects might have had an increased muscle tone in the upper body to facilitate better control of the steering wheel. During the study of Bernardi et al., the subjects were in supine position while writing on a blackboard held in front of them. This procedure probably involved activation of the muscles in the arms and chest.

If we consider only studies in which subjects were not speaking during stress conditions, fewer studies that found a respiration rate increase are included. So although we believe speaking could influence the results, excluding the studies that involved speaking does not result in more consistent outcomes.

As a second feature, we found minute volume or minute ventilation (the total

volume of gas entering the lungs per minute), measured by calibrated belts or a flow meter, as possible stress indicator. This feature was found to increase [139], decrease [10], or show no effect [37, 137]. However, Roth et al. [139], who found an increase, had their subjects talking during the stress condition. We think that speaking might have influenced the outcome in this case.

The last respiration feature that was used in multiple studies is the tidal volume (the volume of gas entering the lungs in one breath), also measured by calibrated belts or a flow meter. The tidal volume was found to increase in stressful conditions [136, 155], or show no effect [137]. Also here, one of the studies involved talking and therefore the increase in tidal volume that Ritz et al. [136] found might have been caused by the talking.

Summary

Respiration rate often increases under the influence of stress, but some contradictory results were found. The results are even more contradictory for the other features. For the real-world study designs and stressors we considered in this section, we found no convincing evidence that stress influences respiration. Furthermore, it should be taken into account that speech influences the respiration signal and therefore it is not as suitable as other physiological parameters for measuring stress in circumstances involving speech.

2.3.4   Skin Conductance

The ten most cited papers investigating the influence of stress on skin conductance (SC) are listed in Table 2.7. The five recent studies on SC and stress that we identified are listed in Table 2.8. Only details related to SC features are listed in the tables.

Study Design

An interesting observation among the SC studies is that three studies offered their subjects different versions of the same stressor. Healey and Picard [63] included three different driving conditions in their test, Breier et al. [17] exposed their subjects to controllable and uncontrollable noise and Sibley et al. [151] exposed their subjects to balance perturbations while standing on a platform at ground level and 1.6 m high. Two authors divided their subject group in two or more groups that each received a different version of the stressor: Renaud and Blondin [133] included three different forms of pacing the Stroop test and Suzuki et al. [158] threatened part of their subject group with an electric shock in case of bad performance in the stress test.

Another observation is that two studies involved a stress task followed by another task and investigated the influence of the first task on the second task. Jackson et

Table 2.7: Ten most cited skin-conductance-related studies, sorted from most cited to least cited

Reference (citations)	Stress elicitation method	Subjects	Study design	Features analyzed	Results
[63] (162)	Driving in different environments	9 subjects	One subject group. Comparison of three different driving conditions. Stress metric based on coded video recordings.	Normalized mean and variance of SC, number of SC responses, sum of response magnitudes, sum of response durations, sum of estimated areas under responses	97.4% classification among three different stress states. Mean SC best correlation with stress metric.
[67] (95)	Divided attention task and risk taking task	20 women	Baseline recording followed by two stress tasks and rest period.	Mean SC, SC range, number of SC fluctuations	↑ mean SC, ↑ SC range, ↑ number of fluctuations from baseline to stress tasks.
[137] (89)	Mental arithmetic and IAPS pictures depicting injuries, mutilation and corpses	8 men, 16 women	Order of stress elicitation methods and film clips inducing different emotions was counterbalanced.	log SCL	↑ SCL during mental arithmetic, = SCL for IAPS slides.
[17] (82)	Controllable and uncontrollable loud noise, anagrams	6 men, 4 women	Controllable and uncontrollable noise sessions on separate days. First noise, followed by rest and mental task.	Mean SCL, number of fluctuations per minute	↑ SCL and ↑ number of fluctuations in uncontrollable noise compared to controllable noise throughout noise and mental task conditions.
[100] (73)	Self-disclosing speech or electric shock on wrist	69 men	Subjects received only one of the stressors. Countdown 360 seconds to speech or shock.	SCL	↑ SCL to both initiation of countdown and to stressor.
[155] (70)	Visual matrix problem-solving, mirror drawing, speech	64 men, 68 women	All subjects performed all three stress tasks.	Skin resistance responses	↑ number of skin resistance responses.
[134] (67)	Classify words into noun or adjective class	8 men, 13 women	Two stress blocks.	SCL	↑ SCL.
[139] (64)	Mental arithmetic	7 men, 19 women	Mental arithmetic in between or before two physical stressors.	SCL, number of SC fluctuations	↑ SCL, ↑ number of SC fluctuations.
[69] (59)	Job speech and mental arithmetic	25 men and 22 women in stress group, 25 men and 23 women in control group	Stressor, 60 min recovery, followed by fear conditioning.	Range-corrected SC response magnitude	Men in the stress group showed larger SC responses to the fear conditioning than men in the control group, but this result was not found for women.
[133] (57)	Stroop test	51 men	Three subject groups received either self-paced, slow-paced or fast-paced Stroop test.	Number of SC responses per minute	No significant difference between control and Stroop tasks. No difference between groups. ↓ number of SC responses over time (habituation).

↑ increase/higher, ↓ decrease/lower, = no change/no difference. IAPS: International Affective Picture System, SC: skin conductance, SCL: skin conductance level

Table 2.8: Five recent and relevant skin-conductance-related studies

Reference (citations)	Stress elicitation method	Subjects	Study design	Features analyzed	Results
[32] (17)	Obstacle avoidance during car driving	100 men	Stroop test, followed by driving test in which a car was suddenly pulled out on the road. SC only measured during driving.	Normalized SCL, Mhomic Perturbation Duration	SCL increased more for subjects who successfully avoided the obstacle and they had longer Mhomic Perturbation Duration.
[158] (9)	Arithmetic task	40 men	Punishment threat for some subjects in second arithmetic session increased distress.	SCL: square root and change score from baseline	$\rightarrow$ SCL for punishment group in session 2 compared to session 1 and $\rightarrow$ SCL compared to non-punishment group.
[132] (1)	PASAT, disgust and fear related IAPS pictures, noise, threat of losing money	80 women	5 min stress in which all elements were embedded.	SCL, NSCF	$\rightarrow$ SCL, $\uparrow$NSCF.
[2] (5)	Teaching using technology	3 male and 6 female teachers	Teachers in classroom observed and videotaped. Interviews after class.	SCL	Some correspondence between stress and SCL, but sometimes $\rightarrow$ SCL without stress or stress without $\rightarrow$ SCL.
[151] (4)	Perturbation leading to postural instability where unsuccessful reaction could lead to fall from 160 cm	5 men, 5 women	Subjects stood on edge of platform at ground level and 160 cm.	EDL, EDR latency, EDR amplitude, time-to-peak EDR	$\rightarrow$ EDL at high height. Nonsignificant trend of shorter EDR latency in high height.

$\rightarrow$ increase/higher. EDL: electrodermal level, EDR: electrodermal response, IAPS: International Affective Picture System, NSCF: Number of skin conductance fluctuations, PASAT: Paced auditory serial addition test, SCL: skin conductance level

al. [69] exposed part of their subject group to a job speech and mental arithmetic task (similar to the TSST, but without a live audience) before fear conditioning for both this stress group and a control group. The stress group showed larger SC responses to the fear conditioning than the control group. Collet et al. [32] had their subjects complete a Stroop test followed by a car driving task. They found that performance on the Stroop test corresponded to performance in the driving task.

Stress Elicitation

Twelve studies involved artificial laboratory stressors. The stressors showed great variety and included mental arithmetic tasks [69, 132, 137, 139, 158], delivery of a speech [69, 100, 155], pictures from the International Affective Picture System (IAPS) [132, 137], loud noise [17, 132], divided attention and risk taking tasks [67], anagrams [17], electric shock [100], visual matrix problem-solving [155], mirror drawing [155], classification of words [134], Stroop test [133], threat of losing money [132] and perturbation leading to postural instability [151].

Merely two studies evaluated SC during real-world stressors: Al-Fudail and Mellar [2] during teaching and Healey and Picard [63] during driving. Collet et al. [32] also had their subjects perform a driving test, but the subjects drove on a private circuit so we do not consider this test as 'real-world'.

Features

There are three types of sweat glands on the body: eccrine, apocrine and apoeccrine. The eccrine sweat glands are innervated by sympathetic nerves and are involved in emotional responses. SC changes with changing electrical properties of the skin caused by sweat gland activity. SC is often measured on the palms and soles, because these areas show the highest density of eccrine sweat glands [39]. In our selection of studies, most SC signals were measured from the hand, which seems to be the most convenient location for SC measurement. Two studies used the sole of the foot to measure SC [63, 132].

The two basic properties that can be determined from SC are the skin conductance level (SCL), measured in μS, and the fluctuations (SC responses or other measures of fluctuation) of the SC signal.

The SC feature that was used most often in the selected studies is SCL. It was found in ten studies that SCL increases in mental stress conditions [17, 63, 67, 100, 132, 134, 137, 139, 151, 158]. Looking at the absolute changes in SCL levels, the increase normally falls in the range of 1–6 μS. This is a relatively large range and, as with some other physiological features, the responses vary among people. An exception is the study of Rief et al. [134], who only found an increase of around 0.4 μS. We believe that this small increase might have been caused by the stressor: classifying words into either noun or adjective class. The stressor was developed

for patients with somatization syndrome and hypochondriasis and was probably not very stressful for the group of healthy control subjects.

The second most often used feature is the number of SC responses. Although different authors use different criteria for an SC response, the overall results are comparable: the number of SC responses increases in mental stress conditions [67, 132, 139, 155]. The only exception is the study by Renaud and Blondin [133], who did not find a significant difference in number of SC responses per minute between the control condition and Stroop test. They did find, however, an increase from baseline to both tasks. They suggest that this increase might be caused by emotional strain related to task performance, regardless of having color-word interference.

All other SC-related features that were used were related to responses and fluctuations of the SC signal. A feature directly related to the fluctuations is the variance of the SC signal (no change reported) [63]. Features related to response amplitude are SC range (increases due to stress) [67] and response magnitude (increases and no response found in stress) [63, 69, 151, 158]. Features related to the timing of the responses are the response duration, Mhomic Perturbation Duration, latency and time-to-peak (no significant changes due to stress reported) [32, 63, 151]. A feature that combines both amplitude and duration of the responses is the sum of the areas under the responses, evaluated by Healey and Picard [63] (no change reported).

Summary

We can state that both SCL and the number of SC responses increased under the influence of mental stress. Other SC features are less popular in stress research and their variability in stress is thus less clear. These features need further investigation before a statement can be made regarding their reactivity to mental stress.

2.4 Discussion

2.4.1 Main Findings

An overview of the results of features reported in at least three of the selected studies is listed in Table 2.9. We identified five noninvasively measurable physiological features from four physiological parameters that are commonly measured and show a clear correlation to mental stress: heart rate, systolic blood pressure, diastolic blood pressure, skin conductance level and number of skin conductance responses. These five features are all relatively easy to determine. This might be one of the reasons that studies often used these features.

Some other features did not show a correlation to mental stress as clear as these five features. For example, heart rate variability did not show up as a convincing

Table 2.9: Overview of main findings for features reported in at least three studies

Feature	Studies	↑	↓	=	Not reported
DBP	15	14	0	1	0
SBP	14	14	0	0	0
Respiration rate	12	7	2	3	0
SCL	12	10	0	1	1
HR	11	11	0	0	0
Number of SC responses	7	5	0	1	1
HF	6	1	4	0	1
LF	5	3	1	0	1
LF/HF	5	2	0	1	2
Minute volume	5	1	1	3	0
SC response magnitude	4	1	0	2	1
SDNN	3	2	0	1	0
Tidal volume	3	2	0	1	0

↑ increase/higher, ↓ decrease/lower, = no change/no difference

parameter for stress measurement. This was unexpected, because HRV should react to the sympathetic activation that is induced by stress. However, in some studies, speech can confound the results because speech influences breathing, which influences HRV. HRV measures can still be useful in stress detection in situations where no speech is involved. More research in controlled conditions is needed to investigate the exact relation between stress and HRV.

Some other features did not show convincing results but still suggested a relation with stress: respiration rate, skin conductance features other than SCL and SC responses. These features need more research so we can find out how exactly they react to stress. The other features related to respiration, except for respiration rate, just do not seem to show a distinct reaction to mental stress.

2.4.2 Challenges

Stress research remains challenging, although we have now identified the most promising physiological parameters in this field. One reason is that there is no adequate gold standard for measurement of stress level available. Some authors verify the effect of a stressor by self-report questionnaires. This is better than not using a reference measure at all, but is still a weak measure. Subjects might hesitate to express their true feelings to the experimenter or might not be conscious of their stress level and express physiological reactions that they are not aware of.

Another reference measure that is used sometimes is the cortisol level in saliva or another body fluid. Cortisol levels increase under the influence of stress and a subject cannot consciously influence this measure. However, cortisol is influenced by many factors besides stress, such as the diurnal rhythm: the cortisol level increases sharply after waking up and decreases during the day. Besides that, there is a delay between the onset of a stressor and an increase of cortisol level, making causal relationships between stressor and cortisol level rise difficult to identify. These issues make it difficult to check if stress was successfully elicited in a study. Because of the lack of a gold standard for stress level, we cannot judge which of the studies found valid results and which actually did not elicit stress successfully, but some studies might not have found the expected results because stress was not successfully induced.

Although we found relations between physiological features and mental stress, this does not imply that a change in physiology always corresponds to a change in mental stress level. Some activities may result in similar autonomic responses and therefore interfere with mental stress detection. Only when all other factors are controlled, a physiological change is related to a change in mental stress level. In practice, it is impossible to control all confounding factors when doing stress research, especially in real-world measurements. Therefore it is recommended to analyze multiple physiological parameters at the same time to get an overall picture of the physiological responses of the body.

It is interesting to see the diversity of stress elicitation methods. Researchers have been creative in designing methods to induce stress in test subjects. We think this is useful, because stress in daily life comes in many different forms. We still found similarities in the main results of the studies and therefore conclude that most stress elicitation methods induced similar physiological responses and thus seem to be valid methods to induce stress.

Although the laboratory stressors and the real-world stressors showed similar reactions, we still found differences in response amplitude for some features. We found larger BP responses to the laboratory stressors than to the real-world stressors. For respiration rate, we found relatively convincing influences of laboratory stress, but mixed results for the real-world stressors. We could not investigate the difference between laboratory and real-world stressors for SC, because there were only two studies involving SC and real-world stress and these studies did not mention quantitative results. Still, the laboratory stressors seem to induce different reactions from the real-world stressors, indicating that there are differences in the experience of laboratory versus real-world stressors. The nature of stressors in laboratory conditions is often much more acute than the real-world stressors that are studied. The real-world stressors often involve work stress, which is of a longer-term nature than the laboratory stressors. Although there are similarities between laboratory stress responses and real-world stress responses, it is of high importance to continue the transition from controlled (laboratory) to uncontrolled (real-world) environments in stress research to confirm if the laboratory findings can be transferred to real-world conditions.

2.4.3   Limitations

We recognize that our study has some limitations. First of all, we chose a subset of physiological parameters by selecting the four parameters with the highest number of results from our search in Scopus. By doing so we aimed at including the most common parameters in our analysis. We cannot be certain that this method actually resulted in selection of the most common parameters and we should keep in mind that we might have missed relevant results from parameters that were not included in our analysis. However, we identified the parameters with the highest number of search results, which are often used in stress research and are likely to be already accepted in the field. If there are other relevant parameters for stress detection, these parameters have not yet been used often in stress studies and still need to prove their credibility.

Furthermore, we chose to only include a subset of all relevant studies in our review. There are many relevant studies on the topic of changes in physiological parameters due to mental stress. As was explained in Section 2.2, in our opinion, it was not feasible to include all relevant studies on this topic. We aimed at including the most influential and relevant studies in the field, but we cannot be certain that we succeeded in doing so. Studies with null effects are less likely to be published than studies that show the expected effects. Moreover, if these studies are published, they are less likely to be cited often and therefore we might have missed them in our review.

Our conclusions are based on 15 studies per physiological parameter. This is a small number, considering that these studies focus on different types of stress and stressors. Although the diversity is a strength for the features for which consistent findings were found, the results are difficult to interpret when there are inconsistencies.

Notwithstanding these limitations, we believe that we were able to identify the physiological features that are most responsive to mental stress. Considering the large number of publications on the topic, our analysis helps to form an overview of the available research and most important outcomes of these studies. These conclusions form a starting point for future stress research and for the development of stress measurement systems and algorithms.

2.4.4   Future Work

In the future, more research is needed on the relation between mental stress and some physiological parameters, as stated above. These parameters might be related to stress, but the results found so far are not convincing enough yet. These parameters include the features related to HRV, some SC features and respiration rate. Regarding respiration rate, researchers found different outcomes, indicating that respiration rate can go up or down due to stress. Some persons might increase their breathing rate in certain stressful situations, while other persons or the same persons during other types of stressors might hold their breath and therefore

decrease breathing rate. Another option is that the respiration pattern becomes more irregular in stressful situations due to a combination of holding breath and fast breathing to take up the amount of oxygen needed by the body.

A challenge for the future is to find a way to deal with the individual differences in physiology. Physiology during rest as well as the physiological stress reaction vary among people [3]. Ideally, a stress detection algorithm should be developed that can learn the characteristics of the person that is using it. This way the algorithm can adapt to the individual. It might be needed to do a baseline measurement before starting to use the algorithm or even a standardized stress protocol to learn about the user's specific stress response. The person specific information can be used to select a subset of physiological features to use for stress measurement in a certain user, or to assign different weights to different features depending on the user. This way a robust stress detection algorithm can be developed that takes multiple features into account and combines all information in a reliable personal stress measure.

Now that we have identified the most suitable physiological parameters for mental stress detection, more research is needed on how to use these parameters in real-world stress situations. Next to the studies that were already conducted in real-world environments, new studies are needed that measure subjects for longer periods. Only then can the long-term physiological effects of stress be measured.

In order to get useful data from these long-term studies, it is necessary to be aware of context during measurement. Not only stress, but for example also physical activity, consumption of coffee, temperature and humidity can affect physiological parameters. It is necessary to measure parameters related to these factors and compensate for them when detecting stress.

Another point of attention is the occurrence of artifacts in the physiological signals due to motion and other uncontrollable circumstances when measuring in real-world environments. Sensors should be designed in such a way that these artifacts are diminished. Also, clever algorithms are needed for artifact removal and robust pattern recognition.

It would be beneficial to implement an algorithm that can calculate stress level in real time. When the user gets real-time feedback on his or her stress level, it is probably easier to change behavior than when the feedback is provided later, after offline analysis of the data. Then the user might not be able to recall the situation and the feelings that were present at that time. This could enable the start of a new approach for treatment in everyday life conditions instead of present stress training in an artificial environment. The feedback mechanism however should be carefully thought through. It is crucial to tell stressed individuals to calm down without increasing their stress levels even more. Additionally, privacy of data is also important. When measuring people's stress level, it is not desirable that this information is available freely. Therefore data should always be transferred over a secured connection. Also the feedback to the user should be delivered privately. The user should be able to choose with whom to share his or her stress information.

2.5 Conclusions

We identified five noninvasively measurable physiological features that are most suitable for detecting mental stress: heart rate, systolic blood pressure, diastolic blood pressure, skin conductance level and number of skin conductance responses. When developing an algorithm for reliable stress level estimation in real-world situations, variability of contexts and individuals should be taken into account. Such an algorithm should take both information from various physiological parameters and context information as inputs. It should be flexible and should be able to adapt to various contexts by including or discarding features based on context information, or by compensating for context factors. The algorithm should also have the option to be personalized to a certain user. Furthermore, a stress detection system should work in real time and provide adequate feedback to its users.

3

Trapezius Muscle EMG as Predictor of Mental Stress

Stress is a growing problem in society and can cause musculoskeletal complaints. It would be useful to measure stress for prevention of stress-related health problems. An experiment is described in which EMG signals of the upper trapezius muscle were measured with a wireless system during three different stressful conditions: a calculation task (the Norinder test), a logical puzzle task and a memory task. The latter two tests were newly designed and aimed at creating circumstances that are similar to work stress. Amplitudes of the EMG signals were significantly higher during stress compared to rest (+2.6% of reference contraction level) and relative time with EMG gaps was lower during stress (-14.3% of time). Also, mean and median frequencies were significantly lower during stress than during rest (-8.6 and -8.8 Hz, respectively). EMG amplitude increased not only from rest to stress conditions, but also during stressful conditions and decreased during relaxation periods. EMG features correlated with subjectively indicated stress levels (correlations of 0.32 with RMS and -0.32 with relative gaptime). The results indicate that EMG is a useful parameter to detect stress. Together with other physiological sensors, EMG sensors can be included in a wireless system for ambulatory monitoring of stress levels.

3.1 Introduction

Low back pain and pain in the neck or shoulders are symptoms that are commonly seen not only in physically demanding jobs, but also for example in office jobs and people working in supermarkets [105]. Causes for these problems are not only monotonous and repetitive work; especially psychosocial factors such as low job satisfaction, lack of autonomy and variation at work, high workload and work pressure have influence on the development of musculoskeletal disorders [105, 135, 142]. It was shown that increased EMG activity can be caused by mental stress in computer work. This increased EMG activity causes musculoskeletal symptoms of the upper extremities [142]. The increasing prevalence of muscle problems is becoming an important economic burden for modern society [135].

Stress is known to activate the sympathetic nervous system [162]. One of the consequences is the elevation of muscle tone, sometimes resulting in shivering [109]. This muscle tone elevation could thus be a predictor of mental stress level.

A lot of research has already been done on the detection of stress from physiological parameters influenced by activation of the sympathetic nervous system. Next to muscle activity, for example heart rate, heart rate variability, galvanic skin response and pupil diameter are found to reflect stress [65, 130, 174]. We focused on detecting stress from muscle activity because of the possible benefits in prevention of musculoskeletal disorders.

Several investigations are reported on the effect of mental stress on muscles [85, 92, 93, 105, 135, 142]. All experiments used the ElectroMyoGram (EMG) from the upper trapezius muscle, from which various features were extracted to quantify muscle activity. The Stroop color word test [157] and different mental arithmetic tasks are the most popular ways to induce mental stress. Previous results indicated that these tests successfully induced stress, which resulted in changes in EMG signals of the trapezius muscle. These changes in EMG included an increase of amplitude and a decrease in the amount of gaps (short periods of relaxation) that were recorded. Both are indications of elevation of muscle activity, caused by the stress tasks.

A system that can measure stress levels based on muscle activity will create interesting applications in real-life situations. Final goal is to design an easily wearable wireless system that can measure real time stress levels. A possible application of such a system could be stress prevention at work.

Before a system that can measure real time stress levels in daily life situations can be made, we need to take an easier case for development of the system. Therefore, our study is done in a controlled laboratory environment. In such an environment, there is no interference from physical activity or environmental factors. If changes are measured in EMG, they are very likely to be caused by stress and not by other factors influencing muscle activity.

The Stroop color word test and mental arithmetic tests can be very stressful, but do not simulate situations that are encountered in real life. Therefore, the first

goal of this study is to propose and validate a new protocol for stress induction in situations mimicking daily work. The second goal of this study is to confirm that stress can be detected from amplitude and temporal features of the EMG of the upper trapezius muscle, as was shown before for this particular muscle [85, 92, 93, 105, 135, 142]. Thirdly, two features of the frequency spectrum are analyzed to investigate if stress influences frequency components of the EMG. Fourth, it is investigated if there are trends over time in the EMG signals. Finally, the data is checked for correlations between perceived stress level and EMG features.

The test protocol, signal acquisition system, data processing, methods for feature extraction and statistical analysis are presented in Section 3.2. Results of the protocol validation and the analysis of the trapezius EMG signals during the tests, in comparison to rest periods, are presented in Section 3.3. The results are discussed in Section 3.4. Finally, conclusions are drawn in Section 3.5.

3.2 Methods

3.2.1 Subjects

A total number of 30 subjects were recruited for the experiment. Subjects had to be of age 18 or older and English-speaking. Subjects with muscle diseases, cardiac problems, or mental disorders such as depression or anxiety were excluded.

The ages of the subjects were in the range 19–53 (mean = 33.1; SD = 7.87); 25 subjects were male and 5 female. All subjects were right-handed. The subjects had 14 different nationalities.

3.2.2 Test Environment

The test was performed in a quiet room with as few distractions as possible. During the experiment, that was performed on a PC, the experimenter could see the test interface on a laptop screen, whereas an extra screen was available for the subject. A cabinet was placed in between the subject and the experimenter, so they could not see each other. The subject sat on a chair behind a desk and had a mouse and keyboard available to answer the questions and do the tests. A sound amplifier was connected to the laptop. The subject could hear the sounds involved in the experiment through headphones connected to the amplifier. A video camera was set up in the room, suggesting that the subject would be video taped during (part of) the test. A picture of the subject area in the experiment room is shown in Figure 3.1.

Figure 3.1: Subject area in the experiment room.

3.2.3 Experimental Protocol

The subject answered some general questions first, and filled in the perceived stress scale (PSS) questionnaire. Then he performed a reference contraction, followed by exposure to three different stress conditions. In between these conditions, resting periods were scheduled to make sure that the conditions did not influence one another. Furthermore, a questionnaire had to be completed before and after each condition. A schematic overview of the experimental protocol is shown in Figure 3.2.

The entire experimental protocol was implemented in a MATLAB interface in which the experimental conditions were executed one after each other automatically. Automated event marking allowed accurate mapping of the recorded signals on the different conditions. The different experimental conditions are explained in the paragraphs that follow.

General Questions and PSS The subject was asked about age, gender, dominant hand and nationality. He then filled in the perceived stress scale (PSS) [31], which provides an indication of his general stress state.

Rest The rest condition was a period of two minutes in which the subject had to sit still and relaxed with his left hand in his lap. During this period, the subject had to type in a number on the keyboard every 10 seconds with his right hand. In this way, the physical activity from the stress conditions was imitated, without any mental stress induced. The subject heard Pachelbel's Canon in D, which is an orchestral piece often used in commercially available 'stress-reduction' tapes [4].

Self-report Questionnaire As a reference measure, self-report questionnaires were used. The use of self-reports as reference for stress induction was reported in other studies as well [105, 135, 142]. After each rest period and
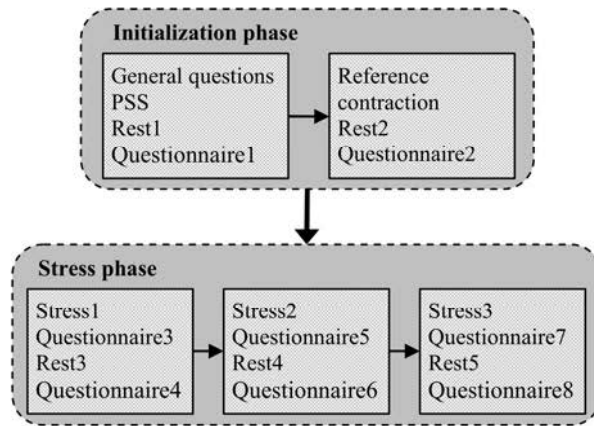
Figure 3.2: Schematic overview of the protocol of the stress experiment.

stress test, the subject had to fill in a questionnaire about his mood state. Visual analogue scales (VAS) were used to let the subject indicate his emotional state by answering 10 questions about 10 different discrete emotions. One of these emotions was "stressed". For each emotion, the scale ranged from "not at all" to "extremely".

Reference Contraction A reference voluntary contraction (RVC) was performed by the subject, for normalization of the EMG amplitude afterwards. The subject was instructed to raise his arms at a sign, while looking straight ahead. The arms were abducted 90° in the frontal plane, with the elbows fully extended, wrists straight and palms down. The arms were held horizontally aside. After 20 seconds, a second sign was given and the subject was told to bring down his arms.

Stress Tests Three stress tests were done during the experiment. The order of the three tests was randomized among the subjects to minimize crossover effects. This means that the order of the three blocks in 'stress phase' in Figure 3.2 changed among subjects.

Test 1, the Norinder test [103], was a calculation test that is commonly used in stress experiments [105, 85]. This test functioned as a 'reference' test to induce mental stress. The test was done under time pressure. The subject had 2:30 min to complete 27 calculations. When an error was made, a red screen appeared, a buzzer sounded and the subject had to calculate the result again. A countdown timer was running while the subject was performing the test. The color of the timer bar faded from yellow to red. Beeps sounded at 0:30, 1:00 and 1:30 min. When time was almost up, the program started beeping every two seconds from 2:00 to 2:15 min and it beeped every second during the last 15 seconds.

Test 2 was designed to only induce mental stress and consisted of five mental challenging tasks, like logical puzzles from an IQ test. The subject had 3:00 min to solve the puzzles. This test was also done under time pressure, with the timer bar and beeps.

Test 3 was designed to induce both mental and psychosocial stress. The subject was shown a very long row of pictures that formed a path. The subject was allowed to have a look at the pictures for one minute, while a timer bar was running again, and then the pictures disappeared. The subject now had to recall from his memory as many situations as he could remember, beginning with the first picture and mention them all in the right order. If a mistake was made or if the subject did not remember any more pictures, his turn was over. In the next turn, the pictures that the subject mentioned correctly were discarded. The row of pictures started at the first picture the subject could not remember anymore. Three turns were done in total. Psychosocial stress was induced by the subjects being told that they were divided in teams and that the next team member would start from where the current subject ended. The team had to try to come further on the path than the other teams. To invoke extra social stress, a video camera was set up in the room and the subject was told that he would be recorded during this test for analysis after the experiment. It was emphasized that the results of the different teams would be published after the experiments had ended.

During all three stress tests, the subject heard BBC news fragments through his headphones to distract him from the tasks and make it harder to fulfill the tests.

Reward At either test 2 or test 3, depending on which one came first, the subject could earn a present. Before the present-winning test started, the subject was told that if he finished the test within a certain amount of time, he would get a reward. If he would not manage to finish within this amount of time, he would be punished. After finishing the present-winning test, the subject was told that he had completed the task well enough to receive the present. At test 2 or test 3, whichever occurred last, the subject was told that he would lose his reward from the first present-winning test if he did not perform well on this test.

3.2.4   Procedure

The subject was asked to sit down and two pregelled Ag/AgCl electrodes were placed on both shoulders in the middle of the imaginary line between the acromion and vertebra C7 as described in the recommendations from SENIAM [148]. Inter-electrode distance was 25 mm and a reference electrode was placed on the spinous process of C7. Then, the EMG electrodes were connected to two wireless sensor nodes; one for the signals from the left shoulder and one for the right shoulder. Each sensor node features a proprietary single channel low-power ASIC for bio-potential amplification and filtering [172]. The wireless sensor nodes used for this

study are based on the body area network platform developed within imec [20].

A chest belt for electrocardiography (ECG) and respiration monitoring and a hand sensor for measuring skin conductivity were placed as well [20]. Next to the EMG, only the ECG signal is used in this study.

Once the subject was equipped with all sensors, a visual signal check was performed to ensure that the sensor nodes were communicating correctly and that the signal quality was good.

After this check, the experiment was started by executing a MATLAB interface in which the experiment was implemented. From then on, the whole experiment proceeded automatically for about 40 minutes, until the subject finished the last questionnaire. After this, all measurement equipment was switched off and the electrodes were removed.

3.2.5   EMG Recording and Signal Processing

The EMG was measured bipolar using pregelled Ag/AgCl ARBO ECG electrodes (model H124SG 24 mm diameter) by Tyco healthcare, at a sampling frequency of 1000 Hz.

All signal processing was done with MATLAB. The EMG signals were filtered with a bandpass filter, passing frequencies from 20 to 450 Hz. After that, the signals were filtered using notch filters at 50, 100, 150, 200, 250 and 350 Hz, to remove residual power line interference.

ECG signal components in the EMG signals of the left trapezius muscle were removed by using a method based on that of Spalding et al. [152]. The simultaneously recorded ECG signals were used to locate the heart beats. A template of the average heart beat artifact was constructed and subtracted from the EMG signal every time a heart beat occurred. The algorithm used by Romero et al. [138] was slightly modified to be able to apply it to remove the ECG components from the EMG signals.

3.2.6   Feature Extraction

An overview of all features that were extracted from the EMG signals can be found in Table 3.1. All feature extraction was done with MATLAB.

The reference contraction was processed following a procedure reported by Krantz et al. [85]. The root mean square (RMS) values for segments of 1s of the middle 15s of the reference contraction were calculated. The mean value of these 15 segments was taken as the RMS reference value. It was set to 100% and all other calculated values were interpreted in relation to this reference value.

Rest and stress conditions were also divided in segments of 1s. The root mean square (RMS) was calculated for these segments individually, and normalized by

Table 3.1: Overview of features extracted from the EMG signals

Feature	Abbreviation	Meaning
Root mean square	RMS	Normalized root mean square value as percentage of the reference contraction
Static load	Static	10th percentile of rank ordered RMS values
Median load	Median	50th percentile of rank ordered RMS values
Peak load	Peak	90th percentile of rank ordered RMS values
Gaps/min	Gaps/min	Average number of gaps per minute
Relative time with gaps	Relative gaptime	Percentage of time in which gaps occurred
Mean frequency	MNF	Mean frequency of the magnitude of the frequency spectrum
Median frequency	MDF	Frequency at which the surface on the left side equals that of the right side of the magnitude of the frequency spectrum
Trends within rest and stress conditions	Trends	Slope of a linear fit through RMS values over time within the rest and stress conditions

the RMS reference contraction. RMS values are thus expressed in percentage of the reference contraction. The mean RMS values were calculated for each condition.

To determine the static, median and peak load, the calculated RMS values for individual 1s segments were rank ordered for each rest and stress condition. The 10th (static load), 50th (median load) and 90th (peak load) percentiles of the rank ordered values were determined.

Also, the number of EMG gaps per minute and the percentage of time in which EMG gaps occurred were calculated. Gaps are short periods in time in which muscle activity drops. The rest and stress conditions were divided into segments of 0.2s and RMS values were calculated for these segments. Gaps were detected as one or more segments in a row that had an RMS value below 5% of the RMS reference contraction [142].

The average number of gaps per minute was calculated by dividing the number of gaps by the duration of the condition. The percentage of time in which EMG gaps occurred was calculated by adding the duration of all gaps and dividing by the duration of the condition. This ratio was multiplied by 100 to get the percentages of time.

The mean and the median frequency were extracted from the frequency domain. First, the 1s segments from the recordings were multiplied with a Hanning window to reduce spectral leakage. Then for each segment the fast Fourier transform (FFT)

was calculated. From these FFTs, the mean frequency (MNF), which is the mean frequency of the magnitude of the frequency spectrum, and the median frequency (MDF), which is the frequency at which the surface on the left side equals that of the right side of the magnitude of the frequency spectrum, were calculated for each segment. The mean MNF and MDF values were calculated for each rest and stress condition.

As stress probably builds up during the stress conditions and stress level decreases again during the rest conditions, it could be possible that there are trends in the EMG signals within these periods. To find these trends, the RMS values calculated for consecutive segments of 1s were plotted as a function of time for each condition. After this, a linear fit was made on the data points in a least squares sense. The slope of this line gave an indication about the trend in the RMS values. A positive slope indicates increasing RMS values over time, whereas a negative slope indicates decreasing RMS values.

3.2.7   Statistical Analysis

Statistical analysis of the obtained results was performed according to the process mentioned in the following. We used MATLAB for all statistic procedures. The significance level was set to 0.05 for all tests.

Test for normality The Lilliefors test showed that most of the calculated parameters from the different subjects in the same condition did not have a normal distribution. Therefore nonparametric tests were used for further analysis of the EMG signals.

Test for significant differences among rest and stress conditions Friedman tests were used to see if there were significant differences in EMG features among the five rest conditions and/or among the three stress conditions. If a significant difference was found in a test, a multiple comparison test was performed using Tukey's honestly significant difference criterion to find for which conditions the means are significantly different from each other. If no significant difference was found, an average value was calculated for the five rest conditions and/or the three stress conditions for the next part of the analysis.

Test for significant differences between stress and rest The purpose of this second set of tests was to investigate possible differences in EMG between rest and stress conditions. Depending on the outcome of the previous tests, the values found during the rest conditions and/or during the stress conditions were put in the next Friedman test separately or as one average value. If significant differences were found in a Friedman test, multiple comparison tests were performed using Tukey's honestly significant difference criterion to find which means were significantly different from each other. If both rest and stress values were averaged, a Wilcoxon signed rank test was used instead of a Friedman test.

3.2.8   Correlations of EMG with Questionnaires

We expected to find that stress levels indicated by the subjects were higher during stress conditions than during rest conditions. We also expected to find that for certain EMG features, the values observed during stress conditions were different from those observed during rest conditions. This raises the question if the indicated stress levels given in the questionnaires correlate with the values of the different features in the EMG signal. We tested this with two EMG features: the RMS and the relative gaptime. These features showed most promising results from first observations of the data. The non-parametric Spearman rank correlation coefficients were used for calculation of the correlations.

We also investigated if general stress state obtained by the PSS questionnaire was related to the values of these EMG features that were obtained during the experiment. Therefore we calculated the correlation coefficients between PSS score and RMS values and PSS score and relative gaptime in the EMG signal during all eight different conditions.

It could be that subjects with higher general stress states reacted differently to the stress tasks than subjects with low general stress states. To investigate whether the difference between rest and stress would correlate with PSS score, the mean values of the features during rest conditions were calculated for all subjects and subtracted from the values measured during the three different stress conditions. Spearman rank correlation coefficients were calculated between PSS score and these differences for RMS and relative gaptime.

3.2.9   Individual Results

In the statistical analysis we are looking at differences at group level. However, for some applications it is useful to look at differences at individual subject level. These applications include for example the development of a personalized stress monitor that keeps track of stress levels over time. Although the amount of data that was collected for each subject is limited, we performed a visual inspection on the individual results. Some examples of the results that were found are discussed in Section 3.3.7.

3.3   Results

After having measured 30 subjects, 22 EMG datasets were included in the analysis. The other 8 were excluded due to bad signal quality (7 datasets) or distractions during the experiment (1 dataset).

Some subjects were using the computer mouse with their right hand during the stress conditions, whereas only the keyboard should have been used, like in the

Table 3.2: Average values of the calculated features for 22 included signals for the five different rest conditions R1–R5, the three different stress conditions S1–S3, mean of the rest conditions, mean of the stress conditions, and differences between stress and rest values. All values were calculated from the left trapezius muscle.

	R1	R2	R3	R4	R5	S1	S2	S3	Mean R	Mean S	S−R
RMS (%RVC)	11.3	14.6	11.6	13.7	13.4	15.7	15.1	15.9	12.9	15.5	+2.6
Static (%RVC)	8.3	11.6	8.1	9.5	9.6	8.1	9.5	8.4	9.4	8.7	-0.8
Median (%RVC)	10.0	13.7	10.8	12.0	12.7	15.4	14.8	14.8	11.8	15.0	+3.1
Peak (%RVC)	16.3	18.8	15.4	18.9	17.8	23.9	19.9	22.7	17.4	22.2	+4.7
Gaps/min	8.8	5.0	4.8	7.8	7.7	6.6	4.2	6.4	6.8	5.7	-1.1
Relative gaptime (%)	44.7	32.3	47.5	41.2	42.7	28.5	20.7	33.1	41.7	27.4	-14.3
MNF (Hz)	147.5	141.4	140.6	149.9	136.3	152.5	140.5	147.2	147.7	139.1	-8.6
MDF (Hz)	112.7	107.7	106.1	116.1	102.8	119.8	106.0	112.4	113.7	105.0	-8.8
Trends (%RVC/s) $*10^{-3}$	-4.1	-29.6	-17.7	-73.6	-58.3	36.1	-2.7	-12.1	-36.7	7.1	+43.8

rest conditions. Therefore only signals from the left trapezius muscle were analyzed, because the movements of the right arm during the experiment could have disturbed the EMG measurements of the right trapezius muscle. The values of the different features were calculated for all five rest conditions and the three different stress conditions. The average values of the calculated features for all included subjects can be found in Table 3.2. Also, the mean values for the five rest conditions, mean of the three stress conditions, and the difference between rest and stress conditions are shown.

Results of the statistical tests can be found in Table 3.3. The obtained p-values for the statistical tests performed among the rest and stress conditions are displayed. Based on these results, the column Test shows which test was chosen for the next test, which was used to find significant differences between stress and rest conditions. The last column gives the p-values obtained from these tests.

Table 3.3: Resulting p-values of statistical tests. Diff = difference; W = Wilcoxon signed rank test; F4 = Friedman test with 4 different conditions; R = rest; S1–S3 = stress conditions 1–3: Norinder test, IQ puzzles, memory test. Significant differences ($p < 0.05$) are stated in bold.

| Feature	Diff. rest	Diff. stress	Test	Diff. rest/stress
VAS stress	0.4047	0.2765	W	**$6.08 \cdot 10^{-5}$** **R and S**
RMS	0.1239	0.5796	W	**0.0262** **R and S**
Static	0.3084	**0.0476** **S2 and S3**	F4	**0.0384** **S2 and S3**
Median	0.4159	0.1947	W	**0.0108** **R and S**
Peak	0.1137	0.9556	W	**0.0081** **R and S**
Gaps/min	0.807	0.4706	W	0.4688
Relative gaptime	0.1372	**0.0344** **S1 and S2**	F4	**0.0006** **R and S2**
MNF	0.9825	0.1947	W	**0.0186** **R and S**
MDF	0.7291	0.075	W	**0.0221** **R and S**
Trends	0.1688	**0.0326** **S1 and S3**	F4	**0.0003** **R and S1** **R and S2**
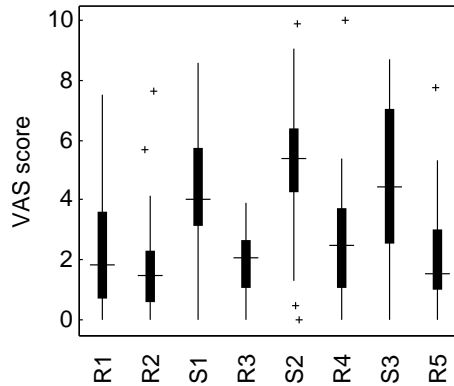
Figure 3.3: Box plots of indicated stress levels during the five different rest conditions R1–R5 and the three different stress conditions S1–S3. The value 0 corresponds to "not stressed at all" and 10 to "extremely stressed". The boxes extend from the lower quartile to the upper quartile values and have a line at the median. Values that lie at more than 1.5 times the interquartile range from the end of the box are considered outliers and are displayed with +. Whiskers extend from each end of the box to the largest and smallest datavalues that are not outliers.

3.3.1 Protocol Validation

To validate the newly designed stress protocol, an analysis was done on the questionnaire answers. In one of the questions of the self-report questionnaires, the subject had to indicate his stress level on a VAS. Box plots of the answers given by all subjects are shown in Figure 3.3. Looking at this figure, it is clear that stress levels after stress conditions on average are higher than after rest conditions. As can be seen in Table 3.3, a statistically significant elevation of stress level after the stress tests was found, in relation to the rest conditions. This proves that the stress levels of the subjects indeed increased due to the newly designed stress tests.

3.3.2 EMG Amplitude

Box plots of the values of the four calculated amplitude features during the eight different rest and stress conditions are displayed in Figure 3.4. In the box plots, relatively large interquartile ranges are visible. However, when looking at the median values for RMS, median and peak load for the different conditions indicated in the box plots, we see that the hypothesis of elevated EMG amplitude during stressful situations seems to be confirmed. The amplitude values of rest 2 are somewhat elevated, compared to the other four rest conditions which do not show differences in median values. This could be explained by the fact that rest 2 started directly after the reference contraction. The muscles may not have been fully relaxed yet after this contraction, resulting in higher EMG amplitudes.
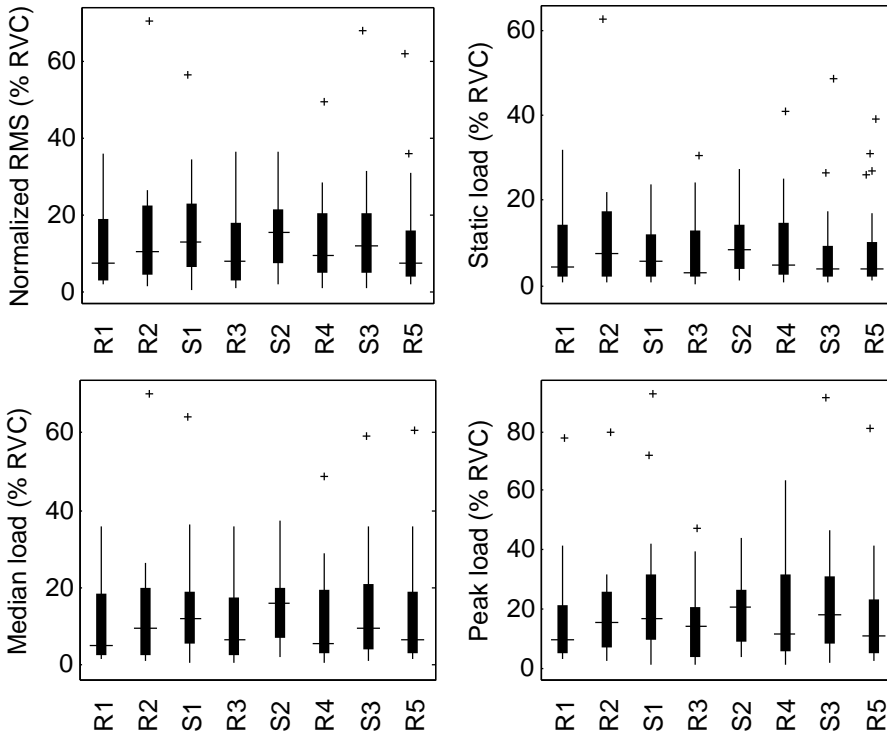
Figure 3.4: Box plots of four amplitude EMG features during the five different rest conditions R1–R5 and the three different stress conditions S1–S3. Upper row from left to right: RMS values and static load. Lower row: median and peak load. For interpretation of the plots, see caption of Figure 3.3.

From the calculated RMS values, it becomes clear that RMS values are higher during stress conditions than during rest conditions. The differences between rest and stress conditions were significant, according to a Wilcoxon test on the RMS values during rest and during stress conditions; see Table 3.3.

Our results show that static load is not a good indicator for stress level, since no significant differences were found between rest conditions and any of the stress tests. The median and peak loads give a better indication of the stress level. Clear differences can be seen between the median values during rest and stress conditions, as confirmed by the Wilcoxon signed rank tests.

3.3.3 EMG Gaps

Box plots of the values of the two calculated gap features during the eight different rest and stress conditions are displayed in Figure 3.5. The first feature that was
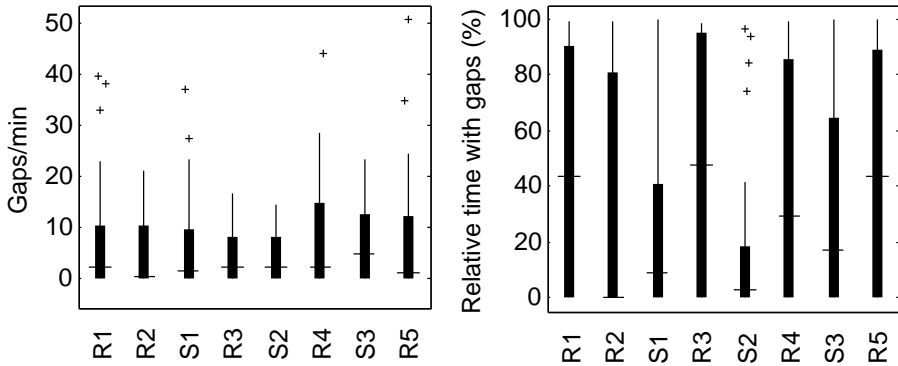
Figure 3.5: Box plots of two gap EMG features during the five different rest conditions R1–R5 and the three different stress conditions S1–S3. Left: number of gaps per minute. Right: relative time with gaps. For interpretation of the plots, see caption of Figure 3.3.

extracted from the detected EMG gaps is the number of gaps that occur per minute. It was expected that fewer gaps would occur during stress than during rest. In the box plots, however, no clear differences can be found between the rest and stress conditions. This is confirmed by the Wilcoxon signed rank test, as can be seen in Table 3.3.

Also for the relative gaptime in the different conditions, it was expected that the values would drop during stress conditions since fewer gaps should occur during stress than during rest. Figure 3.5 seems to confirm this hypothesis, although large interquartile ranges are observed, which indicate large variability among subjects. As was seen before, the second rest condition shows a different behavior from the other rest conditions, possibly due to influence from the reference contraction. The Friedman test with one mean rest value and the three different stress values for relative gaptime as inputs reveals a significant difference between rest conditions and stress test 2, which consisted of the IQ puzzles.

3.3.4 EMG Frequency

Box plots of the two frequency features are displayed in Figure 3.6. Differences between both MNF and MDF values for the different conditions seem to be small (especially when compared to the interquartile range), but when zooming in, a tendency can be found towards lower frequencies during stress conditions, compared to rest conditions. Wilcoxon signed rank tests showed significant differences between rest and stress conditions for MNF and MDF.
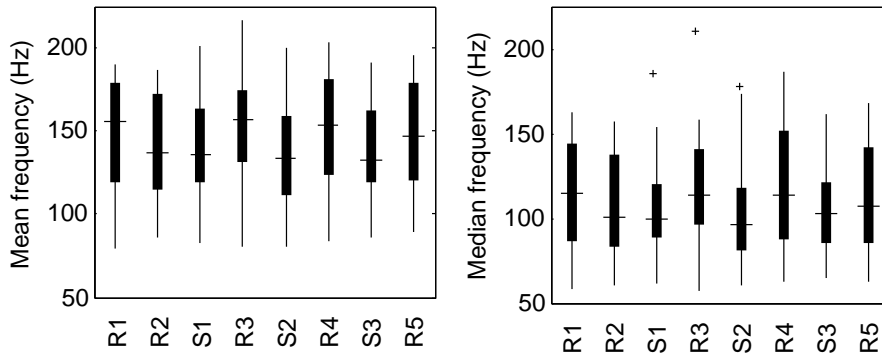
Figure 3.6: Box plots of two EMG features from the frequency domain during the five different rest conditions R1–R5 and the three different stress conditions S1–S3. Left: mean frequency. Right: median frequency. For interpretation of the plots, see caption of Figure 3.3.

3.3.5 Trends within Stress and Rest Conditions

For every condition, the obtained slope values are plotted in a box plot and displayed in Figure 3.7. The median values for stress conditions 1 and 2 are larger than zero. The median value for stress condition 3 is slightly below zero. The median values for the five rest conditions are all below zero. Based on these results, we can conclude that, on average, rest conditions show decreases in EMG amplitude and stress conditions show increases.

There were no significant differences among the slopes of the fitted lines of the five different rest conditions. There was a significant difference between the slopes of stress conditions 1 and 3, which were the Norinder test and the memory task, respectively. The Norinder test showed slopes that were higher than those that were found during the memory task. Therefore, in the next Friedman test, the slopes of the fitted lines of the five rest conditions were taken as one mean value and the three stress conditions were analyzed separately. This test revealed a significant difference between mean slope of the fitted lines during rest conditions and stress condition 1 (Norinder test) and during rest conditions and stress condition 2 (IQ puzzles).

3.3.6 Correlations of EMG with Questionnaires

Plotting all answers given for stress level against their corresponding RMS value results in the picture displayed in the left panel of Figure 3.8. The data shows that there is some positive correlation, which is also shown by the linear fitted line that is plotted between the data points. The Spearman rank correlation coefficient indeed reveals a significant ($p = 1.1003 \cdot 10^{-5}$) positive correlation of 0.3247.
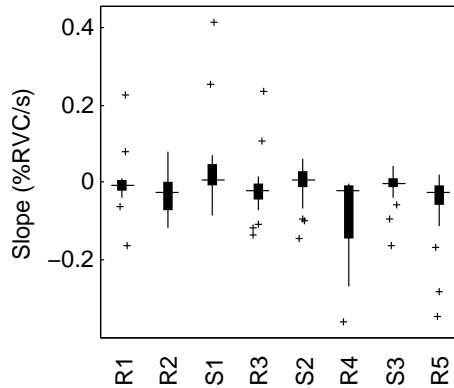
Figure 3.7: Box plots of mean slopes of linear fits on RMS values during the five different rest conditions R1–R5 and the three different stress conditions S1–S3. For interpretation of the plots, see caption of Figure 3.3.

Also, relative gaptime was compared to the answers given for the stress level. The data points and a linear fitted line are displayed in the right panel of Figure 3.8. A significant ($p = 1.4018 \cdot 10^{-5}$) negative Spearman rank correlation coefficient of -0.3210 is found between the stress scores and their corresponding percentages of time with EMG gaps. The correlation is similar in strength to the correlation between RMS values and the answers given for "stressed".

None of the calculated correlation coefficients between PSS score and RMS values and PSS score and relative gaptime in the EMG signal reached significance. General stress state obtained by the PSS questionnaire was thus not related to the values of these EMG features that were obtained during the experiment. Spearman rank correlation coefficients were calculated between PSS score and the differences between rest and stress conditions for RMS and relative gaptime. No significant correlations were found here either.

3.3.7 Individual Results

The individual results of all 22 included subjects were visually inspected to check for differences between rest and stress conditions. When looking at the results in the different conditions for all subjects separately, a lot of variation is visible. Some subjects show the expected pattern of elevated EMG amplitude and reduced amount of EMG gaps during stress, compared to rest conditions; other subjects show an opposite behavior, hardly a difference at all or exceptional high or low values for some of the conditions. Only four subjects showed EMG responses as expected for all conditions. Another eight subjects showed responses that were partly as expected. These subjects only responded as expected to some conditions or not for all features. The remaining ten subjects did not respond as was expected.
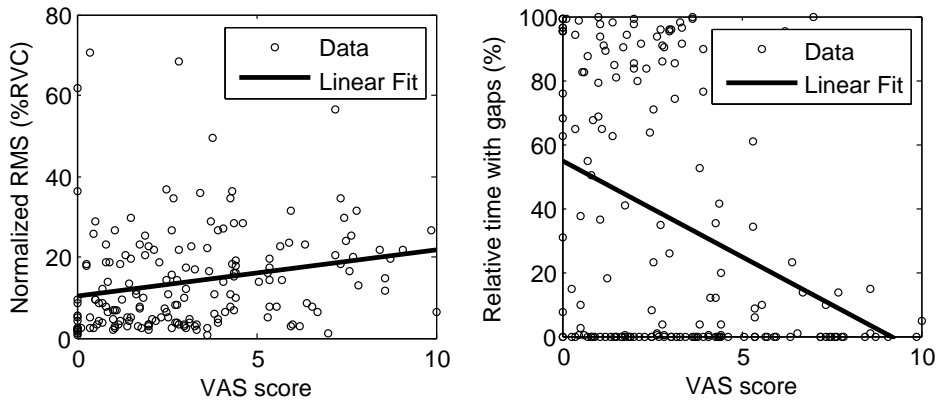
Figure 3.8: Left: Indicated VAS scores for stress levels during all five rest conditions and three stress conditions versus the RMS values during the corresponding conditions. Right: Indicated VAS scores for stress levels and relative gaptime for the corresponding conditions.

Figure 3.9 shows examples of EMG signals of two subjects, of which the upper signal shows the expected reaction and the lower signal shows no clear stress reaction. The indicated stress levels and the calculated peak load, relative gaptime and mean frequency during the different rest and stress conditions for these subjects are shown in Figures 3.10 and 3.11, respectively. In the upper panel of Figure 3.9 we see that during the stress conditions, the amplitude of the EMG signal is higher than during the rest conditions. In Figure 3.10, we see elevated stress levels indicated by the subject and we see these differences in stress level reflected in the EMG features derived from the EMG signal. Peak load is higher during stress than during rest. Relative gaptime and mean frequency are lower during stress conditions, compared to rest conditions. In the lower panel of Figure 3.9, no clear pattern of differences between rest and stress conditions is observable. In Figure 3.11, we see elevations of stress level during stress test 1 and 2 indicated by the subject himself. However, in the features derived from the EMG signal, no clear differences can be found between rest and stress conditions.

3.4 Discussion

3.4.1 Contribution of This Work

The first goal of this study was to propose and validate a new protocol for stress induction in situations mimicking daily work. The second goal was to confirm that stress can be detected from amplitude and temporal features of the EMG of the upper trapezius muscle. Third, two features of the frequency spectrum were analyzed to investigate if stress influences frequency components of the EMG.
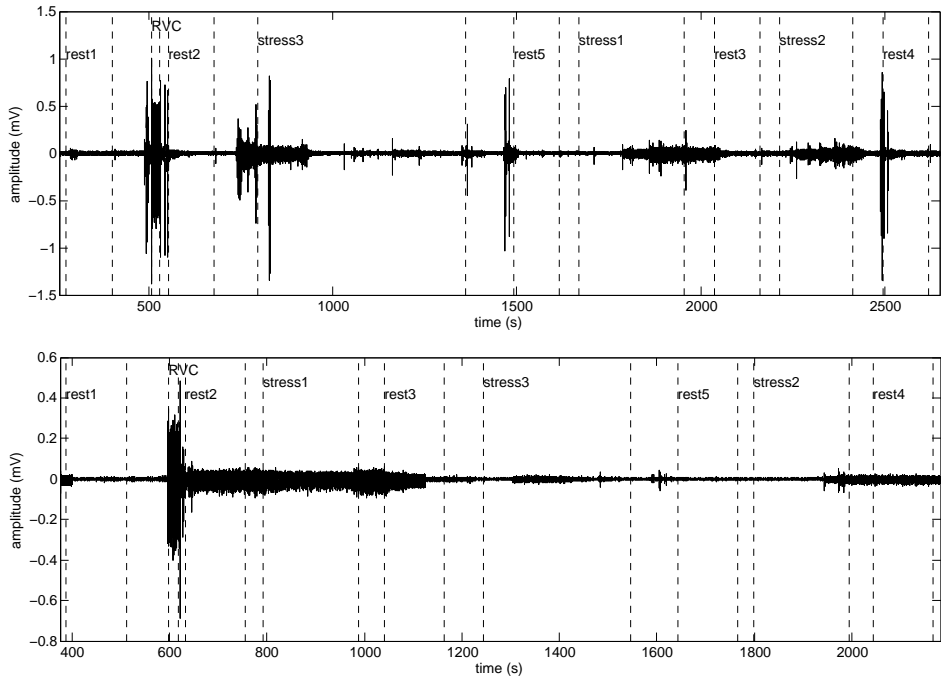
Figure 3.9: Upper panel: EMG signal of one of the subjects who did respond as expected. Lower panel: EMG signal of one of the subjects who did not respond as expected. The signals were filtered as described in Section 3.2.5. The reference contraction (RVC) and the different rest and stress conditions are indicated with vertical dashed lines. Every condition starts at the dashed line with a label mentioning the condition and ends at the next dashed line.
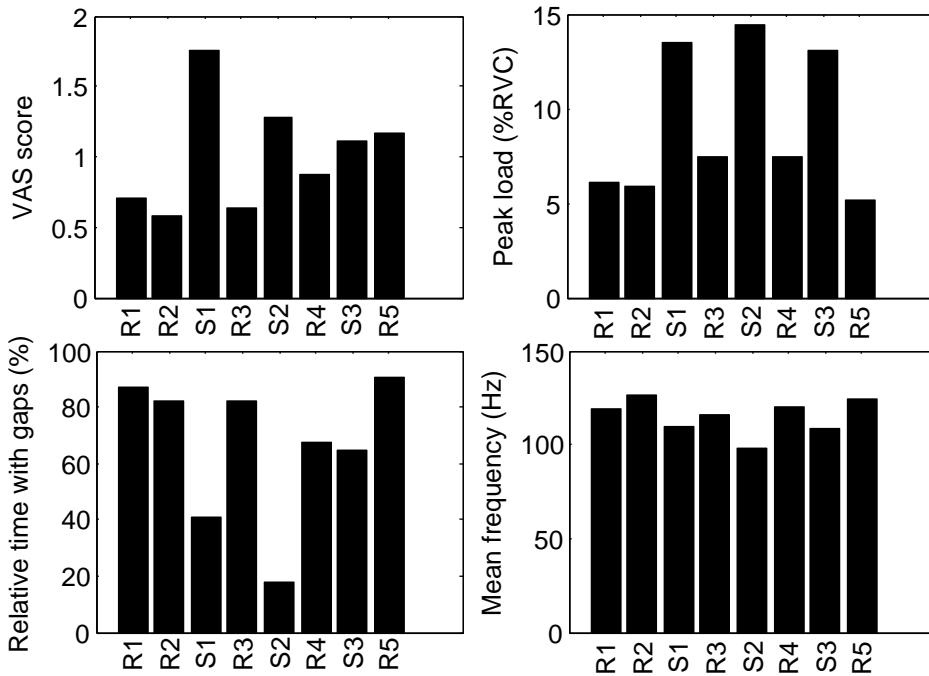
Figure 3.10: Values of indicated stress levels and three EMG features during the five different rest conditions R1–R5 and the three different stress conditions S1–S3 for one of the subjects who did respond as expected. Upper row from left to right: VAS score for stress level and peak load. Lower row: relative gaptime and mean frequency.

Figure 3.11: Values of indicated stress levels and three EMG features during the five different rest conditions R1–R5 and the three different stress conditions S1–S3 for one of the subjects who did not respond as expected. Upper row from left to right: VAS score for stress level and peak load. Lower row: relative gaptime and mean frequency.

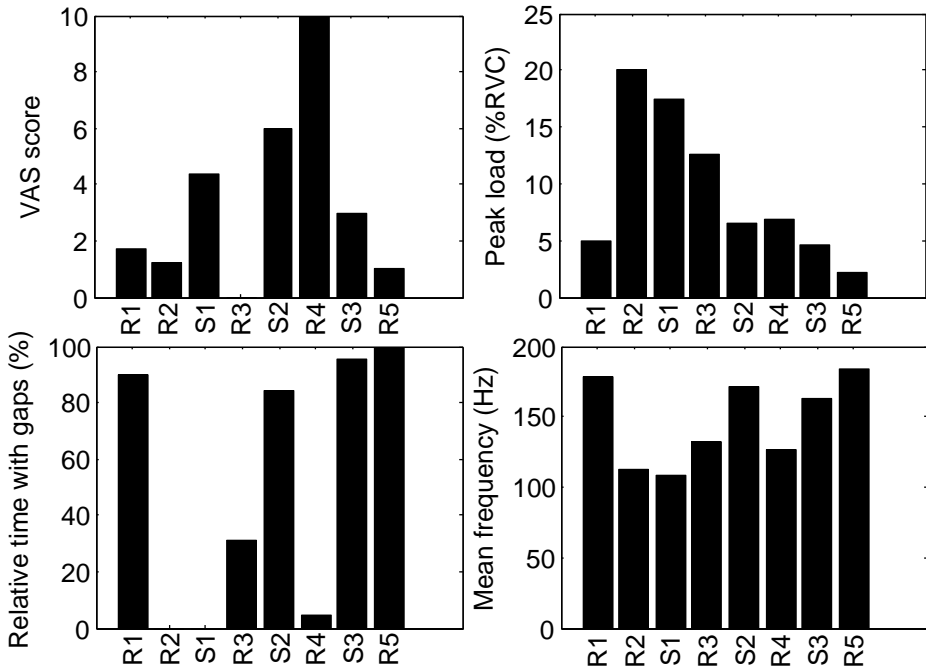Fourth, it was investigated if there are trends over time in the EMG signals. Finally, the data were checked for correlations between perceived stress level and EMG features.

Our first goal was to propose and validate a realistic new protocol for stress induction. According to the results of the analysis on the indicated stress levels of subjects, we succeeded in reaching this goal. Subjects indicated to have significantly higher stress levels directly after the stress conditions than after the rest conditions. This new protocol can be used in future studies to induce stress that is similar to daily work stress.

For all of the investigated EMG features, values found during the five different rest conditions were similar, as was expected. Some differences were found among the three different stress conditions: for static load between stress tests 2 and 3 and for relative gaptime between stress tests 1 and 2. These differences indicate that reactions to the different stress tests were not the same for all three tests. The actual differences in physiological reaction between those tests are unclear, so no statements can be made at this stage about the reason behind these differences. A significant difference was found between mean rest values and mean stress values for RMS, median load and peak load, which shows a relation between stress level and amplitude of the EMG signal. At this point, it is difficult to draw conclusions about the differences between these amplitude features and to tell which feature would be the best predictor for stress level. Further research is needed to investigate the different physiological aspects of the EMG that the different features are showing and to investigate which one would be the most suitable for measuring stress level. The number of EMG gaps per minute is not a good indicator for stress, but in the relative gaptime a highly significant difference between rest conditions and stress test 2 was found. Probably a combination of one of the amplitude features mentioned before and the relative time with gaps could be used as a predictor of stress level.

We found indications that the frequency spectrum of the EMG signal shifts towards lower frequencies during stress. This phenomenon is similar to what happens to the spectrum when a muscle gets fatigued. This is caused by a physiological response to fatigue in which the muscle fiber conduction velocity decreases [116]. We believe the results we found are not caused by fatigue. The frequency values decreased during stress, but not during the rest periods that were also part of the protocol. If the changes would have been caused by fatigue due to the physical effort of sitting behind a desk, the effect should have stayed during the rest conditions. Stress could induce physiological responses in a muscle that are different from fatigue, which could result in a similar shift of the spectrum to the lower frequencies.

Our fourth goal was met, because we found that, on average, the amplitude of the EMG signal decreases during rest conditions and increases during stress conditions. This leads to the conclusion that the reaction to stress and the relaxation afterwards do not occur instantaneously, but are changing gradually by time. The negative slopes during rest conditions are more pronounced than the slopes during stress conditions. This could be an indication that the relaxation process during

rest conditions occurs more gradually than the process in which tension is built up during stress conditions.

Finally, correlations were found between RMS values of the EMG signal and the stress levels indicated by the subjects themselves and between the relative gaptime and the indicated stress levels. This is as was expected and indicates that when subjects feel more stressed, this is reflected in the EMG features. No significant correlations were found between PSS score and different EMG features. A reason for this could be the variation in EMG signals among different subjects. Due to this variation, no absolute stress level can be extracted from the EMG signals. However, also for the differences in EMG feature values between rest and stress levels per subject, no correlations were found between PSS score and these differences. So subjects with a high PSS score do not show higher or lower increases in muscle activity during stress conditions than subjects with a low PSS score.

3.4.2 Comparison with Previous Work

Looking at other reports on the effect of mental stress on muscles, the results that were found in our experiment are quite similar to what was found before. In our experiment, different ways to induce stress were used compared to the other experiments. With our tests, the first step towards more realistic and daily life stress situations was made. Doing the Stroop test or calculations on one hand, as was done in tests used in previous experiments, and solving small problems or trying to remember things for colleagues on the other hand, as was done in our experiment, were proven to show similar stress reactions.

Krantz et al. [85] reported EMG levels as percentage of a reference contraction of the shoulder muscle of the arm that was not used during the experiment, comparable to what was done in our experiment. The difference that Krantz et al. found between rest and mental stress conditions is about 5% of the RVC. The mean difference that was found in our experiment was only around 2.5% of the RVC. It may be that Krantz et al. managed to induce more stress than was induced in our experiment. However, the difference could also be caused by other factors or just be arisen arbitrarily.

Both Laursen et al. [93] and Schleifer et al. [142] performed analysis on EMG gaps during stress. Huge differences are found in the values for the number of gaps that were detected per minute. These differences could be caused by the fact that Laursen et al. used another definition of an EMG gap than the definition used in our experiment, which is the same as was used by Schleifer et al. This makes it hard to compare gap rates found in different experiments. Related to this is the difference in relative time with EMG gaps that was found by Laursen et al. and the values that were found in our experiment. Laursen et al. found a relative gaptime of 68% in rest and 33% during stress; in our experiment mean values of 42% in rest and 27% during stress were found. The values cannot be compared to each other directly, but still it seems that Laursen et al. found a bigger reaction to stress in the relative gaptime than what was found in our experiment. The relative gaptime

that Laursen et al. found during stress is less than half of the value found in rest. In our experiment the value drops "only" 36% from rest to stress.

Larsson et al. [92] did not find differences in mean frequency of the EMG, whereas in our experiment the mean frequency during stress was significantly lower than during rest.

3.4.3 Limitations of This Work

There are some limitations of our work. First of all, the subject was in a static position during the protocol. No physical activity was involved in the protocol. This makes the conditions difficult to compare to a daily life environment. We did not include physical activity in our protocol, because it would have interfered with the stress-related EMG reactivity in the muscles. This interference would have made our analysis too complicated at this point. Extending the environment to a more daily life environment is a point for future work.

Our subject group was relatively small. The available data for analysis was only from 22 subjects. A bigger subject group would be needed to confirm our findings. However, most of our results are in line with results found in literature. This indicates that our results could be generalizable to a wider population.

The newly designed stress protocol was validated by questionnaire answers only. Other authors have used hormone levels to validate stress protocols. For example, for the Trier Social Stress Test (TSST), adrenocorticotropin, cortisol, growth hormone and prolactine were used for validation [79]. The use of hormone levels could have strengthened the conclusions on the validation of the protocol. Validation of the TSST also included analysis of heart rate changes. We do not want to rely on heart rate for validation of our protocol. However, as we recorded heart rate during the experiments, we analyzed the heart rate changes during our stress tests. A significant difference in heart rate from rest to stress conditions was found $(p = 1.36 \cdot 10^{-4})$ [Unpublished work]. This finding adds some confidence that the protocol indeed induced stress in subjects.

3.4.4 Future Work

For all eight features that were derived from the EMG, relatively large variability is observed in the values for these features among the different subjects. Despite the normalization of EMG amplitude to the reference contraction, it is visible in the box plots in Figures 3.4 and 3.5 that interquartile range is relatively large. Also, in Figure 3.6, the ranges of observed mean and median frequencies are relatively large. The data points collected during rest conditions show large overlap with those collected during stress. This suggests, in general, that amplitude, temporal and frequency features of the trapezius EMG, when normalized by the reference contraction, are not suitable for statistical classification between rest and stress

conditions. Whereas the statistical tests that were performed clearly show differences between rest and stress EMG on group level, statistical classification focuses on recognizing different conditions on an individual basis. Due to the variability among the subjects, calculation of an individual real time, absolute stress level is not possible yet either. Future work should focus on finding a better normalization method. If good normalization is not possible to achieve, another way to deal with subject variability in EMG features has to be found. For example a calibration step for each subject could be included in the procedure.

We know that EMG is influenced by many other factors, apart from stress. Static factors like body morphology and dynamic factors like posture and movement influence the recorded EMG signals possibly even more than stress. Therefore it is very difficult to take the measurements out of the laboratory into uncontrolled environments. Physical activity will have huge influences on the EMG signals. A simple normalization step like we did does not seem to compensate for all these factors. Future work must include further investigation into these factors and how to compensate for variations in EMG that are not related to stress. Probably it is needed to retrieve context information on these factors from the environment and from the body.

Furthermore, it might be useful to look at more individual trends next to analysis on group level. To do this, more data is needed per subject to be able to compare a subject's data to his own baseline data. By performing analysis on individual level, compensation for subject specific factors that influence EMG would not be needed anymore.

Finally, whereas we looked at individual EMG features in our analysis, a better stress level estimation could be made by combining multiple features in one stress measure. Even other physiological features, like heart rate, respiration rate and skin conductance level, could be included in such a stress level estimation.

A good stress level estimation facilitates stress detection at an early stage. Feedback can be given to the user about his increasing stress level. If the user reacts to the feedback and changes his behavior accordingly, musculoskeletal and other stress-related disorders can be prevented.

3.5 Conclusions

First, we presented and validated a new protocol for stress induction in situations mimicking daily work. Secondly, it was confirmed that stress changes amplitude and temporal features of the EMG of the upper trapezius muscle. Thirdly, two frequency features were analyzed and it was shown that stress influences frequency components of the EMG. Fourth, trends during stress and rest conditions were found for EMG amplitude. Finally, correlations between subjectively indicated stress levels and features derived from the EMG were found.

The new protocol showed statistically significant increases in stress level due to

newly designed stress tasks. A statistically significant relation was found between stress level and a number of features extracted from EMG signals of the upper trapezius muscle. Amplitude of the EMG signal is significantly higher during stress conditions than during rest conditions. For the relative gaptime a highly significant difference between rest conditions and one of the stress tests was found. In the frequency domain, significant increases of the lower frequency contents in the EMG were observed during stress. The amplitude of the EMG signal increases during stressful conditions and decreases again during relaxation periods. The stress levels indicated by the subjects correlate with the amplitude of the EMG signal (RMS) and correlate negatively with the gaps occurring in the EMG signal.

To be able to classify EMG signals into rest or stress conditions, or to calculate real time stress levels, it is necessary to find a good way of normalization of EMG signal amplitude. A better normalization method might reduce the variability of the features and thus improve statistical classification performance. Also, analysis on individual subject level rather than on group level is likely to improve results.

When this is achieved, a combination of several amplitude, gap and frequency features promises to form a reliable basis for calculation of stress levels. In future work, EMG may be combined with other physiological signals to provide a more robust reliable measurement of stress. A reliable stress level measurement is useful in preventing stress-related disorders like musculoskeletal problems.

4

Wearable Physiological Sensors Reflect Mental Stress State in Office-Like Situations

Timely mental stress detection can help prevent stress-related health problems. The aim of this study was to identify those physiological signals and features suitable for detecting mental stress in office-like situations. Electrocardiogram (ECG), respiration, skin conductance and surface electromyogram (sEMG) of the upper trapezius muscle were measured with a wearable system during three distinctive stress tests. The protocol contained stress tests that were newly designed to represent office-like situations. Two analysis methods were compared: Principal Component Analysis (PCA) and Generalized Estimating Equations (GEE). With PCA, we reached a classification rate of 79.3% between rest and stress conditions of our protocol using a Fisher's Least Square Classifier. Using GEE, on average 74.5% of the cases in the dataset were classified correctly. It should be investigated if these approaches could be used for continuous stress measurement in daily office life to help detect various stress situations at an early stage.

4.1   Introduction

Stress is a major and increasing problem in today's society. Stress, depression and anxiety together are the second most frequently occurring work-related health problems in the European population. Of sickness absences of one month or more, 25% is caused by stress, depression or anxiety [48]. In the US, 30–40% of employees report their job as stressful and the number of lost work days due to anxiety, stress, and neurotic disorders is four times higher than other nonfatal injuries and illnesses [122].

Stress activates the Sympathetic Nervous System (SNS) [162]. Activation of the SNS stimulates sweating, accelerates the heart rate and increases muscle activation, for example [57, 107]. Chronic mental stress results in chronic SNS activation and this can cause health problems such as hypertension, cardiovascular diseases, increased likelihood of infections and depression [153, 147, 162].

Stress-related health problems could be prevented by detecting stress at an early stage, but this requires that stress can be assessed in an objective and quantitative way. One approach to detect mental stress is by measuring relevant physiological parameters, such as muscle tone, heart rate, heart rate variability, skin conductance and eye pupil diameter; see for example [65, 105, 130]. Combinations of several features from these physiological parameters differentiate between stressful and non-stressful situations; see for example [29, 58, 62, 63, 149, 174, 175].

Recent developments in wearable and wireless health monitoring systems facilitate long-term measurement of physiological signals. Body sensor networks can be used to continuously monitor a number of physiological parameters. Mental stress levels could be continuously calculated and updated using these parameters. When real-time stress levels can be estimated robustly, feedback can be provided to the users. The users can change their behavior according to the feedback. In this way a stress monitoring system can prevent health problems caused by chronic stress or it can make chronic disease management possible.

In current stress research, mental stress is often artificially induced in a controlled setting. The Stroop color word test or mental arithmetic tasks are the most popular ways to induce mental stress in these studies. These tests can be very stressful, but do not simulate situations that are encountered in daily life. Besides, usually one type of stressor is used in a study; in daily life typically various stressors are encountered. Therefore we developed and used a new protocol, which was designed to include various realistic stress situations. The protocol has been previously reported and validated using questionnaires (see Chapter 3).

In Chapter 3, we reported on the reaction of the trapezius muscle surface electromyography (sEMG) signal, from now on referred to as electromyography (EMG), on our protocol. In our study, we collected four physiological signals: electrocardiogram (ECG), skin conductance (SC), respiration and EMG of the trapezius muscles. Although, as mentioned before, several studies have investigated the advantages of combining various physiological signals for stress detection, this exact

set of signals is not commonly used. Healey and Picard [62, 63], however, used the same set of signals as we did. They collected physiological data from subjects driving a car and showed a nice example of a study involving a daily life stressor.

The goal of the current study was to identify those physiological signals and features that show the most distinct reaction to the various mental stress situations in our protocol.

Many studies apply statistics to their data to investigate any significant differences between the various conditions of their protocols. However, we use classification to actually classify the data points into one of the conditions. This approach is more relevant for stress detection on an individual basis in the future. It is relevant to know if an *individual* using a stress detection system is stressed, rather than knowing if there are differences between a group of stressed persons and a group of non-stressed persons.

The main contributions of this paper are as follows: (i) we propose a set of nine physiological features, from four different physiological signals, contributing to the stress response; (ii) we demonstrate the possibility to classify mental stress from rest conditions using these features; and (iii) we show evidences that the four physiological signals all contribute to the identification of stress in a robust manner.

4.2   Methods

4.2.1   Experimental Protocol

The experimental protocol is only discussed briefly here. For a detailed description of the protocol, and a validation that stress was actually induced, based on questionnaire answers, the reader is referred to Chapter 3.

A total number of 30 subjects were recruited for the experiment. Subjects had to be aged 18 or older and English-speaking. Subjects with muscle disease, cardiac problems or mental disorders such as depression or anxiety were excluded. The ages of the subjects ranged from 19 to 53 (mean = 33.1; SD = 7.9); 25 subjects were male and 5 were female. All subjects were right-handed. Although there are some indications that women and men react differently to stress (e.g. [85]), the subject group was too small to investigate the differences between men and women. Therefore the subject group was analyzed as a whole.

The test was performed via a PC in a quiet room with as few distractions as possible (Figure 4.1). The subject had a mouse and keyboard available for completing the questions and tests involved in the protocol. The sounds involved in the protocol were played through headphones. A video camera was set up in the room to suggest that the subject would be videotaped during the protocol.

We developed an experimental protocol to study physiological responses to stress
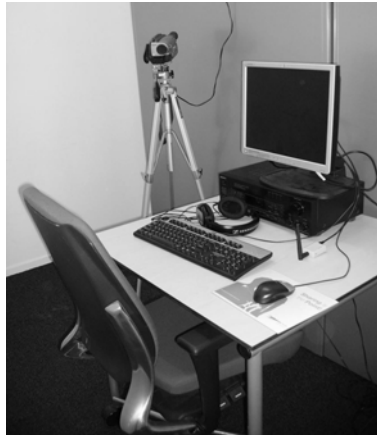
Figure 4.1: Experiment room.

situations similar to stressors encountered in daily office life. The experimental protocol is illustrated in Figure 4.2.

In the initialization phase of the protocol, the subject answered some general questions and completed the perceived stress scale (PSS) questionnaire [31]. Then a rest period was scheduled and the subject completed the first self-report questionnaire. In this questionnaire, the subject indicated his/her emotional state on visual analogue scales (VAS) by answering 10 questions about 10 different discrete emotions. The scales ranged from "not at all" to "extremely". One of the emotions was "stressed". Then the subject performed a reference contraction of the shoulder muscles, followed by another rest period and questionnaire.

In the stress phase, every subject was exposed to three different stress conditions: a calculation task (the Norinder test [103], 2:30 min), a logical puzzle task (3:00 min) and a memory task (approximately 5:00 min). All three tasks were done under time pressure and with distracting news fragments that were played through headphones. Social pressure was induced in the memory task by telling the subjects that their performance would be included in a group result and published to colleagues afterwards. To invoke extra stress, the subject was recorded on video during the memory task as well. It was told that this was done for analysis after the experiment. The order of the three tests was randomized among the subjects to minimize crossover effects. Two-minute resting periods were scheduled in between the stress conditions to let the subject return to baseline physiological and psychological states. The subject heard classical music during these periods. There was some physical activity involved in the rest periods that consisted of typing numbers on the keyboard. This activity was similar to the expected activity during the stress conditions. Self-report questionnaires had to be completed both before and after each stress and rest condition. The results of these self-reports were used to validate that we actually induced stress with our protocol; for details see Chapter 3.
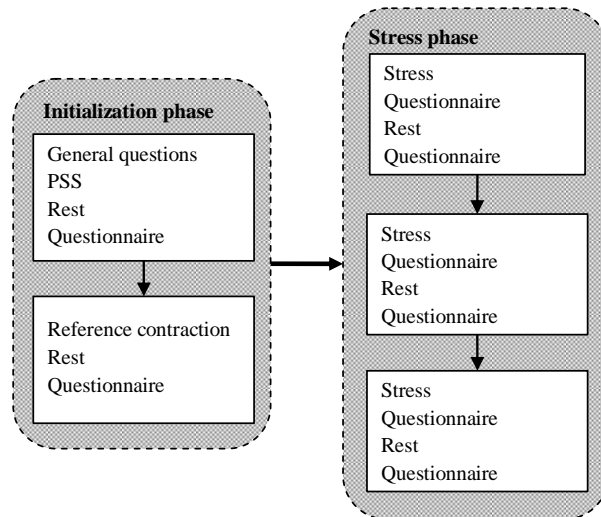
Figure 4.2: Schematic overview of the procedure of the stress experiment. The procedure starts with an initialization phase. The stress phase consists of three different stress conditions, with rest periods and self-report questionnaires scheduled in between. PSS: Perceived Stress Scale.

The entire experimental protocol was implemented in a MATLAB interface. Once the interface was started, the whole experiment proceeded automatically for about 40 minutes, until the subject finished the last questionnaire. Automated event marking allowed accurate mapping of the various conditions on the recorded signals.

4.2.2 Physiological Recordings

The physiological signals were recorded with a wireless body sensor network based on the body area network platform developed within the Human++ program in Holst Centre/imec [20]. The body sensor network facilitated synchronization of all sensors and accurate mapping of the different phases of the test protocol on the signals. Signals were recorded on a micro SD card. One-lead ECG and respiration were recorded at a sampling frequency of 250 Hz using a wireless chest belt. A wireless hand sensor recorded SC with a sampling frequency of 100 Hz. Bipolar EMG signals were measured from the upper trapezius muscles of both shoulders. Two electrodes were placed on the trapezius muscles in the position as described by SENIAM [148]. A reference electrode was placed on the spinous process of vertebra C7. The electrodes were connected to wireless EMG sensor nodes recording at 1000 Hz. More details about the procedure of EMG recording can be found in Chapter 3.

4.2.3   Signal Processing and Feature Extraction

An overview and description of the features that were calculated from the measured signals is shown in Table 4.1. The features were chosen based on the results discussed in Chapter 3 and literature on stress detection from physiological signals (see for example [84, 58, 63, 149, 174]).

For detection of skin conductance response rate (SCRR), the SC signals were smoothed using a 3s moving average filter. SC responses were defined as peaks that followed a derivative value of at least 0.0093 μS/s. This value was empirically found to best detect the responses. Before calculating the signal power in the second difference of the SC signal (SCdiff2), the SC signals were smoothed using a 5s moving average filter. Ohmic Perturbation Duration (OPD) of the SC is defined as the relative time of responsiveness. It varies between 0 and 1 and indicates the relative time that the second difference of the SC signal exceeded a certain threshold. See also [58]. Respiration frequency (RespFreq) was determined as the main frequency component of the power spectral density of the respiration signal below 1 Hz. EMG features were calculated from the signals of the left shoulder. The user gave all input during the rest and stress conditions using the numeric keypad with the right hand. The left arm did not move at all during the entire protocol and therefore no movement interference was expected in the EMG signals of the left side. More details of the calculation of the EMG related features can be found in Chapter 3.

Because a large variability in physiology between the subjects is expected, the feature values were normalized first. The following procedure was followed. All features were calculated over a sliding window of 120s that moved over the signals. The length of the window was equal to the length of the shortest condition: the two-minute rest condition. For most of the features, the window moved in steps of 1s. We believed that a smaller step size would not have led to more accurate results, because we assume that stress levels are not changing within fractions of seconds. For heart rate (HR) and standard deviation of interbeat intervals (SDNN) the window moved from beat to beat. For every position of the window, one value was calculated for each feature for that particular time window. The obtained feature values were normalized by calculating the z-score as follows:

$$z = \frac{(x - \bar{x})}{s} \tag{4.1}$$

With x being the original feature value, z the normalized value, and $\bar{x}$ and s the mean and standard deviation of the feature values over the entire recording, respectively.

After this normalization step, we chose to analyze one average value per rest or stress condition and not to investigate the variability of the feature values *during* the various conditions. For rest conditions, we used the normalized value for the 120s time window exactly overlapping the rest condition. The stress conditions

Table 4.1: Overview of features extracted from the measured signals

	Feature	Extracted from	Abbreviation	Meaning
1	Heart rate	ECG	HR	Mean heart rate
2	Standard deviation interbeat intervals	ECG	SDNN	Mean standard deviation of the interbeat intervals
3	Low frequency heart rate variability	ECG	LF	Heart rate variability in the 0.04–0.15 Hz band
4	High frequency heart rate variability	ECG	HF	Heart rate variability in the 0.15–0.4 Hz band
5	LF/HF ratio heart rate variability	ECG	LFHF	Ratio of the low and high frequency of heart rate variability
6	Skin conductance level	SC	SCL	Mean level of skin conductance
7	Skin conductance response rate	SC	SCRR	Mean number of skin conductance responses per second
8	Skin conductance second difference power	SC	SCdiff2	Signal power in the second difference of the skin conductance signal
9	Ohmic perturbation duration skin conductance	SC	OPD	Relative time of responsiveness of the skin conductance signal
10	Respiration frequency	Resp	RespFreq	Mean respiration frequency
11	RMS of the respiration signal	Resp	RMSResp	Root mean square value of the respiration signal for estimating tidal volume changes
12	RMS of the EMG signal	EMG	RMSEMG	Normalized root mean square value as percentage of the EMG reference contraction
13	Static load	EMG	Static	10th percentile of rank ordered EMG RMS values
14	Median load	EMG	Median	50th percentile of rank ordered EMG RMS values
15	Peak load	EMG	Peak	90th percentile of rank ordered EMG RMS values
16	Gaps/min	EMG	Gaprate	Average number of EMG gaps per minute
17	Relative time with gaps	EMG	Gaptime	Percentage of time in which EMG gaps occurred
18	Mean EMG frequency	EMG	MNF	Mean frequency of the magnitude of the EMG frequency spectrum
19	Median EMG frequency	EMG	MDF	Frequency at which the surface on the left side equals that of the right side of the magnitude of the EMG frequency spectrum

ECG: electrocardiogram, SC: skin conductance, Resp: respiration, EMG: electromyography

all lasted longer than 120s. We assumed that stress would build up during the stress conditions and therefore used the normalized feature values calculated over the last 120s of these conditions for our analysis.

4.2.4   Analysis Methods

Reduction of Feature Set

A total of 19 features were extracted from four physiological signals. It was expected that some of these features would show high correlations, since they are based on the same physiological processes. This introduces redundancy in the dataset and multicollinearity among the features. A selection was made based on the correlations and prior knowledge from a literature study to reduce the number of features and construct a non-redundant feature set for further analysis. Selecting this subset of features also reduced the number of features that needed to be calculated from the data. In a future application of a stress monitoring system this will save processing time and power.

Principal Component Analysis

Next, PCA was applied to the feature subset. The number of principal components (PCs) was chosen such that at least 90% of the variance was explained. These PCs were used to classify the three stress conditions and the three following rest conditions. A 5-fold cross-validation was performed five times to evaluate the classification performance. The average and standard deviation of the error rate over all classifications was calculated. Four different classifiers were used to investigate the differences in performance: a Linear Bayes Normal Classifier, a Quadratic Bayes Normal Classifier, a K-Nearest Neighbor Classifier, and a Fisher's Least Square Linear Classifier.

Generalized Estimating Equations

Apart from PCA, GEE was applied to the feature subset as a second analysis approach. We combined the results of the three different stress conditions and therefore our dependent variable is binomial; it has only two states: "rest" and "stress". Uncorrelated binomial data could have been analyzed with logistic regression, which is a generalized linear model (GLM). However, our dataset contains repeated measures and therefore data points of the same subject are likely to be correlated. We used GEE, because it represents an extension of the GLM to accommodate correlated data. A GEE model with binomial distribution and logit link function was chosen. The order of the three stress tests was randomized, so an exchangeable working correlation matrix was chosen. The analysis was performed using software package "IBM SPSS Statistics 20".

Manual backward feature selection was used to select the best features for differentiation between the rest and stress conditions. The selection was based on significance of the Wald Chi-Square tests of the coefficients. In every step, the parameter with the largest significance value was removed from the model. While selecting features, the goodness-of-fit statistic (Corrected Quasi Likelihood under Independence Model Criterion) was monitored. This statistic should become smaller after every step.

As a training set, 80% of the stress cases and 80% of the rest cases were selected randomly. The other 20% of the stress and rest cases were used as the test set. This splitting procedure was done five times to check if the results obtained with different training and test sets would be consistent.

Using the coefficients obtained by the GEE analysis, the odds of certain normalized feature values to be measured during a stress condition were calculated, using:

$$odds = e^{\sum_{k=1}^{n} B_k \cdot x_k} \tag{4.2}$$

n is the number of features, B_k is the coefficient for feature k and x_k is the normalized feature value of feature k. The probability p was then calculated from the odds as follows:

$$p = \frac{odds}{1 + odds} \tag{4.3}$$

So the outcome of the GEE procedure was a probability between 0 and 1 for each case of being a stress condition. With a cutoff value of 0.5, cases were classified into rest or stress conditions.

4.3   Results

4.3.1   Population

A number of subjects were excluded from analysis because of poor signal quality (1 poor respiration signal, 5 poor SC signals, 5 poor EMG signals; 9 subjects excluded), incomplete data due to failing sensor nodes (2 subjects), or distractions due to other people being present in the room during the experiment (1 subject). Exclusion of these subjects left 18 full recordings of good quality.

4.3.2   Feature Selection

The absolute correlation values among the 19 normalized features are shown in Figure 4.3. High correlations were found for some features. For example, SDNN, low frequency (LF), and high frequency (HF) heart rate variability correlate strongly
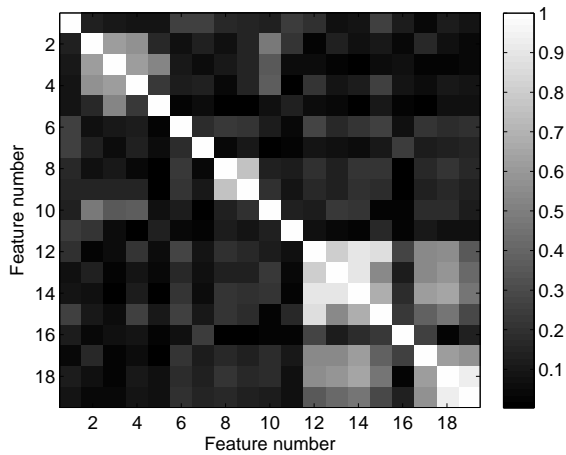
Figure 4.3: Absolute correlation values among the features. The feature numbers correspond to the numbers in Table 4.1.

as do RMS of the EMG signal (RMSEMG), static load, median load, and peak load of the EMG signal. Mean EMG frequency (MNF) and median EMG frequency (MDF) are also highly correlated.

Based on these correlations and on prior knowledge, a subset of 9 features was chosen that did not include absolute correlation values higher than 0.6 between features:

HR Does not correlate strongly to any other feature and has shown to react to stress in many other studies (see for example [85]).

SDNN Was chosen from the cluster SDNN, LF, HF, LFHF. SDNN gave promising results in earlier research [58].

SCL Does not correlate strongly with other features, so this feature is included by itself.

SCRR Does not correlate strongly with other features, so this feature is included by itself.

SCdiff2 Correlated strongly with OPD; SCdiff2 was chosen arbitrarily from the two.

RespFreq Does not correlate strongly with other features, and might be indicative of stress.

Peak Was chosen from RMS, Static, Median, Peak; Peak load gave the best result from this subset in our earlier analysis on EMG signals (Chapter 3).

Gaptime Showed promising results in our earlier analysis (Chapter 3).

MNF Gave a slightly better result than MDF in our earlier analysis (Chapter 3).
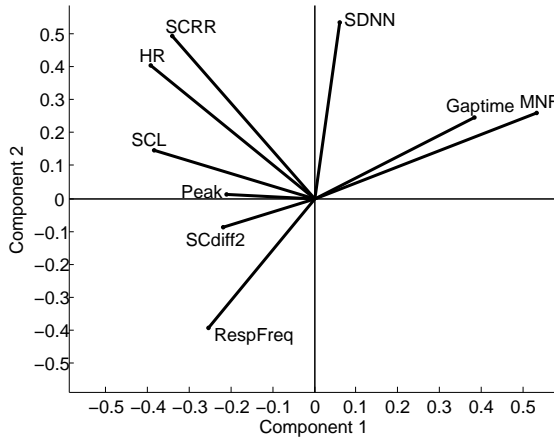
Figure 4.4: Contribution of the nine features to the first and the second principal component.

Table 4.2: Error rates, standard deviations and classification results with four different classifiers

Classifier	Error rate	SD	Correct (%)
Linear Bayes Normal	0.2167	0.0250	78.33
Quadratic Bayes Normal	0.2222	0.0207	77.78
K-Nearest Neighbor	0.2370	0.0168	76.30
Fisher's Least Square	0.2074	0.0140	79.26

4.3.3 Principal Component Analysis

The subset of 9 features was used for PCA. The contributions of all features to the 1st and 2nd PC are graphically displayed in Figure 4.4. The 1st to the 7th PC were selected for further analysis as together they explain 93.6% of the variance. These seven PCs were input for classification between three rest and three stress conditions, treated as two classes. The guessing rate for this classification problem was thus 0.5. The standard deviation (SD) of the error rate gives information about how constant the error rate is. The resulting error rates, SDs and classification percentages are listed in Table 4.2.

4.3.4 Generalized Estimating Equations

We used a manual backward feature selection procedure to select the best features for GEE. Four features (Peak, SCRR, SDNN, MNF) were removed from the analysis. Removing the fifth feature (HR), caused the goodness-of-fit statistic to

Table 4.3: GEE coefficients C1–C5 for the selected features

Feature	C1	C2	C3	C4	C5
HR	0.376	0.277	-0.221*	0.121	0.615
SCL	0.817*	0.871*	0.884*	0.411*	0.959*
SCdiff2	1.391*	1.406*	0.641*	0.638*	1.177*
RespFreq	1.122*	0.813*	0.525*	-0.126*	1.360*
Gaptime	-1.163*	-1.090*	-0.086	0.183*	-0.895*
Intercept	-0.151	-0.021	-0.045	0.082*	-0.135

* $p < 0.05$ Wald Chi-Square test

Table 4.4: Classification results from GEE on the test sets

	Set 1	Set 2	Set 3	Set 4	Set 5	Average (%)
Rest	100.0	81.8	63.6	63.6	90.9	80.0
Stress	63.6	81.8	63.6	63.6	72.7	69.1
Overall	81.8	81.8	63.6	63.6	81.8	74.5

increase, so we decided to keep HR in the model. It had a significance value of 0.103, which is anyway very close to the 0.10 threshold that is commonly used in feature selection procedures.

The five selected features and their coefficients for the different training sets are listed in Table 4.3. The coefficients of SCL and SCdiff2 are all positive and significant. Also the HR and RespFreq coefficients are positive for four out of five training sets. These results confirm the expectations that raised from literature, suggesting that all four features are positively related to stress. The coefficients of Gaptime are negative in four out of five training sets. This indicates that a decrease in Gaptime corresponds to an increase in stress level, as was expected.

Using the calculated probabilities for each case and taking a cutoff value of 0.5, each condition could be classified either in a stress or in a rest condition. The classification results for the five test sets (22 cases in each set) are shown in Table 4.4. There were some variations in the results among the different test sets. On average, 80.0% of the rest cases and 69.1% of the stress cases were classified correctly. Overall, 74.5% of the test set cases were classified in the correct condition.

4.4   Discussion

The goal of this study was to identify those physiological features that show distinct reactions to mental stress in office-like situations. We measured ECG, respiration,

SC and EMG signals during stress and rest conditions. We selected nine non-correlating features from these four signals and analyzed the changes in feature values that were caused by stress. Using seven PCs, the percentage of correctly classified stress and rest cases was 79.3%. Secondly, using GEE, we reduced the feature set to five features and on average 74.5% of the cases of the test set were classified correctly into a rest or stress condition. Both results are higher than the guessing rate of 50%. Although there is not much difference between the results, the PCA approach performed slightly better than the GEE approach. However, the GEE approach only used five features, instead of seven PCs from nine different features in PCA. Using fewer features is an advantage, as especially in small datasets it is preferred to use as few features as possible to avoid the curse of dimensionality. Also, calculating fewer features saves processing time and power.

The fact that all nine features contributed to the first and second PC, see Figure 4.4, shows that all features contain information. Also, the five features selected for GEE were spread over the four different signals. Both observations point to the benefits of multi-signal data collection and analysis. Apparently, all signals contributed to the global stress reaction and together facilitated robust stress detection.

In Table 4.4, we can see that the rest cases are more often classified correctly than the stress cases. A similar phenomenon was observed in the PCA approach. It seems more difficult to detect stress from physiological signals than to identify that someone is at rest. A possible explanation can be found in the stress levels that subjects indicated during the study. Subjects reported low stress levels during the rest conditions, whereas during the stress conditions the reported stress levels covered a larger range (see Figure 3.3). This means that not all subjects indicated high stress levels during the stress conditions. So although it was proven in Chapter 3 that on group level, the stress conditions were perceived more stressful than the rest conditions, this did not hold for every single case. In future work, the individual subjective stress levels should be taken into account in the analysis of the physiological data to get to more personalized stress estimations.

Compared to other papers on classification of mental stress situations, our classification results were relatively low. Classification accuracies of other studies we found range from 80% [175] to 97.4% [63]. However, most other studies only used one type of stressor in their protocols and therefore they were able to tune their algorithms to detect the specific reaction induced by that stressor. In our study, three different stressors were used that were all included in the analysis. Still, an accuracy of 79.3% was achieved with PCA using a Fisher's Least Square Classifier in differentiating these three mental stress conditions from rest conditions. A similar study like ours was done by Choi et al. [29]. Their protocol consisted of five different stress conditions, interleaved with deep breathing relaxation periods. They used four features extracted from skin conductance, respiration and heart rate in a logistic model to classify between rest and stress conditions. Choi et al. obtained a classification accuracy of 81%, which is comparable to our results and

confirms the validity of our results.

The study presented in this paper has the following limitations. Our normalization method worked well for this analysis and we obtained good classification results. However, this method of normalization is not practical for real-time use of the system, because it needs a signal of several minutes including both rest and stress conditions. Clearly, it is preferable to have only a short calibration measurement that is sufficient for normalization of all future recordings. A special protocol could be developed to achieve that. Such a protocol should include baseline recording and also a stressor to be able to normalize for the reactivity to stress.

The entire protocol was performed in a controlled environment. To facilitate future use of a stress measurement system, recordings should be made in uncontrolled, real-life environments. In uncontrolled environments, many factors other than stress influence physiology. This complicates the data analysis. We need to be able to distinguish between changes in physiology due to stress and changes due to other factors. Creating context awareness by monitoring environmental factors and activity will be needed to deal with these factors and to be able to identify stress levels from the physiological signals.

The classification approach we took fits the context of long-term, real-time, continuous stress monitoring, which will be needed for many of the future applications of stress measurement. We want to be able to determine if a person is stressed or not, at any point in time. This information can be used in systems giving personal feedback about stress levels. We think that especially GEE is a useful method to achieve this goal. It is relatively straightforward to calculate a probability from a set of feature values, once the coefficients are determined. In combination with the sliding window, the calculated probabilities might be used as a continuous measure of stress level on a scale from 0 to 1. Of course, this hypothesis will need to be validated first.

In the future, we will continue towards developing methods for long-term physiological monitoring with real-time stress level calculation. A small and non-obtrusive sensor system can be used on a regular, or even daily, basis to monitor stress levels, for example in office environments. When the stress level can be calculated in real-time, direct feedback can be given to the user to help him/her regulate his/her stress level by keeping it in an optimal range.

4.5   Conclusions

We recorded ECG, SC, respiration and EMG signals during a protocol designed to include realistic office-like stress situations. We used these signals for multi-modal stress detection. We achieved a classification rate of 79.3% using PCA and a Fisher's Least Square Classifier and 74.5% using GEE. The PCA approach resulted in a higher classification rate, but the GEE approach used fewer features. Depending on the accuracy needed and the available computing power for ex-

tracting the features, the most suitable approach can be chosen. We still need to overcome the physiological differences among users and add context awareness and activity monitoring to compensate for factors other than stress influencing the physiology. In the end, these methods can be used for long-term, real-time, continuous stress monitoring on an individual basis using a small and non-obtrusive sensor system. Such a system could detect mental stress at an early stage and therefore contribute to stress prevention at work, for example.

5

Psychological and Physiological Effects of Introductory Videos in Simulation-Based Education

Educational outcomes could be improved by detecting optimal emotional arousal level for learning. The subjective experience of reality in simulation-based medical education (SBME) leads to an emotional response that may contribute to the effectiveness of the education. Two randomized controlled trials are described in this chapter, comparing the effect of immersive introductory videos with textual introductions on psychological stress, physiological activity and performance in SBME. ECG and skin conductance were recorded and self-reports were used for psychological stress. Study 1 involved a resuscitation scenario. Introduction to the scenario increased psychological and physiological stress. After the scenario, psychological stress in the video group was higher than in the textual introduction group. HR and LF decreased and HF increased during the introduction in the video group; the changes in the textual introduction group were in opposite direction. Towards the end of the protocol, psychological stress decreased but physiological stress increased, possibly caused by large physical activity during the resuscitation scenario. Study 2 involved simulated laparoscopic surgery with low physical activity. Psychological stress increased during the introduction. The only difference between the groups was in cognitive appraisal after the introduction: the video induced a threat response and the textual introduction induced a challenge response. This difference could influence the learning process and further research is needed on this topic. Individual assessment of psychological and physiological activation status may contribute to optimizing activity level for the most effective learning experience.

5.1 Introduction

Emotional arousal is strongly related to learning performance. According to the Yerkes-Dodson Law, learning performance increases with increment of emotional arousal, but only up to a point. After this point, when levels of emotional arousal become too high, learning performance decreases [173]. Therefore, to reach the best learning performance, it is crucial to detect the point of optimal arousal level, and to create elements that stimulate or inhibit emotional arousal. With integrating this information in simulation-based medical education (SBME), learning outcomes can be improved.

Over the last years SBME is increasingly used as a teaching tool in an effort to prepare health profession students for the stress of clinical work. It is assumed that a high psychological fidelity of the simulator setting creates realistic psychological challenges and thus mimics the problems of daily clinical acute care [50].

The emotional and physiological response in SBME and the effects on learning performance remain unclear, and contradictory results can be found in literature [6, 7, 9, 34, 54, 55, 60, 75, 95, 96, 97, 98, 113, 167]. This suggests that individual differences in emotional assessment of the situation are important when creating the optimal learning circumstances. This assessment, also called cognitive appraisal, is defined as the interaction of demands placed on an individual by his/her environment and a person's resources to meet the demands [52, 94, 160]. When resources are assessed as sufficient to meet the demands, the situation is assessed as a challenge. In contrast, when demands are assessed as outweighing the resources, the situation is assessed as a threat [52, 111, 160]. A difference in cognitive appraisal impacts the emotional and physiological response, which can influence the learning outcome in a positive or negative way.

Effects of acute stress on a person are both psychological and physiological. The effects can be measured in several ways: as self-assessment questionnaires, physiological measures such as heart rate, and neuroendocrine levels such as cortisol or alpha amylase [60, 114]. Because stress is a multidimensional construct it is important to assess multiple parameters when trying to define stress [44]. On top of this, differences in cognitive appraisal correlate with differences in physiological processes and performance [60]. Clarification about the psychological and physiological response to stress-inducing elements in SBME may help to create optimal learning outcomes.

It is not entirely clear which elements may stimulate or inhibit emotional arousal in SBME. It has been argued that the learner's subjective experience of reality provides an emotional response, which contributes to the simulation's effectiveness in embedding the experience in the learner's memory [76]. This subjective experience of reality might be stimulated by the use of an introductory story type, e.g. oral briefing, textual introduction or introductory videos. Results from studies in other domains about the use of varied story types suggested that video and video plus textual versions led to higher levels of both engagement and recall of particular information [83, 171]. This phenomenon has never been investigated

in SBME and can be a useful element in clarifying the learning effectiveness of SBME.

In SBME, several variables are used to measure stress with a lack of consensus. Psychological effects are mostly measured with self-assessment questionnaires, while the physiological response can be divided in variables of sympathetic nervous system activation and activation of hypothalamus-pituitary axis (HPA). Physiological measurements such as heart rate, heart rate variability and blood pressure are frequently used to measure sympathetic activation, while cortisol responses are commonly used as measurement of HPA activation, associated with a threat response. From previous research we know that performance can be enhanced or impaired by high-acuity situations [95]. Elevated stress levels can impede clinical performance on tasks that require divided attention, working memory, retrieval of information from memory, and decision making. On the contrary, memory consolidation can be enhanced with elevated stress levels [95]. Increase in cortisol is more likely to result in performance impairments on tasks of memory, attention, decision making and clinical performance [60]. Physiological stress assessment in literature shows a wide range of measurements and results [14, 118].

Two studies are described in this chapter, aimed at comparing the effect of immersive introductory videos with textual introductions on psychological stress, physiological activity and performance in SBME. Next to these outcomes that are investigated in both studies, the second study also investigates if there is a difference in cognitive appraisal, engagement and motivation between an immersive introductory video and a textual introduction. We hypothesize that an introductory video leads to more engagement causing a higher threat response, increase in psychological and physiological stress response, worse performance, but better learning than a textual introduction.

5.2 Study 1

5.2.1 Methods

We performed a prospective, randomized controlled trial. The study was conducted at a Dutch teaching hospital. This medical centre trains, on rotating basis, medical students from Maastricht University.

Participants

Participants included rotating medical school students (year four, five and six) from all departments of the hospital. Interactive courses in resuscitation are given in medical school, where the students also receive instructions on the use of simulation mannequins. No further course was given before the training. The Medical Ethics Committee of the Máxima Medical Centre Veldhoven decided that ethical approval was not required.

Scenario Introduction

Prior to the simulation scenario, participants were introduced to the scenario by either an introductory video or textual introduction. Medical information was the same in both introductions, although the video included both visual and auditory stimuli, while the textual introduction contained only textual information. The introductory video was designed and recorded by Medsim, a medical simulation centre in Eindhoven, The Netherlands, in cooperation with ambulance staff and emergency phone service. Afterwards, this video was transcribed for the textual introduction. The scenario consists of a pregnant woman involved in a serious car accident. Ambulance staff is taking care of her, calling to doctors in the hospital and finally transporting her to the hospital while her medical status declines. Suddenly an adult observer in the scenario collapses. While the ambulance staff is taking care of the pregnant woman, the participant should start resuscitation on the observer.

Study Protocol

Participants are randomly assigned to the experimental group (viewing an introductory video) or the control group (receiving a textual introduction). After written informed consent is obtained, demographic information, including age, gender, and year of education, is collected. Students do not know what to expect from the scenario before taking part in the training, and have to give agreement not to disclose the scenario to any other participants. Physiological variables (heart rate (HR), HR variability (HRV) and skin conductance (SC)) are continuously measured with an ECG-necklace and a wristband [129, 168]. This equipment is especially designed for measuring stress variables by Holst Centre/imec (Eindhoven, The Netherlands). The protocol starts with 10 minutes of rest (baseline period). Then the introduction starts. After the introduction, students move to the next room where they have to treat the patient. This comprises resuscitation of an adult on the street without specific medical equipment. The simulator used for this scenario is Resusci Anne® (Laerdal, Norway). Questionnaires about psychological stress and suspension of disbelief (engagement in the simulated setting) are filled in during several specific time points (Figure 5.1). Each scenario is immediately followed by a debriefing.

Measurements

Psychological stress, physiological variables, suspension of disbelief and performance are measured before, during and after the introduction.

Psychological Response   The six-item stress subscale of The Dutch (shortened) Profile of Mood States (POMS-NL) is used to measure emotional stress on five-point Likert scales. This questionnaire is filled in before (Q0) and after (Q1)
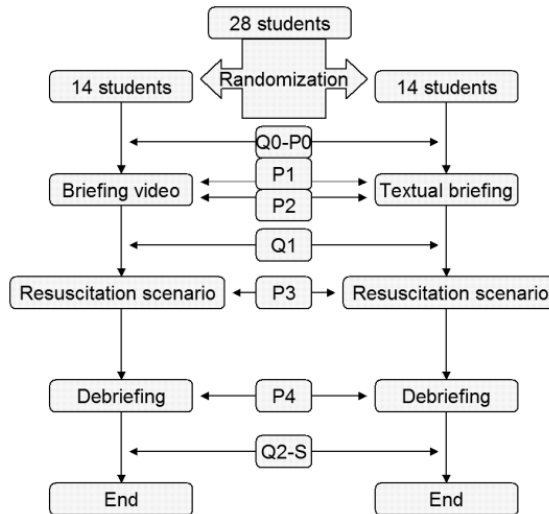
Figure 5.1: Study design with time points of measurements. Q: psychological stress questionnaire, P: physiological variables, S: suspension of disbelief questionnaire.

the introduction, and after the evaluation (Q2) (Figure 5.1). The scores on these six items are summed into a total stress score, which can range from 6 to 30.

Physiological Response The physiological features are calculated from five 45-second periods during the scenario: the last part of the baseline period (P0), halfway through the introduction (P1), the last part of the introduction (P2), the last part of the scenario (P3) and the last part of the debriefing (P4), see Figure 5.1.

HR and HRV HR is determined by direct measurement of normal-to-normal (NN) intervals of QRS complexes resulting from sinus node depolarizations. Time-domain HRV is expressed as standard deviation of the NN intervals (SDNN). Measurements of low frequency (LF) component (0.04–0.15 Hz), and high frequency (HF) component (0.15–0.4 Hz) are calculated using the Fast Fourier Transform (FFT). The results are expressed in normalized units (n.u.). Normalized units represent the relative value of each power component in proportion to the total power of low frequency plus high frequency components [23]. HF activity has been linked to parasympathetic nervous system activity, and LF activity is now accepted to be a mixture of activity of both the sympathetic nervous system and parasympathetic nervous system. The representation of LF to HF (LF/HF ratio) emphasizes the behavior of the autonomic nervous system (ANS).

Skin Conductance The skin conductance (SC), the inverse of skin resistance, is defined as the electrical conductance of the skin in microsiemens. It is used

Table 5.1: Demographics of participants

Demographic	Video	Text	p
Students (N=28)	14	14	
Age, mean (SD)	24.4 (0.9)	23.9 (1.5)	0.23
Female, n (%)	9 (64.3)	8 (57.1)	0.71
Medical school year, mean (SD)	5.2 (0.4)	5.0 (0.4)	0.31

for capturing sweat gland function, which is a variable of the ANS response. SC is expressed as absolute skin conductance level (SCL) and as signal power of the second derivative of the SC signal (SCdiff2) [58].

Suspension of Disbelief An evaluative questionnaire with specific questions about suspension of disbelief is accomplished on a 5-point Likert scale at the end of the experiment (S) (Figure 5.1).

Performance Assessment For assessment of medical performance a checklist is designed, based on Dutch resuscitation guidelines. The scenarios are videotaped and performance is independently scored by two researchers, who are blinded to the group of allocation.

Data Analysis

Statistical analyses are performed using IBM SPSS Statistics. We use the independent samples t-test (2-tailed) to compare the experimental and control group for scores of psychological and physiological variables, suspension of disbelief and performance. For changes over time we use the paired samples t-test. Psychological stress is calculated in scores of POMS-NL. Physiological variables are calculated in changes between P0 and other time points (P1, P2, P3 and P4). P-values <0.05 are considered significant. As there are no comparable investigations addressing the effect of introductory videos on psychological stress and physiological variables compared with textual introduction, sample size calculation is determined by convenience.

5.2.2   Results

A total of 28 participants took part in the scenario, of whom 14 students were randomized to the introductory video group and 14 to the textual introduction group. Participants included 17 females and 11 males. The average age of the participants was 24.1 years, ranging from 22 to 26. See Table 5.1 for complete demographic information of the two groups.
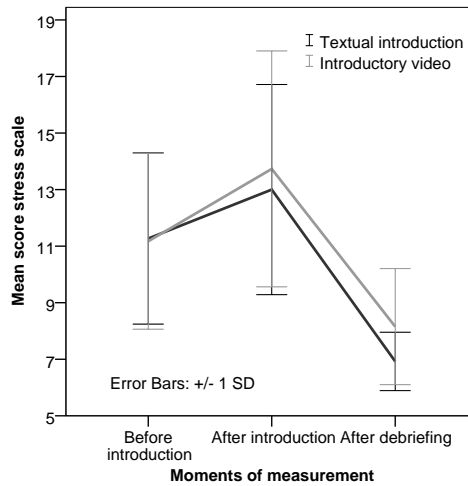
Figure 5.2: Outcome stress subscale (POMS-NL) over time in experimental and control group.

Frequency distribution of the scores of psychological stress, physiological variables, suspension of disbelief and performance were checked on normal distribution. The assumption of normality was not violated.

Mean scores with standard deviation (M(SD)) on the POMS-NL stress subscale prior to the introduction in the experimental and control group were 11.17(3.11) and 11.27(3.02) respectively, p=0.94. Immediately after the introduction, scores were 13.96(4.10) and 13.00(3.72), p=0.53. After the debriefing the experimental group scored significantly higher on the stress subscale than the control group 8.14(1.97) versus 6.86(1.02), p=0.04 (Figure 5.2). When the results of both groups are taken together, there are significant differences in stress level among all three conditions.

Physiological responses are shown in Table 5.2 and Table 5.3. By examining changes of physiological variables over time, we only found two significant differences among the different groups. Heart rate decreases in the introductory video group and increases in the textual introduction group from rest to halfway through introduction (-4.59(8.95) versus 3.43(8.95), p=0.03). The change in normalized LF heart rate variability in the introductory video group is negative whereas in the group with textual introduction the LF change is positive from rest to end of the introduction (-0.201(0.259) versus 0.023(0.241), p=0.03). For normalized HF heart rate variability the change is in the opposite direction and this change is also significantly different between the two groups.

With all participants taken together as one group some significant differences of physiological variables among the various conditions were found (see Table 5.4). The LFnu decreases and HFnu increases from P0 to P1, which indicates a decrease of stress. However, SCL increases in the same period and this indicates an increase

Table 5.2: Psychological responses over time in experimental and control group—changes from before introduction to halfway through introduction and to end of introduction

Feature	Video ΔP0-P1 Mean(SD)	Text ΔP0-P1 Mean(SD)	*p*	Video ΔP0-P2 Mean(SD)	Text ΔP0-P2 Mean(SD)	*p*
HR (bpm)	-4.59(8.98)	3.43(8.95)	0.03*	-1.02(8.55)	3.77(7.97)	0.14
LF (n.u.)	-0.105(0.305)	-0.139(0.202)	0.74	-0.201(0.259)	0.023(0.241)	0.03*
HF (n.u.)	0.105(0.305)	0.139(0.202)	0.74	0.201(0.259)	-0.023(0.241)	0.03*
LF/HF	-31.69 (117.20)	-2.00(4.89)	0.35	-33.63 (116.64)	0.62(4.26)	0.28
SDNN	-0.003(0.034)	-0.022(0.045)	0.16	0.002(0.047)	-0.020(0.043)	0.20
SCL (μS)	0.085(0.212)	0.082(0.157)	0.99	0.04(0.15)	0.11(0.20)	0.35
SCdiff2 ($\cdot 10^{-18}$)	-54712 (188192)	-606(4905)	0.32	-54674 (189766)	9739(24460)	0.22

* $p < 0.05$ independent samples t-test

Table 5.3: Psychological responses over time in experimental and control group—changes from before introduction to scenario and to debriefing

Feature	Video ΔP0-P3 Mean(SD)	Text ΔP0-P3 Mean(SD)	*p*	Video ΔP0-P4 Mean(SD)	Text ΔP0-P4 Mean(SD)	*p*
HR (bpm)	38.56(12.11)	37.72(12.58)	0.86	0.44(5.25)	-2.65(6.26)	0.18
LF (n.u.)	0.157(0.212)	0.190(0.205)	0.69	-0.104(0.231)	0.046(0.248)	0.12
HF (n.u.)	-0.157(0.212)	-0.190(0.205)	0.69	0.104(0.231)	-0.046(0.248)	0.12
LF/HF	-6.66(56.09)	13.09(17.52)	0.24	-1.86(3.36)	0.25(5.18)	0.23
SDNN	-0.03(0.03)	-0.03(0.03)	0.91	-0.03(0.04)	-0.04(0.04)	0.53
SCL (μS)	2.21(1.92)	2.26(2.03)	0.96	2.55(2.79)	1.96(1.52)	0.50
SCdiff2 ($\cdot 10^{-18}$)	557109 (1156977)	1838470 (4852151)	0.36	-3188 (155106)	49064 (92702)	0.29

Table 5.4: Physiological responses over time for all participants

Feature	ΔP0-P1 Mean(SD)	p	ΔP0-P2 Mean(SD)	p	ΔP0-P4 Mean(SD)	p
HR (bpm)	-0.58(9.70)	0.75	1.37(8.47)	0.40	-1.16(5.90)	0.32
LF (n.u.)	-0.122(0.255)	0.02*	-0.089(0.270)	0.09	-0.026(0.247)	0.59
HF (n.u.)	0.122(0.255)	0.02*	0.089(0.270)	0.09	0.026(0.247)	0.59
LF/HF	-16.85(82.79)	0.29	-16.50(82.85)	0.30	-0.765(4.45)	0.38
SDNN	-0.012(0.040)	0.12	-0.009(0.046)	0.31	-0.031(0.038)	<0.01*
SCL (μS)	0.083(0.182)	0.03*	0.078(0.181)	0.04*	2.25(2.20)	<0.01*
SCdiff2 ($\cdot 10^{-18}$)	-26657 (130831)	0.30	-21274 (134147)	0.42	23905(24426)	0.34

Feature	ΔP1-P2 Mean(SD)	p	ΔP1-P4 Mean(SD)	p	ΔP2-P4 Mean(SD)	p
HR (bpm)	1.96(4.43)	0.03*	-0.88(11.0)	0.68	-2.81(10.4)	0.17
LF (n.u.)	0.033(0.231)	0.46	0.075(0.278)	0.17	0.034(0.248)	0.48
HF (n.u.)	-0.033(0.231)	0.46	-0.075(0.278)	0.17	-0.034(0.248)	0.48
LF/HF	0.346(5.08)	0.72	0.470(4.18)	0.57	-0.099(4.54)	0.91
SDNN	0.003(0.038)	0.64	-0.017(0.033)	0.01*	-0.021(0.036)	0.01*
SCL (μS)	-0.006(0.072)	0.69	2.16(2.19)	<0.01*	2.17(2.17)	<0.01*
SCdiff2 ($\cdot 10^{-18}$)	5383(17018)	0.11	50563(96924)	0.01*	45180(96672)	0.02*

* $p < 0.05$ paired samples t-test

of stress. SCL increases throughout the entire protocol. HR increases during the introduction (P1 to P2), but does not change during other conditions of the protocol. Towards the evaluation SDNN decreases and SCL and SCdiff2 increase, all indicating increase of stress.

No significant difference was found in total score of the three items about suspension of disbelief between the two groups (10.29(2.97) versus 10.57(3.41), p=0.39). Surprisingly we found a significant difference in the item "the instructions were clear" in favor of the textual group (2.86(0.95) versus 3.96(1.25), p=0.01). Resuscitation was better performed in the textual group (18.75(4.09) versus 23.49(4.25), p=0.01).

5.2.3 Discussion

The results of this study show that introduction to the scenario seems to contribute to increased psychological stress, with a peak straight after the introduction. A difference in level of psychological stress between video and textual method has not been found. This means that participants of simulation-based training start a simulation scenario on a peak of psychological stress, independent of introductory method. The physiological variables (HR, HRV, and skin conductance) show overall an increase in sympathetic and a decrease in parasympathetic activity, from

baseline to introduction, and from introduction to evaluation. However almost no differences were found between the two introductory methods. This suggests that the introduction itself results in sympathetic activation, but the way of introduction does not matter. The increase in sympathetic activity from introduction to evaluation is surprising, because we expected the stress level to decrease after the scenario and this is also shown in the psychological stress questionnaire responses. The physiological results thus indicate an increase in stress, while the self-reported stress shows a decrease. One cause of this could be the large physical activity during the scenario, which still influences the physiological parameters during the evaluation.

Introduction methods in medical simulation are these days used to influence psychological stress and physiological activity with the aim to create an effective learning situation. Yerkes and Dodson already showed in 1908 that stress, caused by interruptions, could have positive and negative effects on performance [173]. LeBlanc pointed to the positive and negative impact of stress on performance, but also reported positive effects of stress on memory consolidation, the process by which new and fragile memories are rendered into more stable and permanent memories [95].

The impact of the introduction methods on suspension of disbelief and performance is questionable. No differences between the introductory video and textual introduction were found in suspension of disbelief. This means that no superior introduction method exists to stimulate engagement in this scenario. The video group scored lower on performance and higher on psychological stress after the evaluation, compared to the textual group. The elevated stress level could be a possible explanation for the difference in performance. The increase in psychological stress could have reached a tipping point, after which performance drops [140]. On the other hand, the control group also indicated that they knew better what they had to do and this may have improved their performance.

There were several limitations in our study. First, complex logistics involving new technical measurements and availability of students limited our sample size, which may have limited our ability to detect significant differences. The observation of patterns in the students' experience of stress and objective physiological measurements may also be influenced by other confounders not included in this study design, e.g. experience, social support, motivation. Experience or extra courses could especially have affected performance, which could have influenced psychological stress level. Further studies should include other confounders.

New in our study is the possibility of continuous measuring of physiological variables during the whole training. This makes the physiological variables, which are highly sensitive, more specific and reliable. Nevertheless some measurements are missing because of technical problems. In future research reliable instruments are needed for robust physiological recording. Next to this, suspension of disbelief was only scored on three specific items in the evaluation questionnaire. A validated questionnaire is necessary to perform further research about engagement.

5.3   Study 2

5.3.1   Methods

We performed a prospective, randomized controlled study. The study was conducted at a Dutch secondary school in the southern part of the Netherlands. The Medical Ethics Committee of the Máxima Medical Centre Veldhoven approved this study as consent exempt.

Participants

The participants of the study were 38 secondary school students between 17 and 19 years old. They were all studying in their final year, had no previous experience with simulation-based education, and had biology as a subject. The students were all healthy, without any use of medication except for the oral contraceptive pill by some of the female students.

Intervention

One month prior to the actual intervention, a short-term classic education was given about basic clinical aspects of patients and laparoscopic surgery, followed by routinely practicing using predefined levels on the LapSim® (Surgical Science Ltd., Gothenburg, Sweden). The LapSim® is a simulated video-assisted laparoscopic surgery box. After practicing and obtaining informed consent, the students were randomly allocated to either the introductory video group or the textual introduction group by computer-generated random numbers. An immersive introductory video was designed with input from learners, midwives and physicians, and recorded by a medical simulation centre in the Netherlands (Medsim, Eindhoven) with cooperation of a patient actor, a physician, ambulance staff, the emergency phone service and operation room staff. It was then tested in various simulated scenarios. The introductory video simulated a woman involved in a serious car accident. The first part of the introductory video was about ambulance staff taking care of the patient, calling to doctors in the hospital, and transporting the patient to the hospital while the patient's medical status declined. The second part was about the start of a laparoscopic surgery. The video ended after introduction of the instruments in the abdomen, ready to start the surgery. The textual introduction contains a two sentence brief visual introduction in Microsoft PowerPoint. Medical information was more detailed in the introductory video compared with textual introduction. A beamer was used to project both introductory video and textual introduction on a wall. After the introduction, students were dressed like a surgeon and had to perform two predefined levels on the LapSim®.

Outcome Measures

Data collection was accomplished by two secondary school students supervised by one investigator. Data on personal characteristics were gathered using self-report questionnaires.

Cognitive Appraisal   Cognitive appraisal was assessed immediately before and after the scenario, using the method described by Tomaka et al. [160, 161]. Appraisal of demands was examined by asking the participants before and after the scenario to answer the question "How demanding do you expect the upcoming task to be?", and "How demanding was the task you just completed?". Appraisal of resources was measured by asking the participants to answer the question: "How able are/were you to cope with this task?". Answers had to be given on a 10-point Likert scale. Cognitive appraisal was then calculated as the ratio of the primary appraisal (demands) to the secondary appraisal (resources) The situation was appraised as a challenge if the resources were assessed as being equal or greater than the demands (ratio ≤1). The situation was appraised as a threat if the demands were assessed being greater than the resources (ratio >1).

Psychological Response   Psychological stress was measured at the beginning, immediately after the introduction, and directly after the scenario using the six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). Participants had to indicate their level of agreement about how they felt at the given moment on a 4-point scale. Each question was given a score of 1–4, which generates a total score of 6–24. The correlation between scores from this six-item short-form and the full length STAI is 0.95 [108].

Physiological Response   Physiological variables (heart rate, heart rate variability and skin conductance) were measured with an ECG-necklace and a wristband, as in Study 1. Mean values for the sympathetic variables were calculated for a 10 minutes rest period (baseline period), during the introductory method, and during the session on the LapSim®.

HR and HRV HR was determined in the same way as in Study 1. Only frequency domain measures of HRV were calculated because recordings of different durations influence the results of time domain measures such as standard deviation of NN intervals (SDNN) [159]. Measurements of LF component and HF component were expressed in absolute values of power (ms^2) and normalized units (n.u.).

Skin Conductance The same features were extracted from SC as in Study 1: SCL and SCdiff2.

HPA Axis Activation of the HPA axis was measured using salivary cortisol levels, which have shown a close correlation with plasma cortisol levels [41]. Because

of gender differences, the impact of oral contraceptive use, and effects of menstrual cycle phase on cortisol response, salivary cortisol samples were only collected for men [51, 72, 78, 80, 87, 123]. To minimize the effects of diurnal variations in cortisol productions, all the sessions of the men were conducted between 1PM and 4PM. Participants were asked to abstain from eating, drinking, consuming caffeine, using drugs, smoking, brushing teeth, and doing exercise 1 hour prior to the study period as this might interfere with the measurements. Samples were collected at the start and 20 minutes after introductory method because previous studies have shown that cortisol response to stressors peaks during this time [35, 77]. Participants chewed on a roll-shaped saliva collector (Salivettes, Sarstedt, Germany) for 30 seconds. The saliva collector was then placed in a collection tube and frozen until all samples were centrifuged and analyzed at the laboratory of the Máxima Medical Centre using the ELIZA technique.

Engagement and Motivation At the end of each simulation, subjects completed the 19-item "Game Engagement Questionnaire" (GEQ). The GEQ provides a psychometrically strong measure of level of engagement specifically elicited while playing video games. The GEQ has a Rasch estimate (the Rasch analog to Cronbachs alpha) of 0.83 and reliability of 0.96 [18]. Subjects also completed the 22-item Intrinsic Motivation Inventory (IMI). The subscale "interest/enjoyment" is considered the self-report measure of intrinsic motivation.

Performance Assessment The performance of the participants on the LapSim® was rated using the mean score in percentage calculated by LapSim® of two predefined levels. This calculation was based on items as number of errors, distance of the instruments, and time spent to finish the levels.

Learning Assessment Three to four months later participants were asked to play the two previous defined levels all over again. Learning was rated in the same way as performance.

Statistical Analysis

Cognitive appraisal, psychological response, physiological response and performance are expressed as absolute and change median values with interquartile ranges (IQR). Engagement and motivation are expressed as absolute value. Differences in scores between the two groups were compared with the Mann-Whitney U test, and for scores over time the Wilcoxon Signed Ranks test was used. Statistical significance was accepted at a two-sided p-value of 0.05. All analyses were performed using SPSS (Version 21, IBM, San Jose, California). As there are no comparable investigations addressing the differences in cognitive appraisal between

Table 5.5: Demographic comparison of introductory video and textual introduction group

Variable	Introductory video	Textual introduction
Number, (nr, %)	17 (43.6)	22 (56.4)
Age (yr), mean (SD)	17.35 (0.15)	17.18 (0.12)
Gender (m), (nr, %)	11 (50)	7 (41.2)

introductory videos and textual introduction, sample size calculation was determined mainly by modelling the study of Harvey et al. [60] and LeBlanc et al. [97] and by practical aspects.

5.3.2 Results

A total of 39 participants took part in the study, of which 17 students were randomized to the introductory video group and 22 to the textual introduction group. Demographics are shown in Table 5.5. There were no statistical significant differences in age and gender between the groups.

Cognitive Appraisal

Group medians and interquartile ranges (IQR) for cognitive appraisal are listed in Table 5.6. The Mann-Whitney U test showed a statistically significant difference in cognitive appraisal score straight after the introductory method ($p<0.01$). See Figure 5.3 for box plots of the scores of both groups. Post-scenario cognitive appraisal scores were not significantly different between the groups ($p=0.68$).

Psychological Stress Response

Medians and interquartile ranges (IQR) for the psychological stress response are shown in Table 5.6. The Mann-Whitney U test showed no statistically significant differences in score on the STAI-S questionnaire between the two introductory methods. In both groups STAI-S total score increased significantly from starting period to straight after the introduction method (Table 5.7, Figure 5.4).

Physiological Stress Response

Medians and interquartile ranges (IQR) of the physiological response are shown in Table 5.6. No significant differences were found between the two groups on the three points in time. Changes over time are shown in Table 5.7. Statistically significant differences were found in HR (decrease), LFnu (decrease) and HFnu

Table 5.6: Psychological and physiological stress variables in the experimental and control group at different points in time

Variable	Point in time	Video N	Median (IQR)	Text N	Median (IQR)	p
Cognitive appraisal	After introduction	17	1.17 (0.3)	22	1 (0.42)	<0.01*
	After LapSim	17	1.2 (0.91)	22	1.25 (0.85)	0.68
STAI	Before calibration period	17	9 (3.5)	22	7.5 (3.3)	0.32
	After introduction	17	11 (5)	22	10.5 (4)	0.57
	After LapSim	15	11(5)	22	10.5 (3.5)	0.72
HR (bpm)	T0	15	86.28 (23.82)	17	88.17 (13.43)	0.93
	T1	14	83.18 (19.62)	12	83.99 (23.65)	0.64
	T2	14	89.29 (17.1)	17	89.28 (12.63)	0.66
LF (ms^2)	T0	15	4944 (6627)	17	6395 (11094)	0.22
	T1	13	4330 (3816)	12	8381 (8430)	0.12
	T2	13	2570 (2404)	17	3492 (2875)	0.06
HF (ms^2)	T0	15	914 (1647)	17	1486 (2287)	0.44
	T1	13	1703 (2113)	12	1980 (3774)	0.38
	T2	13	624 (1123)	17	770 (716)	0.72
LF (n.u.)	T0	15	0.8225 (0.0781)	17	0.8705 (0.1036)	0.22
	T1	13	0.7305 (0.1453)	12	0.7591 (0.3166)	0.7
	T2	13	0.7886 (0.1623)	17	0.8372 (0.1521)	0.37
HF (n.u.)	T0	15	0.1775 (0.0781)	17	0.1295 (0.1036)	0.22
	T1	13	0.2695 (0.1453)	12	0.2409 (0.3166)	0.7
	T2	13	0.2114 (0.1625)	17	0.1628 (0.1521)	0.37
LF/HF	T0	15	4.64 (2.35)	17	6.72 (5.08)	0.22
	T1	13	2.74 (2.13)	12	3.98 (8.84)	0.7
	T2	13	3.73 (6.01)	17	5.14 (6.34)	0.37
SCL (μS)	T0	10	0.21 (0.85)	13	0.12 (0.5)	0.5
	T1	10	0.3 (0.82)	14	0.09 (0.71)	0.91
	T2	10	0.08 (1.11)	15	0.76 (0.22)	0.96
SCdiff2	T0	10	2.78E-15 (1.96E-14)	13	1.55E-16 (1.91E-14)	0.8
	T1	10	3.55E-16 (2.05E-14)	12	5.80E-17 (1.41E-14)	0.79
	T2	10	8.66E-17 (3.73E-14)	15	2.55E-16 (3.28E-15)	0.7
Salivary cortisol (only men) (mmol/L)	Before calibration period	7	9.66 (12.9)	11	10.13 (4.72)	0.68
	20 minutes after introduction	7	11.24 (9.69)	11	9.99 (15.04)	0.93

T0: Calibration period, T1: Introduction period, T2: LapSim period. * $p < 0.05$ Mann–Whitney U test
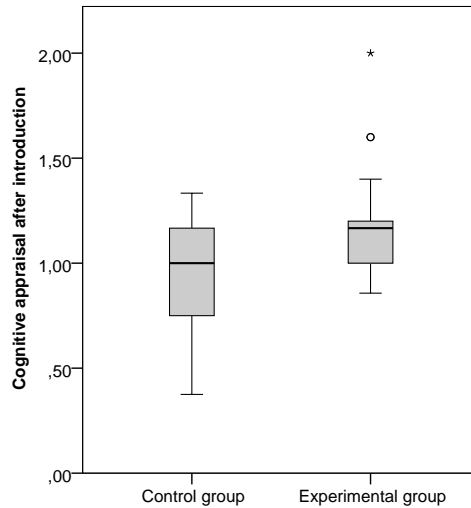
Figure 5.3: Cognitive appraisal of the control group and the experimental group after the introduction.

Table 5.7: Psychological and physiological responses over time in experimental and control group

Variable	Interval between points in time	Video		Text	
		N	p	N	p
Cognitive appraisal	After introduction – After Lapsim	17	0.55	22	0.02*
STAI	Before calibration period – After introduction	17	<0.01*	22	<0.01*
HR	T0–T1	14	<0.01*	12	0.21
LF	T0–T1	13	0.35	12	0.7
HF	T0–T1	13	0.08	12	0.94
LF/HF	T0–T1	13	0.1	12	0.7
SCL	T0–T1	10	0.39	13	0.04*
SCdiff2	T0–T1	10	0.29	12	0.45
Salivary cortisol (only men)	Before calibration period – After introduction	7	0.5	11	0.48
Performance	Experiment – 3 months after experiment	10	0.17	12	0.75

T0: Calibration period, T1: Introduction period. * $p < 0.05$ Wilcoxon Signed Ranks test

Figure 5.4: STAI scores of baseline period and after introduction for both groups.

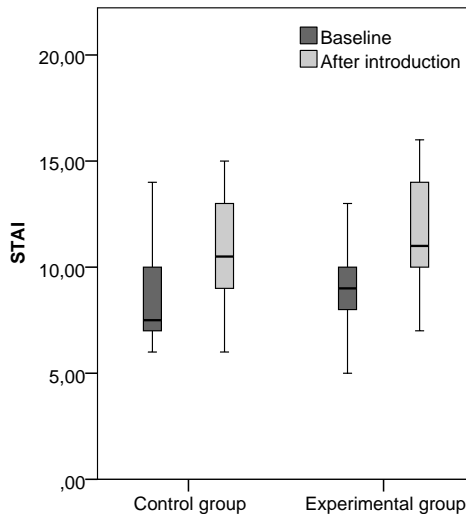(increase) in the introductory video group between starting period and straight after the introduction video. Skin conductance level decreased significantly in the textual introduction group between starting period and straight after the introduction video. No other significant differences were found on scores of physiological stress variables between the two introductory methods and over time. Results of the salivary cortisol measurements are shown in Table 5.6 and Table 5.7. Neither significant differences between the two groups, nor differences over time were found.

Engagement, Performance, Learning and Motivation

Group medians and IQR for engagement, performance assessment, learning assessment and motivation are listed in Table 5.8. No differences between the groups were found. Also the change in performance from initial experiment to 3 months later was not significant for both groups (Table 5.7).

5.3.3   Discussion

This study showed a difference in cognitive appraisal between an immersive introductory video and a textual introduction. An immersive introductory video led to a threat appraisal, while a textual introduction led to a challenge appraisal. However, no difference was found in psychological stress between the groups.

The psychological stress appraisal increased over time in both groups from before the baseline to after the introduction. Two variables of physiological stress changed

Table 5.8: GEQ, performance and motivation in the experimental and control group

| | Video | | Text | | |
Variable	N	Median (IQR)	N	Median (IQR)	p
Total score GEQ	15	30 (6)	22	31 (9.3)	0.84
Performance	16	64 (25.3)	21	62.5 (31)	0.79
Performance after 3 months	10	69.75 (13.9)	12	76.25 (18.6)	0.28
Intrinsic motivation inventory	17	36 (16)	22	35 (10.5)	0.64

over time in the introductory video group: HR and LFnu decreased while HFnu increased from baseline to video introduction. These changes could indicate a decrease in physiological activation. In the textual introduction group, a decrease in skin conductance was found from baseline to textual introduction. This change also indicates a decrease in physiological activation. Finally no difference was found between the two groups and between the different points in time for salivary cortisol, engagement, motivation, performance and learning.

Previous studies have shown that a threat appraisal, with cortisol response, has different effects on memory and attention compared with a challenge appraisal. Conclusions about the effects of cognitive appraisal on three components of memory—working memory, consolidation memory, and retrieval memory—can be drawn from previous studies in SBME and other domains. First, a threat response appears to impair working memory, the capacity to store and manipulate information for brief periods, while a challenge response does not [43, 106]. Secondly, an elevated stress response, especially if cortisol is increased, has been associated with enhancement of memory consolidation, the process in which new and fragile memories are converted into more stable and permanent memories [21, 170]. However, extremely high levels of HPA axis activation will impair this process [1]. Finally, a threat response leads to impairments on memory retrieval, tasks that require remembrance of previously learned information, with no impairments occurring in individuals who show a challenge response [22, 38, 88, 97]. As such it is important to identify factors that are associated with cognitive appraisal in SBME, such as the introductory method, to enhance teaching and learning. It was shown in this study that an immersive introductory video led to a threat appraisal and therefore could be beneficial for use in SBME to reach a better learning effect. However, this effect did not emerge from the performance measures of the LapSim® protocol. No differences in performance were found between the two groups. Also no significant change in cortisol level was found before and after the protocol. The number of subjects for the cortisol analysis was relatively small because only men were included in this analysis. The low number of subjects may have led to a lack of statistical power to prove a cortisol increase.

An interesting observation is the increase in psychological stress for both groups

while at the same time indications of decrease of physiological stress were measured in both groups. The mismatch of two stress modalities is a well-known problem in stress research and is described by Van Doornen and Houtveen [40]. First, there are measurement errors in both the psychological and the physiological stress assessment (e.g. recall bias and random fluctuations of physiological features). Furthermore, there are different physiological stress reactivity patterns among persons, which make a general assessment of stress difficult. Finally, the translation of stress feelings in the brain to the peripheral physiological measures is a complex process that involves many intermediate steps, influenced by various neural and hormonal factors. This all leads to the physiological measures being an indirect and somewhat unreliable indication of the subjective psychological stress experience.

There were no differences between the groups in engagement, performance and motivation. So although more engagement and worse performance were expected from the video group because of a threat appraisal, these hypotheses were not confirmed. The median scores in the GEQ were 30 for the video group and 31 for the text group, whereas the total range of possible scores was 19–57. The median scores were relatively low and indicate that the subjects were not really engaged in the protocol in general. The overall low engagement might have prevented a group difference to occur.

Complex logistics involving new technical measurements and availability of students limited our sample size, which is a limitation of this study and may have limited our ability to detect differences between the two experimental groups and over time.

5.4 General Discussion and Conclusions

We performed two studies on evaluation of introductory methods for simulation based education. The first study showed that both an introductory video and a textual introduction stimulate psychological and physiological stress. Following this first study, there were some recommendations that arose, which were addressed in the second study. As opposed to the first study, the protocol of the second study involved little physical activity. We also tried to eliminate confounders such as experience as much as possible. The participants in the second study were secondary school students who had no medical training and no experience in working with the LapSim®. Also all participants followed biology in school so we expected an interest in medical scenarios and therefore motivation to fulfill the training exercise. As suggested after the first study, we used a dedicated questionnaire for assessment of engagement in the second study.

In both studies the introduction led to an increase in psychological stress level as indicated by self-report questionnaires. This activation is an important process in creating effective simulation training, because minimal or no activation could result in low cognitive engagement and less long-term retention than targeted. There were no differences in psychological stress between the two groups, though.

The introductions were thus effective but no difference was found between video and text introductory methods. Despite these similarities we found a difference in cognitive appraisal between the two groups in the second study. The introductory video led to a threat response, while the textual introduction led to a challenge response. Although no differences in cortisol response, performance and learning were found in this study, this result indicates theoretically that introductory methods can be used to influence the learning process. Further research is necessary for investigation of details of this process.

Measuring physiological stress remains a challenge. Technical problems made it hard to get information for all participants. In future studies researchers in medical simulation should work in close contact with technical personnel to overcome these technical problems. Other factors such as individual differences and artifacts caused by movement also make it hard to draw conclusions. Nevertheless this is the first time that technical wireless equipment is used to measure physiological stress in medical simulation.

Knowledge of individual psychological and physiological activation status and possibilities of inhibiting or stimulating activation can be a focus for effective learning in simulation training. In this process continuous measuring can be of great interest. An opportunity could be created to measure activation level of participants so educators can influence this level to meet the optimal level for the most effective learning experience. This influencing could be especially interesting in team training where participants have different responsibilities and may exhibit differences in activity level. In future research the outcomes of the new specialized techniques presented here can also be combined with outcomes of other stress parameters, such as biochemical levels (cortisol and catecholamines), to make a clear differentiation of sympathetic activity with and without specific stress responses. Reaching the perfect activity level of participants for effective learning can be an educator's goal in future simulation training approaches with the overarching goal of improving patient safety and outcomes.

6

Modelling Mental Stress in Real-World Conditions Using Ambulatory Measured Physiological Variables

The number and severity of health issues due to chronic stress is increasing in recent years. Creating awareness by measuring stress build up and providing feedback on current stress level can help alleviating conditions leading to chronic stress. Assessing the stress levels by monitoring the physiological reaction of the body showed promising results in controlled lab environments. In this work we evaluated the feasibility of using physiological monitoring in assessing stress in an uncontrolled, real-world environment. We designed and executed a study to measure physiological parameters, reference measures for stress and context information. ECG, EMG from the upper trapezius muscles, respiration, skin conductance and skin temperature were monitored. As reference measures we used self-report stress levels and salivary cortisol. We collected information on eating, drinking, posture and physical activity. Analysis showed that there is no consistency among the physiological responses of different subjects when predicting stress levels. However, using regression we developed personal stress models that can predict mental stress levels at an individual level with reasonable accuracy. This shows that the proposed method and such real-world physiological monitoring can be used to assess stress level of a person.

6.1 Introduction

More than a half of all European workers considered work-related stress to be common in their workplace. Stress is the second most frequently reported work-related health problem in Europe [46]. Short-term stress is not considered harmful and often even beneficial. It can increase motivation, adaptation and reaction to the environment. Chronic stress, however, causes various negative effects. Chronic stress (work-related or caused by other factors) can cause several mental and physical health problems, such as burnout and depression, cardiovascular disease and musculoskeletal problems [99]. The key to solving these problems is the prevention of chronic stress. One way to prevent short-term stress from becoming chronic is by measuring stress, raising awareness of stress levels, and, if necessary, creating awareness, so a user is able to change his level of stress. This can be done by measuring stress and providing feedback to the user, but the key assumption here is that we are able to measure stress in an objective, quantitative way.

Stress activates the Sympathetic Nervous System (SNS) [162]. Activation of the SNS induces various physiological reactions in the body, including increase of sweat production, heart rate and muscle activation [57, 107].

Given this physiological activation, many researchers attempted to measure stress by measuring physiological parameters (see for example [65, 105, 130, 174]). Most of these studies on the physiological response to mental stress used a specific, short-term, stressor in a controlled environment. These stressors include for example the Stroop test, mental arithmetic tests and memory tests. However, there is no conclusive evidence that stress in the real world induces the same physiological response as the response that is induced by these short-term and controlled stressors [73]. Stressors in laboratory conditions are often much more acute than real-world stressors. Real-world stress often involves work stress, which is of a longer-term nature than the laboratory stressors. Although there are similarities between physiological responses to laboratory stress and real-world stress, it is of high importance to study stress responses in real-world conditions. This is exactly what our study was designed for: we measure stress in a real-world environment.

There are some examples in literature of studies that measured physiology in an ambulatory setting to assess stress level. Myrtek and Brügner [120] and Myrtek et al. [119] conducted several studies in order to measure emotions in everyday life. They used a 2.5 kg ambulatory system that recorded ECG and motion. Wilhelm and Roth [169] recorded ECG, respiration, skin conductance, finger temperature, ambient temperature and body movement with a 1 kg wearable system. The signals were recorded during a short flight and subjects included phobic and control subjects. There was a study by Healey and Picard in a driving condition [63]. Their sensors were connected to a computer in the car and therefore the system was not usable outside the car. Healey et al. [61] conducted another study in an ambulatory environment, recording skin conductance, heart rate and activity data during five days to detect emotions. Plarre et al. [131] measured ECG, skin conductance, skin temperature, ambient temperature, acceleration and respiration in

two studies. Only ECG and respiration were used in their data analysis, though. The first study was in a lab environment and the stress models that were constructed from this condition were applied to the data collected during the second study in a natural environment.

Our study builds on these previous studies but is more comprehensive in terms of a larger number of data modalities collected and a more practical system of small, wearable and wireless sensors that is used. Also, no initial lab study was done as we immediately started measuring in a natural environment. All physiological variables that we record are used in the analysis and employed to develop models for measuring stress based on a number of reference measures.

The research question that we want to address is: *Can we model stress in real-world conditions using physiological features?* We have developed a data collection protocol for ambulatory recording of various physiological parameters, context parameters and reference parameters to assess real-world stress reactions. In this paper, we elaborate on the protocol and the data we recorded following the protocol. We discuss the results we obtained in developing a general model for stress measurement and when we developed individual models. We show what variety of models do occur and their potential contribution towards effective prevention of chronic stress by measuring and managing stress.

6.2 Methods

6.2.1 Study Design

As stated before, many other studies applied short-term stress and/or were conducted in controlled environments. Our novel study design is developed for long-term, ambulatory monitoring of physiological variables related to real-world stress.

Subjects

For this study it was necessary to recruit subjects who experience fluctuating stress levels. Only when stress levels change over time, the physiological changes associated with stress can be identified. Also, we need to prevent substantial influences of physical activity in the recordings. Ideally, subjects have a sedentary job in which they spend most of their time sitting physically inactive at their desk or in meetings.

Given these two criteria, we chose a group of researchers to be our subjects. They conduct studies independently, deal with participants in their studies, write reports and papers on their results etc. This diverse set of tasks and the corresponding deadlines are likely to induce stress from time to time.

Subjects were healthy volunteers with no cardiovascular disease, known problems with or treatment for high or low blood pressure, muscle diseases, psychological

problems or chronic high stress level / burn out. They did not use medication that could influence any of the recorded parameters.

The study ran for multiple days and much data was collected from every subject. Therefore we chose to include no more than 10 subjects.

Measurement Protocol

The study was designed as a long-term experiment. We recorded for 5 days during working hours (about 8 hours per day).

Before the start of the study, subjects received instructions about the inclusion criteria, procedures and potential risks of participating. All subjects signed informed consent.

One day prior to the data collection, the subjects filled in the Daily Stress Inventory (DSI). This was done because the response to this questionnaire is likely to be different on the first day, compared to consecutive days [16].

On the first day of data collection, the investigator taught the subjects how to install the sensors and the subjects received the saliva sampling equipment (Salivette) and the mobile phone with questionnaires. The experimental procedure was explained. Every next day, the subjects applied the sensors themselves. During the measurement days, the subjects were prompted to answer the questionnaires every 30 minutes and to collect saliva samples every 2 hours. Subjects were encouraged to perform their normal daily activities. At the end of each day, the subjects removed the sensors and recharged them, and they filled in the DSI.

6.2.2 Physiological Parameters

We identified the physiological parameters that clearly correlate with mental stress from the literature study described in Chapter 2. The parameters derived from heart rhythm, blood pressure and skin conductance were most often reported to be influenced by mental stress. Continuous noninvasive blood pressure measurement is not feasible, so we measured heart rhythm and skin conductance in our ambulatory stress monitoring protocol. Furthermore, we also recorded respiration, surface electromyogram (EMG) from the upper trapezius muscles in the shoulders and skin temperature from the wrist. Parameters derived from these signals correlate to mental stress, but not as clearly as the parameters derived from heart rhythm, blood pressure and skin conductance (see Chapter 2). We measured these additional parameters to include some extra information in our data. In real-world conditions it can be useful to record extra information that enables compensation for artifacts and influences from context factors. This principle is called triangulation [19].
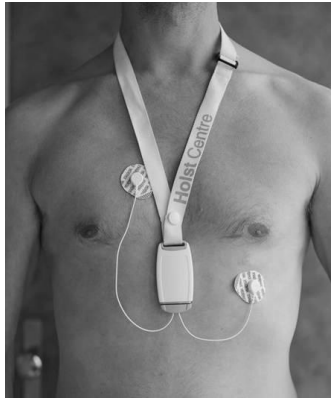
Figure 6.1: ECG necklace sensor node.

Figure 6.2: Wrist band sensor node.

Measurement Equipment

We used the wireless body area network system developed in Holst Centre/imec (Eindhoven, The Netherlands) to record the physiological signals. The system consists of up to six low-power wearable sensor nodes that are synchronized in time and that can measure various physiological parameters. For details, see [129, 168].

The 'ECG necklace' sensor node (see Figure 6.1) records physiological signals and 3D acceleration. We used it for ECG, respiration and EMG recording. The data was stored locally on a microSD card.

The 'wrist band' sensor node (see Figure 6.2) measures the tonic and phasic components of the skin conductance from the wrist, 3D acceleration, skin temperature on the wrist, skin relative humidity, ambient temperature and ambient relative humidity. As for the ECG necklace sensor node, data was stored locally on a microSD card.

Signal Processing and Feature Extraction

An overview and description of the features that were calculated from the measured signals is shown in Table 6.1. The features were chosen based on literature on stress detection from physiological signals (see for example Chapter 3 and [58, 63, 149, 174]). All signal processing was done using MATLAB.

- LF, HF, and LFHF were calculated after applying a Hann window on the interpolated heart rate signal.

- The EMG signals were filtered to remove power-line interference and low frequency noise. A bandpass filter with cutoff frequencies at 20 Hz and 230 Hz and notch filters at 50 Hz and higher harmonics 100, 150 and 200 Hz were used.

- RespFreq was determined as the main frequency component of the power spectral density of the respiration signal in the range 0.05–1 Hz.

- For detection of skin conductance response rate (SCRR), the SC signals were smoothed using a 3s moving average filter. Then the derivative was smoothed using a 1s moving average filter and SC responses were defined as peaks that followed a derivative value of at least 0.0093 μS/s. This value was empirically found to best detect the responses from a dataset collected earlier. Before calculating the signal power in the second difference of the SC signal (SCdiff2), the SC signals were smoothed using a 5s moving average filter and the second derivative was smoothed three times using a 5s moving average filter. Ohmic Perturbation Duration (OPD) of the SC is defined as the relative time of responsiveness. It varies between 0 and 1 and indicates the relative time that the second difference of the SC signal exceeded a certain threshold. See also [58]. ScPh is the power of the phasic skin conductance signal, which is a high-pass filtered version of the tonic skin conductance signal.

6.2.3   Reference Measures

We used two types of reference measure for the 'true' mental stress levels: cortisol and self-reports.

1. The test subjects took saliva samples during the recording days that were analyzed for cortisol levels. The hormone cortisol is released in stressful conditions and is therefore an indicator of stress level [107]. There are aspects that need to be taken into account when using cortisol levels for this purpose. Cortisol follows a diurnal rhythm and therefore the cortisol levels change over time, independent of experienced stress. Also, there is a delay between the occurrence of a stressful event and the corresponding rise in salivary cortisol level. Furthermore, cortisol level is not only influenced by *psychological* stress. Any kind of stressor acting on the body influences corti-

Table 6.1: Overview of features extracted from the measured signals

	Feature	Extracted from	Abbreviation	Meaning
1	Heart rate	ECG	HR	Mean heart rate
2	Standard deviation interbeat intervals	ECG	SDNN	Mean standard deviation of the interbeat intervals
3	Low frequency heart rate variability	ECG	LF	Heart rate variability in the 0.04–0.15 Hz band
4	High frequency heart rate variability	ECG	HF	Heart rate variability in the 0.15–0.4 Hz band
5	LF/HF ratio heart rate variability	ECG	LFHF	Ratio of the low and high frequency of heart rate variability
6	RMS of the EMG signal	EMG	RMSEMGl	Root mean square value of the left EMG
7	Static load	EMG	Staticl	10th percentile of rank ordered left EMG RMS values
8	Median load	EMG	Medianl	50th percentile of rank ordered left EMG RMS values
9	Peak load	EMG	Peakl	90th percentile of rank ordered left EMG RMS values
10	Mean EMG frequency	EMG	MNFl	Mean frequency of the magnitude of the left EMG frequency spectrum
11	Median EMG frequency	EMG	MDFl	Frequency at which the surface on the left side equals that of the right side of the magnitude of the left EMG frequency spectrum
12–17	RMSEMGr, Staticr, Medianr, Peakr, MNFr, MDFr			EMG features from the right side
18	Respiration frequency	Resp	RespFreq	Mean respiration frequency
19	RMS of the respiration signal	Resp	RMSResp	Root mean square value of the respiration signal for estimating tidal volume changes
20	Skin conductance level	SC	SCL	Mean level of skin conductance
21	Skin conductance response rate	SC	SCRR	Mean number of skin conductance responses per second
22	Skin conductance second difference power	SC	SCdiff2	Signal power in the second difference of the skin conductance signal
23	Ohmic perturbation duration skin conductance	SC	OPD	Relative time of responsiveness of the skin conductance signal
24	Phasic skin conductance	SC	SCPh	Power of the phasic skin conductance signal
25	Skin temperature	Temp	Temp	Mean skin temperature measured on the wrist

ECG: electrocardiogram, SC: skin conductance, Resp: respiration, EMG: electromyography, Temp: temperature

Figure 6.3: Screen shot of mobile phone application (in Dutch) asking for stress level from "not at all" to "extremely".

sol level. Saliva samples were taken at 10am, noon, 2pm and 4pm. Subjects did not eat or drink 30 minutes before the sampling time.

2. We kept track of the subjects' stress level during the day by using self-reports. The subjects carried a mobile phone running an application that inquired about the stress level of the user every 30 minutes. Thirty minutes seemed a suitable interval to collect data with a reasonable time window to capture real-life stressors without asking input too often from the subjects. Because it is unclear how the time scales of stress perception and physiological response relate, we chose to ask the subjects to indicate both their current stress level and their average stress level over the past 30 minutes on a Visual Analog Scale (VAS) ranging from "not at all" to "extremely" stressed. For a screen shot of the application where the subjects can indicate their stress level, see Figure 6.3. Furthermore, the subjects filled in the Daily Stress Inventory (DSI) at the end of each measurement day [16].

6.2.4   Context Information

With ambulatory recordings, it is crucial to gather context information next to the physiological data. Therefore the mobile phone application mentioned before also inquired about a number of context factors every 30 minutes: activity level (none, mild, moderate, heavy), posture (sitting, standing), consumption of food, drinks (coffee, tea, energy drink, alcohol) and cigarettes.

Furthermore, we used accelerometers to detect physical activity and posture. All physiological sensor nodes recorded acceleration, so we could detect activity from the wrist, shoulders, chest and hip. Also, we placed a sensor node with accelerom-

eter on the upper leg to determine the orientation of the leg and detect if a person was sitting or standing. Next to acceleration, the wrist band recorded ambient temperature and relative humidity.

6.2.5   Analysis Methods

Since we have three reference measures for stress, we took a different segment of physiological data for each reference measure:

- For the self-report on current stress level we calculated the feature values from two minutes of data just before the self-reported stress level was given.

- For the self-report on average stress level over the past 30 minutes we calculated the feature values from 30 minutes of data just before the self-reported stress level was given.

- For the cortisol values we calculated the feature values from 50 minutes before the sampling time to 20 minutes before the sampling time (30 minute segment). This was done because cortisol level has a delay of approximately 20–30 minutes between a stressful situation and the increase of salivary cortisol level [80].

The physiological signals we measure are influenced by a number of other factors next to mental stress. Data points that could be influenced by some of these factors were excluded from the analysis to create a higher chance that the physiological changes that were observed were from mental stress. The data points just after the subjects had eaten, drunk coffee, tea, energy drink or alcohol, smoked or performed moderate or heavy physical activity were excluded.

As a next step, scatter plots were made of every reference measure against the feature values of every feature, for all subjects separately. Possible outlier points were identified from the scatter plots. The original physiological data of these possible outliers was checked and if the data seemed to be of bad quality in this segment (due to malfunctioning sensors, motion artifacts, etc.), the data point was excluded from further analysis. Sometimes the quality of a particular signal was bad over an entire recording day, then that recording was excluded completely from the analysis.

After these two steps of excluding data points influenced by context factors and excluding outliers, the size of the dataset was about half of the original dataset. A substantial amount of data was excluded, but the data that was left was clean and of good quality.

Because we often derived multiple features from the same physiological signal, one could expect that some of these features correlate. Such a situation is undesirable for further statistical analysis, because it causes multicollinearity, which can lead to overfitting of the regression models. To prevent multicollinearity, the correlations among all features were investigated for every reference measure and for every subject. Absolute correlation values higher than 0.7 were considered problematic.

If necessary, features were excluded from analysis such that no correlations higher than 0.7 existed in the remaining feature set. No problematic correlations were found among features from different sensors.

Finally, we used forward regression in SPSS to investigate the relations between the physiological features and the reference measures. In a forward regression, the features that show high partial correlation to the outcome variable are entered one by one (we chose an F-to-enter criterion of 0.10). We applied forward regression to all data points together and to every subject individually.

While performing the regression analyses, it turned out that for some features for certain subjects there were too many data points excluded. A very low number of valid cases was left to be used in the regression resulting in a perfect fitting model. When this happened, we deleted the feature(s) with the low number of valid cases from the list of features eligible to be entered into the regression model and ran the regression analysis again.

6.3   Results

In this section we first discuss the results of the regression on the self-reported current stress values. Both on a general level with data of all subjects included, and on an individual level by modelling stress separately for every subject. Next, we do the same for the regression on the self-reported average stress level over the past 30 minutes. Finally, we discuss the results of the regression on cortisol levels.

The results of the regression of the physiological feature values on the self-reported values for current stress level are shown in Table 6.2. The features that were selected by the selection criterion of the forward regression vary considerably among subjects. Also the adjusted R^2 values vary. Only the model for subject 10 shows a lower R^2 value than the general model for all subjects. The ANOVA tests were significant ($p < 0.05$) for most of the models, except for subjects 3 ($p=0.062$), 9 ($p=0.058$) and 10 ($p=0.093$). For subject 8, none of the features met the inclusion criterion, so no features were entered into the equation.

The results of the regression on the self-reported average stress level over the past 30 minutes are shown in Table 6.3. Again the feature sets that were selected and the adjusted R^2 values differ among subjects. Again, only one subject's model shows a lower R^2 value than the general model. The ANOVA tests were all significant, except for subject 10 ($p=0.076$). As in the regression on the current stress level, for subject 8 no features were entered into the equation because none of the features met the inclusion criterion.

In the regression of the physiological feature values on the cortisol values, time was entered as variable to control for the diurnal rhythm that cortisol level follows. Results of the regression are shown in Table 6.4. No features were entered into the regression model for subjects 7 and 8 because the features did not meet the inclusion criterion. Also for subject 5, no model could be fit on the available

Table 6.2: Results of regression analysis on current stress level

Subject	Number of input features	Selected features	Adjusted R^2
all	25	HR, HF, SCRR	0.112
1	14	Staticl, RMSResp, OPD	0.291
2	17	SDNN, HF, Medianl, RMSEMGr, RMSResp	0.464
3	16	SCL	0.161
4	17	MNFl	0.340
5	12	RMSResp	0.174
6	12	HR, Staticr, Temp	0.481
7	11	Temp	0.223
8	15	–	–
9	16	Staticr, Peakr	0.135
10	13	OPD	0.102

Table 6.3: Results of regression analysis on average stress level over the past 30 minutes

Subject	Number of input features	Selected features	Adjusted R^2
all	25	HR, HF, Temp	0.142
1	14	HR, RMSEMGl, MDFr, RespFreq, Temp	0.630
2	15	HR	0.125
3	10	HR, HF	0.225
4	10	MDFr, MNFr, RMSResp	0.268
5	12	RMSResp	0.197
6	11	HR, SDNN, LF, Staticr, MDFr, RespFreq, RMSResp, SCL, SCRR, Temp	0.921
7	9	Temp	0.368
8	16	–	–
9	15	MNFl, SCRR	0.181
10	11	Peakr	0.151

Table 6.4: Results of regression analysis on cortisol level

Subject	Number of input features	Selected features	Adjusted R^2
all	25	HR, SDNN, MNFr, MDFr, RespFreq	0.493
1	13	SDNN, Staticl	0.640
2	16	SDNN, Staticl, Temp	0.725
3	15	SDNN	0.364
4	14	SDNN	0.621
5	15	–	–
6	7	LFHF, SCL	0.736
7	11	–	–
8	17	–	–
9	7	OPD	0.875
10	11	HR	0.382

data. Results vary among subjects, but heart rate variability (SDNN and LFHF) is selected in five of the models.

The personal regression models generally show better results than the overall model for all subjects. Therefore we believe that a personal stress model is needed to predict stress in a reliable way. To conclude this section, we will show an example that illustrates the predictive power of the regression model. We selected the subject with the highest adjusted R^2 value from the regression of physiology on current stress level. This is subject 6 with an adjusted R^2 value of 0.481 (see Table 6.2). We selected the data points of measurement day 1–4 and ran a regression on these points. This resulted in a model with features *Staticr* and *Temp*. We used the coefficients of this model to predict the stress level corresponding to the data points of day 5. The self-reported stress levels and the predicted stress levels are shown in Figure 6.4. We found a correlation of 0.588 (p=0.125) for these data points.

6.4   Discussion

We have developed and conducted a novel study protocol for long-term ambulatory physiological monitoring for stress assessment. We collected a database of physiological signals, context parameters and reference measures for stress level. We recorded five different physiological modalities (ECG, EMG, respiration, skin conductance and skin temperature), five context parameters (ambient temperature, ambient relative humidity, activity level, posture, consumption) and two types of
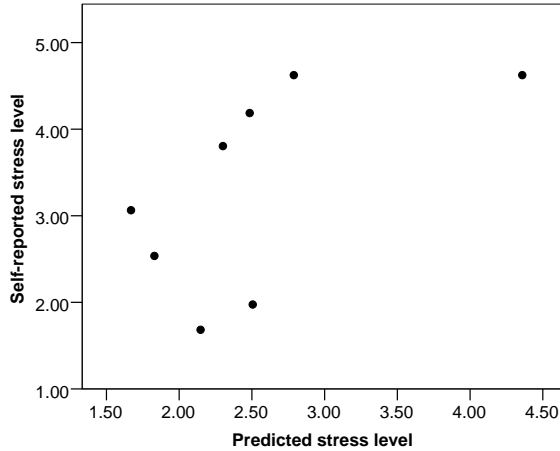
Figure 6.4: Self-reported versus predicted stress level for subject 6.

reference measures (questionnaires and cortisol). The data collection was done in a real-world environment (office environment). This ambulatory setting is a necessity for developing a stress prediction model that produces valuable results in real-world environments. The data were recorded for five working days, which is relatively long for this kind of study. Altogether, this database contains very valuable data for the research we did and for future research.

Our results show that stress can be modeled using physiological parameters. It is possible to develop a general model for stress detection, but the accuracy is relatively low. So it seems necessary to personalize the model. For most of the subjects, we found a significant individual model to predict stress. However, the models we found differed substantially among subjects regarding the selected feature set and the accuracy. This suggests that reaction to stress might be reflected in different physiological signals for different persons. Interestingly, Lacey et al. [89, 90] already found in the 1950s that the stress response differs among persons. This is also confirmed by the poor performance of the regression analyses we applied to all data points of all subjects together. Consequently, personal models are needed for reliable stress detection. To determine a personal model, a personal calibration would be needed before valid results can be produced.

The selected feature sets do not seem to form any pattern. Features from all five physiological signals were selected across subjects. For some subjects, there were features selected from just one signal. For other subjects, a feature set was selected that combined multiple physiological signals. At this point, no conclusions can be drawn on which features in general are best for predicting stress. The number of physiological signals we recorded was relatively large compared to the amount of data. The forward regression method prevented models with too many variables. More data could create more certainty about the composition of the personal feature sets.

Data points that could have been influenced by eating, drinking coffee, tea, energy drink or alcohol, smoking or physical activity were excluded from our analysis. This approach can be justified for data analysis purposes, but is not realistic in real-world environments because eating, drinking and smoking are difficult to detect. Therefore, as a next step in this research field it would be valuable to first find out if these context factors indeed influence the measurements, and in which way they influence the measurements. Once this is known, it can be determined how to deal with these influences. For example, the data from the accelerometers can be used to detect physical activity and to automatically exclude affected data parts.

We used self-reports and cortisol levels as reference measures. One could think of other measures including life events that are available through digital sources nowadays. Meetings and other events are logged in digital calendars, social media, emails, etc. This information can be used to obtain insights in the relation of these meetings and life events to physiological stress responses. See for example [82] for a study relating skin conductance levels to digital calendar information.

We showed an example of how a personal stress model could be determined and applied to new data of the same person. Although we recorded five days from every subject, we did not have enough data points to thoroughly verify this procedure. More longitudinal recordings from the same subjects would be needed to truly verify this procedure and to investigate if the personal stress models are consistent over longer times. Furthermore, if data is recorded from a larger population, stratification of people could be made. High responsive versus low responsive groups could be identified and characterized. Or different response patterns such as HR reactive versus SC reactive could be found.

Finally, it turned out that some subjects showed very little variation in stress level throughout the study. For future research, we suggest to select a subject group or certain conditions that guarantee significant fluctuations in experienced stress level. Subjects that have stressful jobs on a daily basis, such as air traffic controllers or high-level managers could fit this requirement. A stress prediction model will perform better if it was developed based on more extreme stress levels.

6.5 Conclusions

We developed and executed a study that takes a novel approach to estimating the impact of stress in real-world environments. We recorded various physiological parameters, reference measures and context information. We used wearable, wireless sensors during five working days in a real-world environment. Analysis of the data showed that stress can be predicted to a certain extent using regression models. Results also showed that personal stress models generally perform better than a general stress model. To answer our research question: the outcome of this pioneering study shows a good indication that modelling stress in real-world conditions using physiological features is possible, but we need personalized models.

7

Discussion

In Chapter 1, the general goal of this thesis was formulated as follows:

to assess the feasibility of constructing personal models for the relation between mental stress and physiological variables, for use in ambulatory stress management systems.

This goal can be subdivided in three aspects of which the feasibility should be assessed:

1. using physiological variables for stress detection

2. personalization of stress models

3. ambulatory measurements

The feasibility of using physiological variables for stress detection was shown in Chapter 2, Chapter 3 and Chapter 4. Chapter 2 reported on the research findings in literature on this topic. It was shown in many studies that there is a physiological reaction to stress that can be measured from certain physiological variables. Chapter 3 and Chapter 4 reported on a new experiment that verified and confirmed the findings from literature in various office-like situations. Also, in this experiment it was shown that physiological reactions to stress can be measured and can be used for classification of the results.

In Chapter 3, differences in EMG features between rest and stress conditions were found on group level without personalization of the analysis. However, large variability of feature values among subjects was also found. Therefore, personalization of stress models is necessary to improve accuracy of individual models. Personal normalization was done successfully in Chapter 4 for classification on individual subject level.

Ambulatory measurements are certainly feasible as was shown in the studies described in Chapter 5 and Chapter 6. In Chapter 5, the subjects were monitored

during medical simulation training. Although the environments were simulated, the training was real and the simulation resembled the real-world environment in a convincing way. The subjects who participated in the study described in Chapter 6 were continuously monitored for five days during daily office work. Although these studies show that ambulatory recording of physiological signals is feasible, the measurement system is not ready for long-term and commercial use yet. The sensors are too bulky and not comfortable. It is not desired to measure at many different locations on the body (shoulders, chest, wrist). Ideally an ambulatory recording system should consist of only one sensor device that is small, unobtrusive and can easily be applied to only one location on the body.

7.1 Conflicting Results and Unexpected Findings

In Chapter 3 it was shown that self-reported stress correlates to RMS EMG value and relative EMG gaptime. Unfortunately, the correlation values were not that high and no similar results were found in the other studies that were executed. Occasional correlations were found in the other studies between certain features and the self-reported stress values, but there seemed to be no clear pattern among these correlations.

There are two factors that may play a role in the lack of correspondence between self-reported stress levels and the measured physiological features. First, the self-reported values may not be accurate. When subjects are asked to indicate their stress level in a questionnaire, they need to interpret their stress feelings and indicate these subjective feelings on some kind of objective scale like a VAS or Likert scale. Obviously this interpretation is not straightforward and can cause inaccuracies. Second, the stress feelings in the mind are translated to measurable physiological signals through many intermediate steps such as nerves and hormonal reactions. Also, the sensation of the physiological stress reaction can, in turn, have impact on the mental stress experience. Because many factors play a role in the interplay between body and mind, all kinds of unpredictable effects may occur between the stress feelings in the mind and the peripheral physiological variables that are measured.

This inconsistency between self-reported stress and the peripheral physiological reaction contributes to one of the fundamental challenges in stress research: there is no recognized reference measure. Many researchers have already tried to measure stress, but nobody knows what the 'real' stress measure should be to which the models are tuned. Apart from various—validated or custom made—questionnaires, hormone levels are also commonly used as a reference for stress level. These currently available reference measures are the most reliable indicators of stress level, so these measures are what research in the area of stress detection is based on. Nevertheless, none of these measures is close enough to the actual mental stress feelings that only the subjects themselves can experience. Researchers working in this field therefore need to keep in mind that they are not working with a perfect

reference measure.

In Chapter 5 an increase in physiological stress was found while self-reported stress levels decreased over the same time frame. So besides the fact that often no correlation is found between the self-reported stress and physiological signals, here the two measures even moved in opposite directions. In this case, the cause of the inconsistency was most probably the physical activity that was involved with the protocol of Study 1. During the evaluation period at the end of the protocol, the influence of the physical activity was still present in the physiological signals while the mental stress was reduced compared to the baseline value. While this is an example which demanded relatively large physical effort, the phenomenon described here still highlights the importance of controlling physical activity and other context factors while performing physiological stress measurements. Many context factors can influence the physiological measurements and it is important to either discard the data that may have been influenced by context factors or to compensate for the influence of context factors before analyzing the physiological data for stress level.

The analysis of the data collected during the last study, described in Chapter 6, was performed on personal level for every subject. It was unexpected to see such substantial differences in the selection of the subset of features for the regression models. More data would be needed to verify if these differences are consistent, i.e. if one person always shows the same response pattern. Also it could be investigated if certain subgroups could be distinguished that show similar physiological response patterns. Some personal characteristics that were already found to cause different physiological responses are gender [5, 13, 15, 102], age [26, 30, 163], race [101, 150], aerobic fitness [53] and certain physical and psychological conditions such as posttraumatic stress disorder [12], panic disorder [12, 66], anxiety disorder [67], autism [56], self-mutilation [126] and obesity [154]. However, other studies found contradicting results. For example, Blangsted et al. and Krantz et al. did not find different responses for subgroups based on gender [11, 85]. Therefore more research would be needed to unravel the specific physiological mechanisms that comprise the different stress reactions. Still, some of these characteristics—such as gender, age and aerobic fitness—might have played a role in the subject group that was used in the study described in Chapter 6 and are likely to have partially caused the differences in regression models. Furthermore, the study took place in an ambulatory real-world environment in which different types of stressors could have occurred. Not only the personal characteristics, but also different types of stressors can cause differences in physiological response [5].

7.2   Physiological Features for Stress Detection

There is a large number of studies that investigated the physiological reaction to mental stress. A selection of these studies was described in Chapter 2. This literature review identified five features that correlate to mental stress: heart rate,

systolic and diastolic blood pressure, skin conductance level and number of skin conductance responses.

Blood pressure was not measured during the studies described in this thesis because no wearable continuous blood pressure sensor was available for this purpose. Heart rate and skin conductance level turned out to be relevant features for stress detection, as was described in Chapter 4. The second difference of the skin conductance signal was selected here instead of number of skin conductance responses. The information that these two features give is somewhat comparable as they both give information about the fluctuations in the skin conductance signal.

Another feature that was selected in the study in Chapter 4 was respiration frequency. In the literature review this feature was found to correlate with stress often, but some contradictory results were also found. Although respiration frequency was selected in this study, one should be careful when applying it in real-world situations as it is influenced by speaking.

The literature study showed that most of the studies in which HRV features were evaluated found changes of these features in the expected direction. In the PCA analysis in the study described in Chapter 4 SDNN contributed substantially to the second component, which indicates that there was relevant information in this feature. There was no heart rate variability feature selected in the GEE analysis. Apparently inclusion of the SDNN feature did not add any new information to the already selected features for this analysis.

In Chapter 5, only ECG and skin conductance were measured. In the resuscitation study both signals change between the different conditions. It was interesting to see that during the evaluation phase after the actual simulation, skin conductance level remained elevated for a long time whereas heart rate returned to baseline quicker.

The results of Chapter 6 show many differences among the feature sets that were selected for the various subjects. Some features are commonly used in stress research, but other features are not that common. Features that are not common, such as the EMG features and skin temperature, are the key features for the regression model for some subjects but do not play a role in the models of other subjects. These individual differences can explain the contradictory results that were reported in literature for these features. Heterogeneity of the subject group and the personal physiological stress reaction profiles of the subjects could have caused contradictions in the results. Hence, individual models are needed for accurate stress level estimation.

7.3 Limitations

Baseline physiological values recorded in laboratory settings are generally higher than values measured in situations in which the subject is more relaxed [67]. Furthermore, laboratory stressors tend to induce weaker responses than real-life stress

situations [67]. Therefore the protocols of the studies described in Chapter 3, Chapter 4 and Chapter 5 did probably not induce stress levels similar to real-world situations. The study described in Chapter 6 did take place in real-world situations. However, some subjects did not experience much fluctuation in stress levels during the measurement days. If not much fluctuation of stress can be measured, the stress reaction cannot be modeled accurately either.

In the study described in Chapter 6 we sampled saliva for analysis of cortisol levels. The samples were taken four times a day for five days, resulting in twenty samples per subject. More frequent sampling was not feasible because of costs and because subjects could not eat or drink anything 30 minutes before the sampling. Frequent sampling would mean that subjects hardly would have had time to eat or drink during the day. However, more frequent sampling could have made the regression analysis stronger and could have led to a more robust stress model.

The regression analysis that was performed in Chapter 6 assumed linearity between the physiological features and the reference variables. The actual relationship between these variables is still unknown though. The translation from a stress sensation in the mind to the measurable physiological variables is a complex process that probably does not exhibit linear behavior.

Five measurement days is relatively short when trying to assess chronic stress levels. Therefore the last study focused on shorter episodes of stress. If multiple short stress episodes lead to chronic elevation of stress this would be a valid approach. However, to test this hypothesis longer measurement periods and/or follow-up studies would be necessary.

The subject groups that were chosen for the studies described in this thesis were rather specific. The studies described in Chapter 3, Chapter 4 and Chapter 6 included researchers in the subject groups. The studies described in Chapter 5 included medical students and high school students in the subject groups. The results that were found in these studies can therefore not be generalized to the whole population. For such a generalization one would need a number of subjects for subgroups with similar physiological response patterns, as indicated in Section 7.1. Examples of subgroups are men vs. women, age groups, healthy subjects vs. subjects with certain medical conditions and groups based on race. More studies with diverse subject groups would be necessary to facilitate generalization to the whole population.

7.4 Conclusions

The main conclusion of this thesis is that it is feasible to use physiological variables for personalized stress detection in ambulatory conditions. This conclusion implies the potential for development of commercial stress management systems for applications in personal life. Such a system could be used in a work environment, during psychological treatment, in car driving etc.

Only few studies have developed a method to estimate stress level on a continuous scale and continuous in time. Most studies just investigate differences between baseline and stress conditions using statistical tests or they use classification algorithms to classify the two conditions. In the studies described in this thesis two ways for continuous stress estimation were suggested. First, the outcome of the Generalized Estimating Equations (GEE) analysis could be used for continuous stress level estimation using the probability of stress and a sliding window. Second, regression analysis can be used to model self-reported stress levels indicated on a VAS to physiological features. This way stress levels can be estimated on a continuous scale.

To the best of our knowledge, the real-world study that was done is the first study in an ambulatory setting in which this large number of physiological, context and reference variables were recorded. Five different physiological signals, eight context variables and three reference measures were collected. These variables were collected from ten subjects during five days, in an ambulatory setting during working hours.

The final study was set up in such a way that individual analysis of the data was possible. The subject group was relatively small but the amount of data collected per person was large. This setup facilitates the development of a personalized model for every subject. It was shown that these personalized models differ significantly among subjects and that these models perform better than a general model that is fit to the data of all subjects together. So a general model could be used as a starting point in a feature stress management system. The model should be individualized to the user as soon as possible though.

7.5 Future Research

This thesis presents studies towards real-world stress detection. The studies raised a number of directions for future work, which will be described here.

From the literature review in Chapter 2 it turned out that blood pressure measurement could have great potential for stress monitoring. An extension of the recorded signals in order to include blood pressure would be desirable to improve stress estimation models. Unfortunately, no wearable continuous blood pressure sensors are available for this purpose. Ambulatory blood pressure estimation based on pulse arrival time seems promising, but validation of this technique is still on-going (e.g. in Samsung's Simband).

Certain context factors influence physiological signals and could therefore impede the stress detection. More investigation of this influence could improve the stress estimation models. In the last study, for example, information about physical activity, posture, consumption of food, drinks and cigarettes, temperature and relative humidity were measured. These could all have had an influence on the results. Analyzing these factors and the physiological signals might bring understanding of

the exact influence of these factors on the physiology. When this understanding is achieved, it can be decided if data influenced by these factors should be excluded from analysis or if the estimation of stress level can be compensated for these factors.

The timing of the responses to stressors is different for various physiological mechanisms. For example, skin conductance latency is much higher than EMG latency. This phenomenon applies both to the response time and the recovery time. Thorough insight in the timing patterns of the various mechanisms can help to improve the accuracy of stress level estimation.

Also a more thorough investigation into individual stress response patterns could bring more understanding of the origin of the individual differences. As was already pointed out in Section 7.1, there are many population subgroups that can be distinguished based on personal characteristics. If different response patterns could be identified for these subgroups, this would help in defining accurate and personalized stress estimation algorithms.

An important topic in stress research is the method for individual calibration of the stress estimation algorithm. The most straightforward method is to invite users to a controlled (laboratory) environment and have them undergo a standardized set of stressors. The physiological response to these stressors can then be used for calibration of the algorithm to improve the accuracy of stress estimation. Another option is to develop a self-learning stress measurement system. Such a system would initially estimate stress using a general algorithm. During use, the system would prompt the user from time to time and inquire about his or her stress level. Based on this user input, the algorithm can be personalized and will become more accurate.

Two final topics that are pointed out here for future research are real-time stress level calculation and effective feedback. As was already indicated in the Introduction (Chapter 1), it would be useful to develop a personal stress management system that could measure stress and give effective feedback. To fulfill this goal it is necessary to develop algorithms that can calculate stress level from physiological signals in real time. At the same time, the outcome of such algorithms should be converted into understandable feedback to the user. The preferred ways to give effective feedback that creates awareness of stress level in a pleasant way need to be investigated.

Bibliography

[1] H. Abercrombie, N. Kalin, M. Thurow, M. Rosenkranz, and R. Davidson. Cortisol variation in humans affects memory for emotionally laden and neutral information. *Behavioral neuroscience*, 117(3):505–516, 2003.

[2] M. Al-Fudail and H. Mellar. Investigating teacher stress when using technology. *Computers and Education*, 51(3):1103–1110, 2008.

[3] F. Alamudun, J. Choi, R. Gutierrez-Osuna, H. Khan, and B. Ahmed. Removal of subject-dependent and activity-dependent variation in physiological measures of stress. In *6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pages 115–122, 2012.

[4] K. Allen and J. Blascovich. Effects of music on cardiovascular reactivity among surgeons. *Journal Of the American Medical Association*, 272:882–884, 1994.

[5] M. Allen, C. Stoney, J. Owens, and K. Matthews. Hemodynamic adjustments to laboratory stress: the influence of gender and personality. *Psychosomatic Medicine*, 55(6):505–517, 1993.

[6] P. Andreatta, M. Hillard, and L. Krain. The impact of stress factors in simulation-based laparoscopic training. *Surgery*, 147(5):631–639, 2010.

[7] S. Arora, N. Sevdalis, D. Nestel, M. Woloshynowych, A. Darzi, and R. Kneebone. The impact of stress on surgical performance: a systematic review of the literature. *Surgery*, 147(3):318–30, 330.e1–6, 2010.

[8] H. Baek, H. Lee, J. Kim, J. Choi, K. Kim, and K. Park. Nonintrusive biological signal monitoring in a car to evaluate a driver's stress and health state. *Telemedicine and e-Health*, 15(2):182–189, 2009.

[9] K. Beischel. Variables affecting learning in a simulation experience: a mixed methods study. *Western journal of nursing research*, 35(2):226–247, 2013.

[10] L. Bernardi, J. Wdowczyk-Szulc, C. Valenti, S. Castoldi, C. Passino, G. Spadacini, and P. Sleight. Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. *Journal of the American College of Cardiology*, 35(6):1462–1469, 2000.

[11] A. K. Blangsted, K. Hansen, and C. Jensen. Muscle activity during computer-based office work in relation to self-reported job demands and gender. *European Journal of Applied Physiology*, 89(3-4):352–358, 2003.

[12] J. Blechert, T. Michael, P. Grossman, M. Lajtman, and F. H. Wilhelm. Autonomic and respiratory characteristics of posttraumatic stress disorder and panic disorder. *Psychosomatic Medicine*, 69(9):935–943, 2007.

[13] J. Blumenthal, E. T. Thyrum, and W. Siegel. Contribution of job strain, job status and marital status to laboratory and ambulatory blood pressure in patients with mild hypertension. *Journal of Psychosomatic Research*, 39(2):133–144, 1995.

[14] C. Bong, J. Lightdale, M. Fredette, and P. Weinstock. Effects of simulation versus traditional tutorial-based training on physiologic stress levels among clinicians: a pilot study. *Simulation in Healthcare*, 5(5):272–278, 2010.

[15] E. Bouma, H. Riese, J. Ormel, F. Verhulst, and A. Oldehinkel. Adolescents' cortisol responses to awakening and social stress; effects of gender, menstrual phase and oral contraceptives. the TRAILS study. *Psychoneuroendocrinology*, 34(6):884–893, 2009.

[16] P. J. Brantley, C. D. Waggoner, G. N. Jones, and N. B. Rappaport. A daily stress inventory: Development, reliability, and validity. *Journal of Behavioral Medicine*, 10(1):61–74, 1987.

[17] A. Breier, M. Albus, D. Pickar, T. Zahn, O. Wolkowitz, and S. Paul. Controllable and uncontrollable stress in humans: Alterations in mood and neuroendocrine and psychophysiological function. *American Journal of Psychiatry*, 144(11):1419–1425, 1987.

[18] J. Brockmyer, C. Fox, K. Curtiss, E. McBroom, K. Burkhart, and J. Pidruzny. The development of the game engagement questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology*, 45(4):624–634, 2009.

[19] E. L. van den Broek, J. H. Janssen, J. H. Westerink, and J. A. Healey. Prerequisites for affective signal processing (ASP). In P. Encarnacao and A. Veloso, editors, *International Conference on Bio-Inspired Systems and Signal Processing, Biosignals 2009*, pages 426–433, Portugal, January 2009. INSTICC Press.

[20] L. Brown, B. Grundlehner, J. van de Molengraft, J. Penders, and B. Gyselinckx. Body area network for monitoring autonomic nervous system responses. In *Pervasive Computing Technologies for Healthcare*, 2009.

[21] T. Buchanan and W. Lovallo. Enhanced memory for emotional material following stress-level cortisol treatment in humans. *Psychoneuroendocrinology*, 26(3):307–317, 2001.

[22] T. Buchanan, D. Tranel, and R. Adolphs. Impaired memory retrieval correlates with individual differences in cortisol response but not autonomic response. *Learning & memory*, 13(3):382–387, 2006.

[23] R. Burr. Interpretation of normalized spectral heart rate variability indices in sleep research: a critical review. *Sleep*, 30(7):913–919, 2007.

[24] J. Cacioppo, M. Burleson, K. Poehlmann, W. Malarkey, J. Kiecolt-Glaser, G. Berntson, B. Uchino, and R. Glaser. Autonomic and neuroendocrine responses to mild psychological stressors: Effects of chronic stress on older women. *Annals of Behavioral Medicine*, 22:140–148, 2000.

[25] D. Carroll, A. Phillips, G. Der, K. Hunt, and M. Benzeval. Blood pressure reactions to acute mental stress and future blood pressure status: Data from the 12-year follow-up of the West of Scotland Study. *Psychosomatic Medicine*, 73(9):737–742, 2011.

[26] D. Carroll, C. Ring, K. Hunt, G. Ford, and S. Macintyre. Blood pressure reactions to stress and the prediction of future blood pressure: effects of sex, age and socioeconomic position. *Psychosomatic Medicine*, 65(6):1058–1064, 2003.

[27] T. Chandola, A. Britton, E. Brunner, H. Hemingway, M. Malik, M. Kumari, E. Badrick, M. Kivimaki, and M. Marmot. Work stress and coronary heart disease: What are the mechanisms? *European Heart Journal*, 29(5):640–648, 2008.

[28] Y. Chida and M. Hamer. Chronic psychosocial factors and acute physiological responses to laboratory-induced stress in healthy populations: A quantitative review of 30 years of investigations. *Psychological Bulletin*, 134(6):829–885, 2008.

[29] J. Choi, B. Ahmed, and R. Gutierrez-Osuna. Development and evaluation of an ambulatory stress monitor based on wearable sensors. *IEEE Transactions on Information Technology in Biomedicine*, 16(2):279–286, 2012.

[30] E. Clays, F. Leynen, D. De Bacquer, M. Kornitzer, F. Kittel, R. Karasek, and G. De Backer. High job strain and ambulatory blood pressure in middle-aged men and women from the Belgian job stress study. *Journal of Occupational and Environmental Medicine*, 49(4):360–367, 2007.

[31] S. Cohen, T. Kamarck, and R. Mermelstein. A global measure of perceived stress. *Journal of Health and Social Behavior*, 24:385–396, 1983.

[32] C. Collet, C. Petit, A. Priez, and A. Dittmar. Stroop color-word test, arousal, electrodermal activity and performance in a critical driving situation. *Biological Psychology*, 69(2):195–203, 2005.

[33] T. Cox, A. Griffiths, and E. Rial-González. Research on work-related stress. Technical report, European Agency for Safety and Health at Work, 2000.

[34] S. Demaria, Jr, E. O. Bryson, T. J. Mooney, J. H. Silverstein, D. L. Reich, C. Bodian, and A. I. Levine. Adding emotional stressors to training in simulated cardiopulmonary arrest enhances participant performance. *Medical Education*, 44(10):1006–1015, 2010.

[35] S. Dickerson and M. Kemeny. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological bulletin*, 130(3):355–391, 2004.

[36] G. Dikecligil and L. Mujica-Parodi. Ambulatory and challenge-associated heart rate variability measures predict cardiac responses to real-world acute emotional stress. *Biological Psychiatry*, 67(12):1185–1190, 2010.

[37] R. Dishman, Y. Nakamura, M. Garcia, R. Thompson, A. Dunn, and S. Blair. Heart rate variability, trait anxiety, and perceived stress among physically fit men and women. *International Journal of Psychophysiology*, 37(2):121–133, 2000.

[38] G. Domes, M. Heinrichs, U. Rimmele, U. Reichwald, and M. Hautzinger. Acute stress impairs recognition for positive words—association with stress-induced cortisol secretion. *Stress*, 7(3):173–181, 2004.

[39] M. van Dooren, J. de Vries, and J. H. Janssen. Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior*, 106(2):298–304, 2012.

[40] L. van Doornen and J. Houtveen. Fysiologische stressmetingen in de werksituatie: nut, beperkingen en bevindingen. *Gedrag & Organisatie*, 22(3):275–293, 2009. Reeks 'De Werkplaats'.

[41] L. Dorn, J. Lucke, T. Loucks, and S. Berga. Salivary cortisol reflects serum cortisol: analysis of circadian profiles. *Annals of Clinical Biochemistry*, 44(Pt 3):281–284, 2007.

[42] Elsevier B.V. Scopus. http://www.scopus.com/.

[43] B. Elzinga and K. Roelofs. Cortisol-induced impairments of working memory require acute sympathetic activation. *Behavioral neuroscience*, 119(1):98–103, 2005.

[44] V. Engert, S. Vogel, S. I. Efanov, A. Duchesne, V. Corbo, N. Ali, and J. C. Pruessner. Investigation into the cross-correlation of salivary cortisol and alpha-amylase responses to psychological stress. *Psychoneuroendocrinology*, 36(9):1294–1302, 2011.

[45] T. Esch, G. Stefano, G. Fricchione, and H. Benson. Stress in cardiovascular diseases. *Medical Science Monitor*, 8(5):RA93–RA101, 2002.

[46] European Agency for Safety and Health at Work – EU-OHSA. Healthy workplaces manage stress. https://www.healthy-workplaces.eu/en.

[47] European Agency for Safety and Health at Work – EU-OHSA. European opinion poll on occupational safety and health. Technical report, May 2013.

[48] European Commission. *Health and safety at work in Europe—A statistical portrait*. Publications Office of the European Union, Luxembourg, 2010.

[49] W. Fenz and S. Epstein. Gradients of physiological arousal in parachutists as a function of an approaching jump. *Psychosomatic Medicine*, 29(1):33–51, 1967.

[50] B. Flanagan, D. Nestel, and M. Joseph. Making patient safety the focus: crisis resource management in the undergraduate curriculum. *Medical Education*, 38(1):56–66, 2004.

[51] P. Foley and C. Kirschbaum. Human hypothalamus-pituitary-adrenal axis responses to acute psychosocial stress in laboratory settings. *Neuroscience and Biobehavioral Reviews*, 35(1):91–96, 2010.

[52] S. Folkman, R. Lazarus, C. Dunkel-Schetter, A. DeLongis, and R. Gruen. Dynamics of a stressful encounter: cognitive appraisal, coping, and encounter outcomes. *Journal of personality and social psychology*, 50(5):992–1003, 1986.

[53] E. de Geus, L. van Doornen, and J. Orlebeke. Regular exercise and aerobic fitness in relation to psychological make-up and physiological stress reactivity. *Psychosomatic Medicine*, 55(4):347–363, 1993.

[54] K. Gildea, T. Schneider, and W. Shebilske. Stress appraisals and training performance on a complex laboratory task. *Human factors*, 49(4):745–758, 2007.

[55] D. Girzadas, S. Delis, S. Bose, J. Hall, K. Rzechula, and E. Kulstad. Measures of stress and learning seem to be equally affected among all roles in a simulation scenario. *Simulation in Healthcare*, 4(3):149–54, 2009.

[56] M. S. Goodwin, J. Groden, W. F. Velicer, L. P. Lipsitt, M. G. Baron, S. G. Hofmann, and G. Groden. Cardiovascular arousal in individuals with autism. *Focus on Autism and Other Developmental Disabilities*, 21(2):100–123, 2006.

[57] K. Griffiths, M. Mackey, and B. Adamson. The impact of a computerized work environment on professional occupational groups and behavioural and physiological risk factors for musculoskeletal symptoms: A literature review. *Journal of Occupational Rehabilitation*, 17:743–765, 2007.

[58] B. Grundlehner, L. Brown, J. Penders, and B. Gyselinckx. The design and analysis of a real-time, continuous arousal monitor. In *Sixth International*

Workshop on Wearable and Implantable Body Sensor Networks, pages 156–161, 2009.

[59] M. Hall, R. Vasko, D. Buysse, H. Ombao, Q. Chen, J. Cashmere, D. Kupfer, and J. Thayer. Acute stress affects heart rate variability during sleep. *Psychosomatic Medicine*, 66(1):56–62, 2004.

[60] A. Harvey, A. Nathens, G. Bandiera, and V. Leblanc. Threat and challenge: cognitive appraisal and stress responses in simulated trauma resuscitations. *Medical education*, 44(6):587–594, 2010.

[61] J. Healey, L. Nachman, S. Subramanian, J. Shahabdeen, and M. Morris. Out of the lab and into the fray: Towards modeling emotion in everyday life. In *Pervasive Computing*, volume 6030 of *Lecture Notes in Computer Science*, pages 156–173. Springer Berlin Heidelberg, 2010.

[62] J. Healey and R. Picard. Smartcar: detecting driver stress. In *15th International Pattern Recognition Conference*, volume 4, pages 218–221, 2000.

[63] J. Healey and R. Picard. Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, 6(2):156–166, 2005.

[64] N. Hjortskov, A. H. Garde, P. Ørbæk, and Å. M. Hansen. Evaluation of salivary cortisol as a biomarker of self-reported mental stress in field studies. *Stress and Health*, 20(2):91–98, 2004.

[65] N. Hjortskov, D. Rissén, A. Blangsted, N. Fallentin, U. Lundberg, and K. Søgaard. The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology*, 92(1):84–89, 2004.

[66] T. Hoehn, S. Braune, G. Scheibe, and M. Albus. Physiological, biochemical and subjective parameters in anxiety patients with panic disorder during stress exposure as compared with healthy controls. *European Archives of Psychiatry and Clinical Neuroscience*, 247(5):264–274, 1997.

[67] R. Hoehn-Saric, D. McLeod, and W. Zimmerli. Somatic manifestations in women with generalized anxiety disorder. psychophysiological responses to psychological stress. *Archives of General Psychiatry*, 46(12):1113–1119, 1989.

[68] Institute of Work, Health & Organisations. Towards the development of a European framework for psychosocial risk management at the workplace, 2008.

[69] E. Jackson, J. Payne, L. Nadel, and W. Jacobs. Stress differentially modulates fear conditioning in healthy men and women. *Biological Psychiatry*, 59(6):516–522, 2006.

[70] C. Jern, E. Eriksson, L. Tengborn, B. Riberg, H. Wadenvik, and S. Jern. Changes of plasma coagulation and fibrinolysis in respone to mental stress. *Thrombosis and Haemostasis*, 62(2):767–771, 1989.

[71] R.-P. Juster, B. S. McEwen, and S. J. Lupien. Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience and Biobehavioral Reviews*, 35(1):2–16, 2010.

[72] E. Kajantie and D. Phillips. The effects of sex and hormonal status on the physiological response to acute psychosocial stress. *Psychoneuroendocrinology*, 31(2):151–178, 2006.

[73] T. W. Kamarck and W. R. Lovallo. Cardiovascular reactivity to psychological challenge: Conceptual and measurement considerations. *Psychosomatic Medicine*, 65(1):9–21, 2003.

[74] K. Karlsson, P. Niemelä, and A. Jonsson. Heart rate as a marker of stress in ambulance personnel: a pilot study of the body's response to the ambulance alarm. *Prehospital and disaster medicine*, 26(1):21–26, 2011.

[75] A. Keitel, M. Ringleb, I. Schwartges, U. Weik, O. Picker, U. Stockhorst, and R. Deinzer. Endocrine and psychological stress responses in a simulated emergency situation. *Psychoneuroendocrinology*, 36(1):98–108, 2011.

[76] M. Kharasch, P. Aitchison, C. Pettineo, L. Pettineo, and E. Wang. Physiological stress responses of emergency medicine residents during an immersive medical simulation scenario. *Disease-a-month: DM*, 57(11):700–705, 2011.

[77] C. Kirschbaum, B. Kudielka, J. Gaab, N. Schommer, and D. Hellhammer. Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic medicine*, 61(2):154–162, 1999.

[78] C. Kirschbaum, K. Pirke, and D. Hellhammer. Preliminary evidence for reduced cortisol responsivity to psychological stress in women using oral contraceptive medication. *Psychoneuroendocrinology*, 20(5):509–514, 1995.

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

[80] C. Kirschbaum, S. Wüst, and D. Hellhammer. Consistent sex differences in cortisol responses to psychological stress. *Psychosomatic Medicine*, 54(6):648–657, 1992.

[81] S. Kjeldsen, K. Knudsen, G. Ekrem, T. Fure, P. Movinckel, and J. Erikssen. Is there an association between severe job strain, transient rise in blood pressure and increased mortality? *Blood Pressure*, 15(2):93–100, 2006.

[82] R. Kocielnik, N. Sidorova, F. Maggi, M. Ouwerkerk, and J. Westerink. Smart technologies for long-term stress monitoring at work. In *IEEE 26th Inter-*

national Symposium on Computer-Based Medical Systems (CBMS), pages 53–58, 2013.

[83] M. J. Koehler, A. Yadav, M. Phillips, and S. Cavazos-Kottke. What is video good for? Examining how media and story genre interact. *Journal of Educational Multimedia and Hypermedia*, 14(3):249–272, 2005.

[84] F. Kok, H. Westenberg, J. Thijssen, and J. van Ree. Endocrine and cardiovascular responses to a series of graded physical and psychological stress stimuli in healthy volunteers. *European Neuropsychopharmacology*, 5(4):515–522, 1995.

[85] G. Krantz, M. Forsman, and U. Lundberg. Consistency in physiological stress responses and electromyographic activity during induced stress exposure in women and men. *Integrative Psychological and Behavioral Science*, 39(2):105–118, 2004.

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

[87] B. Kudielka and C. Kirschbaum. Sex differences in HPA axis responses to stress: a review. *Biological psychology*, 69(1):113–132, 2005.

[88] S. Kuhlmann, M. Piel, and O. Wolf. Impaired memory retrieval after psychosocial stress in healthy young men. *Journal of Neuroscience*, 25(11):2977–2982, 2005.

[89] J. I. Lacey, D. E. Bateman, and R. VanLehn. Autonomic response specificity: An experimental study. *Psychosomatic Medicine*, 15(1):8–21, 1953.

[90] J. I. Lacey and B. C. Lacey. Verification and extension of the principle of autonomic response-stereotypy. *The American Journal of Psychology*, 71(1):50–73, 1958.

[91] P. Landsbergis, P. Schnall, T. Pickering, K. Warren, and J. Schwartz. Life-course exposure to job strain and ambulatory blood pressure in men. *American Journal of Epidemiology*, 157(11):998–1006, 2003.

[92] S.-E. Larsson, R. Larsson, Q. Zhang, H. Cai, and P. Å. Öberg. Effects of psychophysiological stress on trapezius muscles blood flow and electromyography during static load. *European Journal of Applied Physiology and Occupational Physiology*, 71(6):493–498, 1995.

[93] B. Laursen, B. Jensen, A. Garde, and A. Jørgensen. Effect of mental and physical demands on muscular activity during the use of a computer mouse and a keyboard. *Scandinavian Journal of Work, Environment & Health*, 28:215–221, 2002.

[94] R. Lazarus and S. Folkman. *Stress, Appraisal and Coping.* Springer Publishing Company, 1984.

[95] V. LeBlanc. The effects of acute stress on performance: implications for health professions education. *Academic medicine: journal of the Association of American Medical Colleges*, 84(10 Suppl):S25–33, 2009.

[96] V. LeBlanc and G. Bandiera. The effects of examination stress on the performance of emergency medicine residents. *Medical education*, 41(6):556–564, 2007.

[97] V. LeBlanc, R. MacDonald, B. McArthur, K. King, and T. Lepine. Paramedic performance in calculating drug dosages following stressful scenarios in a human patient simulator. *Prehospital emergency care: official journal of the National Association of EMS Physicians and the National Association of State EMS Directors*, 9(4):439–444, 2005.

[98] V. Leblanc, C. Regehr, W. Tavares, A. Scott, R. Macdonald, and K. King. The impact of stress on paramedic performance during simulated critical events. *Prehospital and Disaster Medicine*, 27(4):369–374, 2012.

[99] S. Leka and A. Jain. Health impact of psychosocial hazards at work: An overview. Technical report, World Health Organization, 2010.

[100] R. Levenson. Alcohol and stress response dampening: Pharmacological effects, expectancy, and tension reduction. *Journal of Abnormal Psychology*, 89(4):528–538, 1980.

[101] K. Light, P. Obrist, A. Sherwood, S. James, and D. Strogatz. Effects of race and marginally elevated blood pressure on responses to stress. *Hypertension*, 10(6):555–563, 1987.

[102] K. Light, J. Turner, and A. Hinderliter. Job strain and ambulatory work blood pressure in healthy young men and women. *Hypertension*, 20(2):214–218, 1992.

[103] M. Loeb, D. Holding, and M. Baker. Noise stress and circadian arousal in self-paced computation. *Motivation and Emotion*, 6:43–48, 1982.

[104] D. Lucini, G. Norbiato, M. Clerici, and M. Pagani. Hemodynamic and autonomic adjustments to real life stress conditions in humans. *Hypertension*, 39(1):184–188, 2002.

[105] U. Lundberg, R. Kadefors, B. Melin, G. Palmerud, P. Hassmen, M. Engstrom, and I. Dohns. Psychophysiological stress and EMG activity of the trapezius muscle. *International Journal of Behavioral Medicine*, 1:354–370, 1994.

[106] S. Lupien, C. Gillin, and R. Hauger. Working memory is more sensitive than declarative memory to the acute effects of corticosteroids: a dose-response study in humans. *Behavioral neuroscience*, 113(3):420–430, 1999.

[107] E. Marieb. The autonomic nervous system. In *Human anatomy and physiology*, pages 531–553. Pearson Benjamin Cummings, San Fransisco, 6th edition, 2004.

[108] T. Marteau and H. Bekker. The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, 31(Pt 3):301–306, 1992.

[109] F. Martini. *Fundamentals of Anatomy and Physiology*. Prentice Hall, 4th edition, 1998.

[110] K. Matthews, S. Zhu, D. Tucker, and M. Whooley. Blood pressure reactivity to psychological stress and coronary calcification in the coronary artery risk development in young adults study. *Hypertension*, 47(3):391–395, 2006.

[111] B. McEwen. Protective and damaging effects of stress mediators. *The New England Journal of Medicine*, 338(3):171–179, 1998.

[112] B. S. McEwen. The neurobiology of stress: from serendipity to clinical relevance. *Brain Research*, 886(1–2):172–189, 2000.

[113] L. McGraw, D. Out, J. Hammermeister, C. Ohlson, M. Pickering, and D. Granger. Nature, correlates, and consequences of stress-related biological reactivity and regulation in army nurses during combat casualty simulation. *Psychoneuroendocrinology*, 38(1):135–144, 2013.

[114] K. A. McKay, J. E. Buen, K. J. Bohan, and J. P. Maye. Determining the relationship of acute stress, anxiety, and salivary alpha-amylase level with performance of student nurse anesthetists during human-based anesthesia simulator training. *AANA Journal*, 78(4):301–309, 2010.

[115] B. Mehler, B. Reimer, J. Coughlin, and J. Dusek. Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record*, 2138:6–12, 2009.

[116] R. Merletti and P. Parker. *Electromyography—Physiology, Engineering and Noninvasive Applications*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.

[117] D. B. Miller and J. P. O'Callaghan. Neuroendocrine aspects of the response to stress. *Metabolism*, 51(6, Part 2):5–10, 2002.

[118] M. Muller, M. Hansel, A. Fichtner, F. Hardt, S. Weber, C. Kirschbaum, S. Ruder, F. Walcher, T. Koch, and C. Eich. Excellence in performance and stress reduction during two different full scale simulator training courses: a pilot study. *Resuscitation*, 80(8):919–924, 2009.

[119] M. Myrtek, E. Aschenbrenner, and G. Brügner. Emotions in everyday life: an ambulatory monitoring study with female students. *Biological Psychology*, 68(3):237–255, 2005.

[120] M. Myrtek and G. Brügner. Perception of emotions in everyday life: studies with patients and normals. *Biological Psychology*, 42(1–2):147–164, 1996.

[121] U. Nater, N. Rohleder, J. Gaab, S. Berger, A. Jud, C. Kirschbaum, and U. Ehlert. Human salivary alpha-amylase reactivity in a psychosocial stress paradigm. *International Journal of Psychophysiology*, 55(3):333–342, 2005.

[122] National Institute for Occupational Safety and Health. Stress...at work. U.S. Department of Health and Human Services; Public Health Service; Centers for Disease Control and Prevention, 1999.

[123] P. Nepomnaschy, T. Lee, L. Zeng, and C. Dean. Who is stressed? Comparing cortisol levels between individuals. *American Journal of Human Biology*, 24(4):515–525, 2012.

[124] K. Nilsen, T. Sand, L. Stovner, R. Leistad, and R. Westgaard. Autonomic and muscular responses and recovery to one-hour laboratory mental stress in healthy subjects. *BMC Musculoskeletal Disorders*, 8:81, 2007.

[125] R. E. Noble. Diagnosis of stress. *Metabolism*, 51(6, Part B):37–39, 2002.

[126] M. K. Nock and W. B. Mendes. Physiological arousal, distress tolerance, and social problem-solving deficits among adolescent self-injurers. *Journal of Consulting and Clinical Psychology*, 76(1):28–38, 2008.

[127] G. Oliver, J. Wardle, and E. Gibson. Stress and food choice: A laboratory study. *Psychosomatic Medicine*, 62(6):853–865, 2000.

[128] K. Pacák and M. Palkovits. Stressor specificity of central neuroendocrine responses: Implications for stress-related disorders. *Endocrine Reviews*, 22(4):502–548, 2001.

[129] J. Penders, M. Altini, J. van de Molengraft, F. Yazicioglu, and C. Van Hoof. A low-power wireless ECG necklace for reliable cardiac activity monitoring on-the-move. In *International IEEE EMBS Conference*, 2011.

[130] M. L. Peters, G. L. Godaert, R. E. Ballieux, M. van Vliet, J. J. Willemsen, F. C. Sweep, and C. J. Heijnen. Cardiovascular and endocrine responses to experimental stress: Effects of mental effort and controllability. *Psychoneuroendocrinology*, 23(1):1–17, 1998.

[131] K. Plarre, A. Raij, S. Hossain, A. Ali, M. Nakajima, M. al'Absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smailagic, and L. Wittmers. Continuous inference of psychological stress from sensory measurements collected in the natural environment. In *10th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 97–108, 2011.

[132] T. Reinhardt, C. Schmahl, S. Wüst, and M. Bohus. Salivary cortisol, heart rate, electrodermal activity and subjective stress responses to the Mannheim Multicomponent Stress Test (MMST). *Psychiatry Research*, 198(1):106–111, 2012.

[133] P. Renaud and J.-P. Blondin. The stress of Stroop performance: Physiological and emotional responses to color-word interference, task pacing, and pacing speed. *International Journal of Psychophysiology*, 27(2):87–97, 1997.

[134] W. Rief, R. Shaw, and M. Fichter. Elevated levels of psychophysiological arousal and cortisol in patients with somatization syndrome. *Psychosomatic Medicine*, 60(2):198–203, 1998.

[135] D. Rissén, B. Melin, L. Sandsjö, I. Dohns, and U. Lundberg. Surface EMG and psychophysiological stress reactions in women during repetitive work. *European Journal of Applied Physiology*, 83:215–222, 2000.

[136] T. Ritz, E. Simon, and A. Trueba. Stress-induced respiratory pattern changes in asthma. *Psychosomatic Medicine*, 73(6):514–521, 2011.

[137] T. Ritz, A. Steptoe, S. DeWilde, and M. Costa. Emotions and stress increase respiratory resistance in asthma. *Psychosomatic Medicine*, 62(3):401–412, 2000.

[138] I. Romero, M. Xie, A. Link, R. Bousseljot, D. Kreiseler, C. Elster, M. Bär, and H. Koch. Analysis of the atrial fibrillation main frequency in the PTB ECG database. In *European Study Group on Cardiovascular Oscillations*, pages 120–122, 2006.

[139] W. Roth, J. Margraf, A. Ehlers, C. Taylor, R. Maddock, S. Davies, and W. Agras. Stress test reactivity in panic disorder. *Archives of General Psychiatry*, 49(4):301–310, 1992.

[140] J. Rudolph and N. Repenning. Disaster dynamics: Understanding the role of quantity in organizational collapse. *Administrative Science Quarterly*, 47(1):1–30, 2002.

[141] L. Salahuddin, J. Cho, M. Jeong, and D. Kim. Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4656–4659, 2007.

[142] L. Schleifer, T. Spalding, S. Kerick, J. Cram, R. Ley, and B. Hatfield. Mental stress and trapezius muscle activation under psychomotor challenge: a focus on EMG gaps during computer work. *Psychophysiology*, 45:356–365, 2008.

[143] P. Schnall, C. Pieper, J. Schwartz, R. Karasek, Y. Schlussel, R. Devereux, A. Ganau, M. Alderman, K. Warren, and T. Pickering. The relationship between 'job strain', workplace diastolic blood pressure, and left ventricular mass index. Results of a case-control study. *Journal of the American Medical Association*, 263(14):1929–1935, 1990.

[144] P. Schnall, J. Schwartz, P. Landsbergis, K. Warren, and T. Pickering. Relation between job strain, alcohol, and ambulatory blood pressure. *Hypertension*, 19(5):488–494, 1992.

[145] P. Schnall, J. Schwartz, P. Landsbergis, K. Warren, and T. Pickering. A longitudinal study of job strain and ambulatory blood pressure: Results from a three-year follow-up. *Psychosomatic Medicine*, 60(6):697–706, 1998.

[146] N. Schommer, D. Hellhammer, and C. Kirschbaum. Dissociation between reactivity of the hypothalamus-pituitary-adrenal axis and the sympathetic-adrenal-medullary system to repeated psychosocial stress. *Psychosomatic Medicine*, 65(3):450–460, 2003.

[147] C. Schubert, M. Lambertz, R. Nelesen, W. Bardwell, J.-B. Choi, and J. Dimsdale. Effects of stress on heart rate complexity—a comparison between short-term and chronic stress. *Biological Psychology*, 80(3):325–332, 2009.

[148] SENIAM. Recommendations for sensor locations in shoulder or neck muscles. http://www.seniam.org/, 1999.

[149] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on information technology in biomedicine*, 14(2):410–417, 2010.

[150] B.-J. Shen, L. R. Stroud, and R. Niaura. Stress differences in cardiovascular responses to laboratory stress: a comparison between Asian and white Americans. *International Journal of Behavioral Medicine*, 11(3):181–186, 2004.

[151] K. Sibley, G. Mochizuki, J. Frank, and W. McIlroy. The relationship between physiological arousal and cortical and autonomic responses to postural instability. *Experimental Brain Research*, 203(3):533–540, 2010.

[152] T. Spalding, L. Schleifer, B. Hatfield, S. Kerick, and J. Cram. Removing the influence of the heart from surface-recorded EMG. *Biofeedback*, 31:6–10, 2003.

[153] A. Steptoe. Invited review: The links between stress and illness. *Journal of Psychosomatic Research*, 35(6):633–644, 1991.

[154] A. Steptoe, M. Cropley, J. Griffith, and K. Joekes. The influence of abdominal obesity and chronic work stress on ambulatory blood pressure in men and women. *International Journal of Obesity and Related Metabolic Disorders*, 23(11):1184–1191, 1999.

[155] A. Steptoe, G. Fieldman, O. Evans, and L. Perry. Cardiovascular risk and responsivity to mental stress: The influence of age, gender and risk factors. *Journal of Cardiovascular Risk*, 3(1):83–93, 1996.

[156] A. Steptoe, G. Willemsen, N. Owen, L. Flower, and V. Mohamed-Ali. Acute mental stress elicits delayed increases in circulating inflammatory cytokine levels. *Clinical Science*, 101(2):185–192, 2001.

[157] J. Stroop. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6):643–662, 1935.

[158] S.-I. Suzuki, H. Kumano, and Y. Sakano. Effects of effort and distress coping processes on psychophysiological and psychological stress responses. *International Journal of Psychophysiology*, 47(2):117–128, 2003.

[159] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability. standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3):354–381, 1996.

[160] J. Tomaka, J. Blascovich, R. Kelsey, and C. Leitten. Subjective, physiological, and behavioral effects of threat and challenge appraisal. *Journal of Personality and Social Psychology*, 65:248–260, 1993.

[161] J. Tomaka, J. Blascovich, J. Kibler, and J. Ernst. Cognitive and physiological antecedents of threat and challenge appraisal. *Journal of personality and social psychology*, 73(1):63–72, 1997.

[162] C. Tsigos and G. Chrousos. Hypothalamic-pituitary-adrenal axis, neuroendocrine factors and stress. *Journal of Psychosomatic Research*, 53:865–871, 2002.

[163] B. Uchino, C. Berg, T. Smith, G. Pearce, and M. Skinner. Age-related differences in ambulatory blood pressure during daily stress: evidence for greater blood pressure reactivity with age. *Psychology and Aging*, 21(2):231–239, 2006.

[164] A. Väänänen, E. Anttila, J. Turtiainen, and P. Varje. Formulation of work stress in 1960–2000: Analysis of scientific works from the perspective of historical sociology. *Social Science & Medicine*, 75(5):784–794, 2012.

[165] T. Vrijkotte, L. van Doornen, and E. de Geus. Effects of work stress on ambulatory blood pressure, heart rate, and heart rate variability. *Hypertension*, 35(4):880–886, 2000.

[166] E. Wethington, D. Almeida, G. W. Brown, E. Frank, and R. C. Kessler. The assessment of stressor exposure. In A. Vingerhoets, editor, *Assessment in Behavioral Medicine*, pages 113–134. Taylor & Francis, 2004.

[167] C. Wetzel, S. Black, G. Hanna, T. Athanasiou, R. Kneebone, D. Nestel, J. Wolfe, and M. Woloshynowych. The effects of stress and coping on surgical performance during simulations. *Annals of Surgery*, 251(1):171–176, 2010.

[168] J. Wijsman, R. Vullers, S. Polito, C. Agell, J. Penders, and H. Hermens. Towards ambulatory mental stress measurement from physiological parameters. In *Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 564–569, Sept 2013.

[169] F. H. Wilhelm and W. T. Roth. Taking the laboratory to the skies: Ambulatory assessment of self-report, autonomic, and respiratory responses in flying phobia. *Psychophysiology*, 35(5):596–606, 1998.

[170] O. Wolf. The influence of stress hormones on emotional memory: relevance for psychopathology. *Acta Psychologica*, 127(3):513–531, 2008.

[171] A. Yadav, M. Phillips, M. Lundeberg, M. Koehler, K. Clouse, and K. Dirkin. If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. *Journal of Computing in Higher Education.*, 23(1):15–37, 2011.

[172] R. Yazicioglu, P. Merken, R. Puers, and C. Van Hoof. A 60 μW 60 nV/$\sqrt{}$Hz readout front-end for portable biopotential acquisition systems. *IEEE Journal of Solid-State Circuits*, 42:1100–1110, 2007.

[173] R. Yerkes and J. Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurological Psychology*, 18:459–482, 1908.

[174] J. Zhai and A. Barreto. Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1355–1358, 2006.

[175] J. Zhai, A. Barreto, C. Chin, and C. Li. Realization of stress detection using psychophysiological signals for improvement of human-computer interactions. In *IEEE SoutheastCon*, pages 415–420, 2005.

Summary

Recently, general concern about work-related stress is increasing. The number of people affected by work-related stress has grown since the second half of the 20th century. Examples of causes of work stress are a bad person-environment fit and a highly demanding job with low control. If such an imbalance between person and environment and/or demand and control persists, chronic stress will occur. Chronic stress induces a number of mental and physical health problems that impact personal life, organizations and society. Timely detection and reduction of stress could prevent these health problems and their negative effects. There are three general approaches for stress detection: questionnaires, biochemical measures and physiological measures. This thesis focuses on the detection of stress using physiological measures. Stress causes rapid activation of the autonomic nervous system and this activation can be measured in a number of physiological variables. The goal of this thesis is:

to assess the feasibility of constructing personal models for the relation between mental stress and physiological variables, for use in ambulatory stress management systems.

The most commonly measured physiological variables for stress detection are reviewed in Chapter 2: heart rhythm, blood pressure, respiration and skin conductance. Five features that were found to be related to mental stress are: heart rate, systolic and diastolic blood pressure, skin conductance level and number of skin conductance responses.

In Chapter 3 and Chapter 4, an experiment is described in which physiological signals were measured with a wireless system during three distinctive stress tests that were designed to resemble office-like situations. Electrocardiogram (ECG), respiration, skin conductance and surface electromyogram (EMG) of the upper trapezius muscle were measured and analyzed. EMG amplitude was higher during

stress than during rest and relative time with EMG gaps, mean frequency and median frequency of the EMG signal were lower during stress than during rest (Chapter 3). Using all four physiological signals, classification between stress and rest conditions reached an accuracy of 79.3% using Principal Component Analysis and a Fisher's Least Square Classifier, and an accuracy of 74.5% using Generalized Estimating Equations (Chapter 4).

In Chapter 5, two randomized controlled trials are described that compare the use of immersive introductory videos with textual introduction in simulation-based medical education (SBME). Both studies showed that the introduction induced increased psychological and physiological stress levels. There were indications that the video introduction induced higher psychological stress than the textual introduction and the task that followed the video was appraised upfront as a threat rather than a challenge. These differences could influence the effect of education and need to be investigated further so the optimal activity level for most effective learning can be found.

The final study was performed in an uncontrolled, real-world environment and is described in Chapter 6. ECG, EMG from the upper trapezius muscles, respiration, skin conductance and skin temperature were monitored during five days in an ambulatory setting. Self-reported stress levels and salivary cortisol were collected as reference measures. Context information on eating, drinking, posture and physical activity was also collected. Results show that personal stress models can be determined that predict mental stress levels at an individual level. However, there was not much consistency among the physiological responses of different subjects.

To the best of our knowledge, this real-world study is the first study in an ambulatory setting in which this large number of physiological, context and reference variables were recorded. Five different physiological signals, eight context variables and three reference measures were collected from ten subjects during five days.

The main conclusions of this thesis are that using physiological variables for mental stress detection is feasible, personalization is necessary due to large variations among persons, and that ambulatory measurements are feasible if an unobtrusive and low-power sensor is available. Two options have been provided for estimating stress level on a continuous scale: Generalized Estimating Equations and regression. The most common features used in stress estimation are blood pressure, heart rate and skin conductance. Other features such as heart rate variability, EMG and temperature are relevant for some subjects, but not relevant for others. Respiration frequency could be a useful feature but is heavily influenced by speech.

There are some difficulties in the research field that need attention in future work. First of all, there is no recognized reference measure available that is known to resemble actual mental stress level accurately. Secondly, there is much potential influence on the measurements by context factors such as physical activity, coffee consumption and ambient temperature. Furthermore, additional physiological variables such as blood pressure can be investigated, as well as the prospects for

real-time stress level estimation and ways to give effective feedback to the user. Finally, the characteristics of subjects should be chosen carefully and long-term measurements are needed to investigate chronic stress.

Samenvatting

De laatste tijd neemt de algehele bezorgdheid over werkgerelateerde stress toe. Het aantal mensen dat getroffen wordt door werkgerelateerde stress groeit sinds de tweede helft van de 20e eeuw. Voorbeelden van oorzaken van werkstress zijn het slecht bij elkaar passen van persoonlijkheid en omgeving of een baan die veel vraagt, maar weinig controle biedt. Als een dergelijke onbalans tussen een persoon en zijn omgeving en/of vraag en controle aanhoudt, zal chronische stress optreden. Chronische stress veroorzaakt een aantal mentale en fysieke gezondheidsproblemen die invloed hebben op privéleven, organisaties en de samenleving. Tijdige opsporing en vermindering van stress zouden deze gezondheidsproblemen en de negatieve effecten ervan kunnen voorkomen. Er zijn drie gangbare benaderingen voor het opsporen van stress: vragenlijsten, biochemische metingen en fysiologische metingen. Stress zorgt voor een acute activering van het autonome zenuwstelsel en deze activering kan worden gemeten in een aantal fysiologische variabelen. Het doel van dit proefschrift is:

het bepalen van de haalbaarheid van het ontwikkelen van persoonlijke modellen die de relatie tussen mentale stress en fysiologische variabelen beschrijven, ten behoeve van gebruik in ambulante systemen voor stressregulatie.

De meest gebruikelijke fysiologische variabelen voor stressdetectie worden besproken in Hoofdstuk 2: hartritme, bloeddruk, ademhaling en huidgeleiding. Vijf karakteristieken waarvoor een relatie met mentale stress werd gevonden zijn: hartslag, systolische en diastolische bloeddruk, niveau van huidgeleiding en aantal gemeten reacties in het huidgeleidingssignaal.

In Hoofdstuk 3 en Hoofdstuk 4 wordt een experiment beschreven waarin fysiologische signalen werden gemeten met een draadloos systeem tijdens drie verschillende stress tests die werden ontworpen om situaties na te bootsen zoals ze in een kantoor

optreden. Elektrocardiogram (ECG), ademhaling, huidgeleiding en oppervlakte elektromyogram (EMG) van de musculus trapezius pars descendens (monnikskapspier) werden gemeten en geanalyseerd. De amplitude van het EMG was hoger tijdens stress dan tijdens rust. De relatieve tijd met EMG 'gaten', gemiddelde frequentie en mediaanfrequentie van het EMG signaal waren lager tijdens stress dan tijdens rust (Hoofdstuk 3). Bij gebruik van alle vier de signalen voor classificatie tussen stress- en rustcondities werd een nauwkeurigheid van 79.3% behaald met een *Principal Component Analysis* en een *Fisher's Least Square Classifier*, en een nauwkeurigheid van 74.5% met gebruik van *Generalized Estimating Equations* (Hoofdstuk 4).

In Hoofdstuk 5 worden twee gerandomiseerde onderzoeken met controlegroepen beschreven waarin het gebruik van introductievideo's vergeleken werd met het gebruik van een introductietekst in medische educatie met behulp van simulaties. Beide studies toonden aan dat de introductie zorgde voor verhoogde psychologische en fysiologische stressniveaus. Er waren indicaties dat de introductievideo hogere psychologische stress veroorzaakte dan de introductietekst en de taak die na de video uitgevoerd moest worden werd van tevoren gewaardeerd als een bedreiging en niet als een uitdaging. Deze verschillen kunnen invloed hebben op het effect van educatie en moeten verder worden onderzocht zodat het optimale niveau van activering voor het meest effectieve leerproces kan worden vastgesteld.

De laatste studie vond plaats in een ongecontroleerde natuurlijke omgeving en wordt beschreven in Hoofdstuk 6. ECG, EMG van de monnikskapspier, ademhaling, huidgeleiding en huidtemperatuur werden gedurende vijf dagen gemeten in een ambulante omgeving. Door de proefpersonen gerapporteerde stressniveaus en cortisol in speeksel werden verzameld als referentiemetingen. Ook werd contextinformatie over eten, drinken, lichaamshouding en fysieke activiteit verzameld. De resultaten toonden aan dat persoonlijke stressmodellen kunnen worden bepaald die mentale stress op individueel niveau voorspellen. Er was echter niet veel samenhang tussen de fysiologische reacties van de verschillende proefpersonen.

Naar eer en geweten wordt gesteld dat deze studie de eerste ambulante studie in een natuurlijke omgeving is waarbij dit grote aantal fysiologische, context- en referentievariabelen werd gemeten. Vijf verschillende fysiologische signalen, acht contextvariabelen en drie referentiemetingen werden verzameld van tien proefpersonen gedurende vijf dagen.

De belangrijkste conclusies van dit proefschrift zijn dat het haalbaar is om fysiologische variabelen te gebruiken voor het detecteren van mentale stress, dat personalisatie nodig is door de grote variatie tussen personen, en dat ambulante metingen haalbaar zijn als een niet hinderlijke en energiezuinige sensor beschikbaar is. Er zijn twee opties genoemd voor het schatten van het stressniveau op een continue schaal: *Generalized Estimating Equations* en regressie. De meest voorkomende kenmerken die bij het schatten van stress worden gebruikt zijn bloeddruk, hartslag en huidgeleiding. Andere kenmerken zoals hartslagvariabiliteit, EMG en temperatuur zijn weliswaar voor sommige proefpersonen relevant, maar voor anderen niet. Ademhalingsfrequentie kan bruikbaar zijn maar wordt aanzienlijk

beïnvloed door spreken.

Een aantal problemen in het onderzoeksdomein behoeft aandacht in toekomstig onderzoek. Ten eerste is er geen erkende referentiemeting beschikbaar die bekend-staat als een accurate weergave van het werkelijke stressniveau. Ten tweede is er veel potentiele invloed op de metingen door contextfactoren zoals fysieke activiteit, koffieconsumptie en omgevingstemperatuur. Verder kunnen extra fysiologische va-riabelen zoals bloeddruk worden onderzocht, en kan onderzoek worden gedaan naar mogelijkheden voor actuele schatting van het stressniveau en naar manieren om effectieve feedback te geven aan de gebruiker. Ten slotte moeten de eigenschappen van proefpersonen zorgvuldig worden gekozen en zijn langetermijnmetingen nodig om chronische stress te onderzoeken.

Acknowledgements

At last, the time has come to write the final words of my thesis. Many people have contributed to this work and my well-being in different ways.

First, of course, I need to thank all subjects who participated in my studies. Without your cooperation it would have been impossible to carry out the research and I would not have been able to write this thesis.

Thanks as well to my promotor, Hermie, who gave me the opportunity to do this PhD project. He always provided useful feedback to my work and tried many times to make me realize that I was doing a good job.

Over the years, the position of supervisor at Holst Centre was fulfilled by several people. Bernard, Julien, Ruud and Vojkan, thank you for your useful guidance throughout the project.

Then there are the other (former) colleagues of the Body Area Network group. Many thanks to all of you for your help: the technical support, but also the many coffee breaks, lunches and other activities we did together. You are not only my colleagues, but also good friends. There were many days that I was only looking forward to the next break I could spend with you. Seulki, some special words for you. I have enjoyed all our chats about various topics from work to traveling to relationships and much more. I am very happy that you are my paranymph!

Special thanks to the office mates over the years: Aline, Belén, Marco, Di, Carlos, Vojkan, Lindsay, Eva and Bernard. The experience of sharing an office was very different with all of you, but always a pleasure and I had a lot of fun with all of you.

Of course I cannot forget my other friends from outside work. The moments we had together made me gain the energy to finish this project. My non-Dutch

colleagues were always surprised if I had yet another birthday party of a friend during the weekend. I realized that it is a Dutch tradition to keep celebrating birthdays as a grown-up, but it is a tradition I enjoy a lot! Also the non-birthday drinks, dinners, and other meetings were always a lot of fun. Kirsten, the paths of our lives have run parallel over the past decade: we started at university together, got a boyfriend, received our bachelor and master degrees, got married and now I catch up to you again by finishing my PhD. I always enjoyed our dates at UT when I was in Enschede and our talks on various topics, including the complaints about doing a PhD. Thank you for being my paranymph!

Members of my choir Dississit, I also want to thank you. It often takes some effort to get off the couch on Friday evening for a rehearsal, but it has always been worth it. It is such a good feeling to close the week with singing (and often also the afterparty).

Some of the most important people in my life are my family and they definitely deserve a few special words. First of all my grandparents. I admire how you manage to keep up with all new technical developments. You always show interest in my work and more than once I received a relevant article in the mail that you found in the newspaper and signed with my name. My brothers, Rob and Marc, and my sister, Monique, thank you for the fun during our brother-sister-weekends and all other occasions. Although the physical distance between us has grown larger, I feel that our bonds have become stronger. And of course my parents, thank you so much for all your love, care and support throughout my life. Thank you for putting your trust in me and for bringing me up to independence.

Joost, thank you for your love, your support, for cheering me up when needed, for listening to my worries, and for making me put everything in perspective from time to time.