Dissertation

Understanding Stress Responses Related to Digital Technologies

Romina Poguntke

Universität Stuttgart

Understanding Stress Responses Related To Digital Technologies

Von der Fakultät für Informatik, Elektrotechnik und Informationstechnik und dem Stuttgart Research Centre for Simulation Technology (SRC SimTech) der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Romina Poguntke geb. Kettner

aus Moers, Deutschland

Hauptberichter:	Prof. Dr. Albrecht Schmidt
Mitberichter:	Prof. Dr. Jonna Häkkilä
Mitberichter:	Prof. Dr. Niels Henze
Tag der mündlichen Prüfung:	09.12.2019

Institut für Visualisierung und Interaktive Systeme
der Universität Stuttgart

2020

ABSTRACT

Feeling stressed is a common phenomenon which is known among all age groups, cultures, and societal classes. Although its perception and recognition may follow universal rules too, humans experience the reasons for stress and the success of stress mitigating techniques differently. Research in various domains, such as psychology, biology, or neuroscience have made great progress in understanding the evolutionary roots of stress and its stressors, namely stimuli that elicit stress reactions. However, the nature of stressors has changed significantly over time. While thousands of years ago humans experienced stress physiologically when facing a dangerous animal in the wilderness, nowadays stress is a harmful consequence of the permanent challenge for our brain to process the constant stream of information. As a result, the modern human feels stressed by the omnipresent access to information resulting in information overload, going hand in hand with the societal expectations to be reachable and so-called "online" anytime. The consequential problems for users, as well as design challenges for designers and developers to avoid such disagreeable effects have been recently addressed in the field of Human-Computer Interaction (HCI).

In this thesis, I identify, describe and successively explore the character of different inventions for reducing the stress level of users and their impact on them. Hereby, I present three exemplary research probes differing in their degree of privateness and their degree of using digital technologies. Thus, my contribution in the human-computer interaction domain lies in the investigation of how to design stress mitigating interventions for interactive systems and respectively, what impact the manipulation of potential stressors can have on users. One prerequisite for reducing stress is, knowing when a user is stressed. Thus, I will provide the foundation for the research being conducted with physiological measurements in this thesis by showing that subjective self-measures of stress have physiology-based correlates (Chapter 3). As the measurements' quality depends, among other factors, on the hardware it is being carried out with, I will further emphasize that the measurability goes along with certain constraints which I clustered and summarized forming the "*Design Space for Physiology Aware Systems*" (Chapter 4). Given the reliable detection of stress using physiological data, I explored tactile feedback for notifying users about stress (Chapter 5). While the first approach concentrated on the comparison of an existing feedback mechanism, namely vibration which is already used for smartphones and wrist-worn wearables, to pressured-based feedback, the second exploration exploited the advantages of thermal stimulation and focused on investigating the stimuli's properties to be preferred for giving stress feedback. Concluding from the aforementioned explorations that feedback on stress is conditionally desirable from the user's perspective, I designed three exemplary interventions manipulating stressors (Chapter 6). First, an Android application was implemented delaying smartphone notifications and thereby, eliminated potential stressors, here notifications. For the second intervention cognitive load, as a another stressor was visualized on an ambient display putting the user in a passive role requiring no additional attention or action. In contrast to this, in the third investigation the user was asked to adjust a physical wearable display reflecting the self-reported feeling of being busy. Here, the user was required to take action him- or herself which further facilitated conversations, since the display was designed as necklace visible for anyone.

By performing the described research activities representing distinct degrees of privateness and in their use of digital technologies, I show that the manipulation of a stressor supports self-reflection in users. Hereby, three settings, namely from private to semi-public to public, as well as three exemplary tools, namely the smartphone, the ambient display, and the physical display have been employed. I show that the conscious confrontation with the stressors helps users to self-reflect about future behavior regarding the coping with stressors. Further, valuable insights on side-effects when visualizing stress-related information can be inferred from the user feedback implying also privacy preservation. In conclusion, in this thesis I present the "*Design Space for Physiology-Aware Systems*" and the derived "*Design Recommendations for Stress-Aware Interactive Systems*", which contribute to a better understanding of what should be considered when designing stress mitigating techniques. Further, I demonstrate that manipulating stressors has a promising potential paving the way to future systems to support self-reflection and mutual consideration of one's stress.

ZUSAMMENFASSUNG

Das Gefühl gestresst zu sein, ist ein alltägliches Phänomen, das unter allen Altersgruppen, Kulturen und sozialen Schichten bekannt ist. Obwohl die Wahrnehmung und das Erkennen von Stresssituation ebenfalls für viele Menschen ähnlich sein dürfte, so unterscheidet sich die Einschätzung des Erfolgs von stressreduzierenden Interventionen. Denn trotz der großen Fortschritte, die die Wissenschaft bezüglich des Verständnisses von Stress und Stressoren in verschiedensden Domänen, wie Psychologie, Biologie or Neurowissenschaften gemacht hat, so hat sich die Natur des Stressoren in den vergangenen Jahrzehnten drastisch verändert, wobei man unter Stressoren Stimuli versteht, die Stressreaktionen auslösen. Während vor Tausenden von Jahren Menschen die physiologischen Indikatoren für Stress erlebten, wenn sie einem wilden Tier gegenüber standen, so ist Stress heutzutage eine Konsequenz der permanenten Herausforderung unseres Gehirns den konstanten Informationsfluss zu verarbeiten. Daraus resultiert, dass sich der moderne Mensch durch den omnipräsenten Zugang zu Informationen gestresst fühlt, was widerum zu einem Empfinden von Überfordung führt. Damit einher geht das Gefühl ständig erreichbar oder online sein zu müssen, um soziale Erwartungen zu erfüllen. Die daraus entstehenden Probleme für Nutzer, wie ebenso die Herausforderungen für Designer und Entwickler um solche unangenehmen Effekte zu vermeiden, werden bereits seit einiger Zeit in dem Bereich der Mensch-Computer-Interaktion behandelt.

Mit dem Ziel eine angemessene und effektive Interventionstechnik zur Reduktion von Stress zu finden, exploriert die vorliegende Arbeit sukzessive verschiedene Interventionsarten und deren Einfluss auf den Nutzer. Hierzu präsentiere ich drei beispielhafte Anwendungen, die sich in ihrer Ausprägung der Privatssphäre und Digitalisierung unterscheiden. Damit liegt mein wissenschaftlicher Beitrag im Feld der Mensch-Computer-Interaktion, in der Erforschung, wie sich Stressreduzierende Interventionen für interaktive Systeme gestalten lassen und welchen Einfluss die Manipulation von Stressoren auf deren Nutzer hat.
Eine Voraussetzung zur Reduktion von Stress ist es, zu wissen wann der Nutzer gestresst ist. Daher bildet das Zeigen des Zusammenhangs zwischen der subjektiven Selbstbeurteilung des Stresszustandes und den auf physiologischen Indikatoren basierenden Korrelaten die Grundlage für die in dieser Arbeit durchgeführten physiologischen Messungen (Kapitel 3). Dadurch, dass Messgüte dieser Erhebungen unter anderem von der Wahl der Instrumente abhängt,

gehe ich des Weiteren auf die damit verbundenen Einschränkungen ein, welche in dem "*Designraum für physiologisch-basierte Systeme*"gruppiert und zusammengefasst werden (Kapitel 4). Vor dem Hintergund der zuverlässigen Erkennung von Stresszuständen auf Grundlage von physiologischen Daten, habe ich die Benachrichtung von Nutzern über ihren Stresszustand mit Hilfe von taktilem Feedback exploriert (Kapitel 5). Während sich der erste Ansatz auf den Vergleich zwischen bestehenden Feedbackemethoden wie Vibration, was bereits für Smartphones und Wearables verwendet wird, und Druckbasiertem Feedback konzentrierte, so zielte die zweite Exploration auf die Untersuchung der Vorteile von Temperaturbasierter Stimulation ab. Dabei fokussierte ich mich auf die Erforschung der Eigenschaften des Stimulus, um ein Nutzerfreundliches Feedback geben zu können.
Mit der auf den zuvor erwähnten Untersuchungen basierenden Schlussfolderung, dass explizites Feedback bezüglich des Stresszustandes nur bedingt wünschenswert ist aus Nutzersicht, gestaltete ich drei exemplarische Anwendungen zur Manipulation von Stressoren (Kapitel 6). In der ersten Studie wurde eine Android Anwendung implementiert, die Smartphone-Benachrichtigungen verzögert und damit potentielle Stressoren, in diesen Fall Benachrichtigungen, eliminiert. Für die zweite Intervention wurde kognitive Last, als ein weiterer Stressor auf einem ambienten Bildschirm visualisiert, was den Nutzer in einer passiven Rolle ließ, da keine spezielle Aufmerksamkeit oder Aktion gefordert war. Im Gegensatz dazu wurde der Nutzer in der dritten Untersuchung gebeten auf einem physikalischen tragbaren Prototypen das selbsteingeschätzte Gefühl von Geschäftigkeit anzugeben. Hierbei war es notwendig, dass der Nutzer selbstständig agierte, was ebenfalls Konversationen erleichterte, da der Protoyp als Kette sichtbar für alle war.

Mit der Umsetzung der beschriebenen Forschungsarbeiten, die verschiedene Abstufungen der Privatsphäre von privat über semi-öffentlich bis hin zu öffentlich sowie der Nutzung von Smartphones, ambienten Bildschirmen und physikalischen Prototypen als beispielhafte digitalisierte Werkzeuge umfasst, konnte ich demonstrieren, dass die Manipulation von Stressoren die Selbstreflektionsfähigkeiten von Nutzern unterstützt. Es zeigt sich, dass die bewusste Konfrontation mit Stressoren den Nutzern hilft über ihr zukünftiges Verhalten im Umgang mit den Stressoren zu reflektieren. Zudem konnten umfassende Erkenntnisse über die Nebeneffekte der Visualisierung Stress-bezogener Informationen aus den Nutzerrückmeldungen gewonnen werden, was ebenfalls Implikationen für den Schutz der Privatsphäre hat. Abschließend lässt sich festhalten, dass die vorliegende Arbeit zu einem besseren Verständnis beiträgt was für die Gestaltung stressreduzierender Interventionen zu beachten ist. Darüber hinaus wird deutlich, dass die Manipulation von Stressoren viel Potenzial bietet um auch in zukünftigen System zum Einsatz zu kommen.

PREFACE

This thesis contains work created from 2016 till 2019 at the University of Stuttgart. Since studying of stress responses requires different types of expertise from distinct disciplines, this thesis has been done in close collaboration with experts from the University of Stuttgart, projects partners within the EU ERC-AMPLIFY (683008) and SFB-TRR 161 projects, as well as external collaborators. These collaborations resulted in publications which are a core part of this thesis. The contributing authors (i.e., co-authors of papers) are clearly stated at the beginning of each chapter together with the reference to the publication if applicable. To keep the consistency throughout the thesis, I use the term "I" instead of "we" when referring to myself.

ACKNOWLEDGMENTS

First, I would like to thank my supervisor Albrecht Schmidt who had confidence in me. Thank you for encouraging me when things did not go as planned and for pursuing me to finish this work while the "natural" deadline was approaching.

My special thanks also goes to Jonna Häkkilä who did not only agree to examine this thesis, but who was a great host during my stay in Rovaniemi at the University of Lapland.

In addition, I would like to thank Niels Henze for investing time in my thesis and for asking good but sometimes weird questions which often turned out to provide helpful feedback.

Throughout my Ph.D. journey so many interesting things happened and I encountered so many inspiring and great minds. Although the former Stuttgart HCI lab slowly melted away and dispersed around the world (or at least Germany), I want to say thank you to the people who shaped my everyday life for a very long time.

Thank you Stefan Schneegaß for the endless discussions about my ideas and visions - however crazy they were, you mostly listened patiently before you told me they were perhaps not the best. At the very least, we got something from almost each project we started and sometimes the outcome was more experience-based than scientific. I enjoyed working with you and having another "Pottkind" in the lab who shared my passion for football.

Thank you Thomas Kosch for making motionEAP a memorable experience. Based on my observations from countless Nerf gun battles, exciting project meetings and our trips to Berlin, I reached the conclusion that you were the craziest hacker in our group which made it so fun working with you.

Thank you Francisco Kiss, my last office mate. Although our conversations did not always make sense, I enjoyed working with you and philosophizing about the meaning of life.

Not to forget, the three Egyptian ladies deserve a special thank you. We shared rooms and from time to time offices, we shared food and we figured out that we also shared hopes, concerns and dreams beyond research life. Thanks to Mariam Hassib, Yomna Abdelrahman and Passant El-Agroudy I discovered my love for arabic food and aggressive sounding languages. Not only, did you make my Bachelorette's Party unforgettable, but also the trip to New York, which we all happily survived.

Thank you Christina Schneegaß for offering me a safe place next to Pino in your Munich apartment anytime. Funnily, you marked the starting point of my Ph.D. career back in 2015, when we shared rooms as student volunteers during the "Mensch und Computer" Conference in Stuttgart.

My very special thanks also goes to the people with whom I had endless Skype meetings because you were located in different places and time zones. What started as an ambitious project idea, and was mostly successful, resulted in friendships and I am grateful to have met you. Special thanks to Katrin Hänsel who became much more than just a co-author. Thank you also to my Finnish colleagues, Aku Visuri and Olli Korhonen, with whom I spent an amazing week in Oulu at the UBIS Summer School back when we were all naive, yet joyful, Ph.D. students. Another important person I would like to thank is Emmi Harjuniemi. You made my research stay in Rovaniemi one of the best experiences in my life. Thanks to you, I will not forget Santa's advice to have fun each day of the year.

To the Stuttgart colleagues who are now spread all over the world: Patrick Bader, Lars Lischke, Markus Funk, Florian Alt, Pacal Knierim, Miriam Greis, Tonja Machulla, Pawel Woźniak, Rufat Rzayev, Huy Viet Le, Sven Mayer, Vali Schwind, Alex Voit, Dominik Weber, Jakob Karolus and Tilman Dingler, thank you.

I would also like to thank Anja Mebus and Murielle Naud-Barthelmeß who kept the group together for so long and were always open for good conversations over a warm cup of tea.

Last, but certainly not least, I would like to say thank you to my family that embraces so many more people now - thanks to Tim's huge family. Thank you for keeping me up, supporting me with whatever I needed, whether it be for annoying experiments, last minute language corrections, jovial gatherings and a neverending open ear. My parents also deserve a special thanks, Jutta and Norbert Kettner, you are so special to me.

Finally, there is one person for whom I cannot express how thankful I am having him in my life. Thank you Tim for walking with me on the road of life and giving me a helping hand, even I do not ask for it and even if it just means to bring me my favorite chocolate. I am looking forward to what lies ahead of us and to the exciting adventures this journey holds for us and our family. This brings me to someone who is yet too young to realize what value he adds to my life. Thank you Max, not only for determining a "natural" deadline for finishing this thesis, but for bringing so much joy and happiness to my life.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ANS Autonomous Nervous System

API Application Programming Interface

BVP Blood Volume Pulse

CMDQ Cornell Musculoskeletal Discomfort Questionnaire

CSS Cascading Style Sheets

CSV Comma-Separated Values

ECG Electrocardiography

EDA Electrodermal activity

EEG Electroencephalography

EMG Electromyography

EMS Electrical Muscle Stimulation

EOG Electrooculography

ESM Experience Sampling Method

FI Fixed Interval

FOMO Fear of Missing Out

GAS General Adaption Syndrom

GSR Galvanic Skin Response

HCI Human-Computer Interaction

HF/LF Ratio High Frequency/Low Frequency Ratio

HPA Axis Hypothalamic-Pituitary-Adrenal Axis

HR Heart Rate

HRV Heart Rate Variability

HTML HyperText Markup Language

HTTP Hypertext Transfer Protocol

Hz Hertz

ICT Information Communication Technology

JSON JavaScript Object Notation

MAT Mental Arithmetic Tasks

NASA-TLX NASA Task Load Index

PPG Photoplethysmography

rMSSD Root Mean Square of Successive Differences

RQ Research Question

SAM Self-Assessment Manikin Scale

SD Sender-Dependent

SDK Software Developer Kit

SDNN Standard Deviation of Normal Sinus Beats

SNS Sympathetic Nervous System

SSSQ Short Stress State Questionnaire

ST Skin Temperature

UDI User-Defined Interval

UUID Universally Unique Identifier

Chapter 1

Introduction

Stress is a complex phenomenon which has been examined intensively in the past decades. Many research disciplines, such as biology, psychology, neuroscience, and medicine have investigated, simulated, and modeled the nature of stress. Despite the extensive and ambitious attempts to open up the field of stress, even modern science has difficulties to capture all facettes of such a complex construct as stress is [198, 229, 233]. In the course of exploring its characteristics, several correlates have been discovered. These related constructs have been identified as so-called stressors. A stressor can be understood as a source of stress eliciting an unpleasant feeling of being stressed in a person. According to Lepine, Podsakoff, and Lepine [158] stressors initiate the start of a stress process which results in strains. Such strains can take different forms, ranging from workload, job demands, role conflicts and lead to severe consequences like anxiety, exhaustion or depression [125]. These illnesses are provoked by chronic stress much more often than by one-time life events referred to as acute stress [176]. As a consequence of chronic stress, psychiatric disorders like the aforementioned, have increased dramatically within the past years. When the market research institute Gallup asked people from 142 countries about their negative feelings, they found in their 2019 global emotions report that 39% of the respondents experienced "a lot" of worry and another 35% agreed to experiencing "a lot" of stress [78]. Furthermore, with the investigation of stress as a harmful trigger for depressions, anxiety, and cardiovascular diseases, also its stressors have been studied in length [16, 176, 215, 238, 252]. Besides unhealthy lifestyles, such as too little physical activity, too much unhealthy food, there are also sources of stress derived from our technology usage [16, 134, 152, 154, 242, 289].

A systematic review of 35 studies suggested that there is a correlation between the use of Information Communication Technology (ICT) and stress, although the users' age does not seem to mediate this relation [16]. Moreover, we know that the continuous stream of information being regarded as stimuli to be processed from a neuroscientific perspective, is one main source of stress in our contemporary society [22, 146, 154]. The resulting multitasking, as a technique to cope with the constant flow of electronic ubiquitous stimuli has shown to worsen the ability to switch between tasks [191] which further affects their attention span and memory negatively [86]. With the market launch of smartphones, the large-scale "technification" had reached another stage which has not been known before including the "invasive" consequences, as Böhme calls it [22]. Considerable differences from past cultural living, took place with the ubiquity of digital technologies of the everyday life. Facing the evolvement of the internet from an from-anywhere-accessible information system to a mass medium enabling people to receive, generate, and share information, the cognitive demands and challenges the human brain has to deal with changed completely and rapidly. While cortical anatomical structures were shaped and adapted to the environmental conditions over thousands of years, the fast technological developments did not grant sufficient time to adapt to the novel challenges. As a result the brain has difficulties processing the vast amount of information flooding the modern human every day. This information overload leads to stress [14, 112, 154] affecting personal wellbeing.
Consequently, there is a high demand for stress mitigating applications. Prior research has shown that stress can be reduced through design as highlighted by the following examples. Design that supported wellbeing [199] or technologies helping to cope with stress [69, 194, 283] has been in the focus of previous work. Moreover, the role of stressors has been largely neglected in related HCI literature, except for some works that aimed to permeate what defines a stressor [119, 134]. Like the work by Alonso, Varkevisser, and Keyson shows, the investigation of "expressive stress relievers" [30] how they refer to it, ranges from comparing different tactile qualities appealing prototypes to evaluating stress-representing biofeedback applications [76, 155, 290].

In contrast to these examples, this thesis concentrates on a more holistic approach given the highly individual character of stress lying subjectively in the eye of the beholder. As pointed out, the investigation of stress and its correlates as well as related constructs remains a difficult and complex field. Even with the fusion of experts from different disciplines working on technical or conventional interventions to fighting stress, there has not been a universal solution found yet. Thus, despite all the valuable attempts made in industry and research alike, we still lack an effective method to mitigate stress when exposed to unpleasant situations. Instead of evaluating specific applications, the present work systematically builds the prerequisite for sensing stress before identifying the shortcomings of prevalent physiology-aware hardware. As part of the exploration of two different holistic techniques to mitigate stress, this thesis focalizes the investigation of the effects on the users when testing a prototype aiming to reduce stress. Consequently, in the course of my research the users' subjective experience with the designed system has been focused. As a result of these evaluations using various research methods in the laboratory and in the wild, this thesis contributes a *Design Space for Physiology-Aware Systems"* charting the challenges when detecting stress reactions based on physiological data and an insightful investigation of the effectiveness of manipulating the potential source of stress instead of providing feedback on it. Hereby, I show that manipulating the stressor leads to higher self-reflection and an increased awareness about the preoccupation with stressors in the users. Further, the conscious confrontation, particularly when the user takes an active role in manipulating the visualization of the stressor, even supports mutual regard and can stimulate vital conversations provoking the dedicated consideration sources of stress. By concluding with design recommendations addressing the overall improvement of physiology-aware hardware, and further providing concrete suggestions on how to realize these when for example implementing a mobile application, this thesis provides important insights for designers, developers and user experience researchers. From the inferred conclusions responding to the ten research questions an overview over the considerable aspects is gained and by the exploration of suitable stress mitigating interventions the way for an effective stress-aware system is paved.

1.1 Research Questions

When elaborating on the understanding of stress responses in relation to digital systems, there are different aspects to be considered. While a significant part of prior work has concentrated on examining the consequences of using technology [16, 146, 154], such as the domain of technostress, the efforts of the affective computing community to physiologically recognize stress [90, 147, 198] and develop user-adaptive systems [116, 166, 250] represent another branch. Building upon previous work, this thesis aims to answer in total ten research questions tackling the following three aspects: **I** Physiological measurement of stress, **II** Exploring tactile feedback for stress states, and **III** Designing interventions to reduce stress. In Table 1.1 the corresponding research question this thesis strives to answer is summarized and will be explained in the following.

Despite various related work using physiological measures to sense mental states, emotions, or stress levels, in the HCI community there is still a lack of consensus which parameters suit best for detecting stress. Since the basic prerequisite for developing and designing stress mitigating systems is to reliably detect when users are stressed, the first two Research Questions (RQs) **RQ1a** and **RQ1b** provide an explanation on how stress responses can be physiologically measured and physiological and subjective data indicating stress correlate with each other. As revealed in the presented user study and inspired by related literature remarking the shortcomings of physiological sensing hardware, the research question **RQ1c** has been phrased to carve out the constraints of physiological sensing devices for interactive systems. Grounded on the answers to this questions the subsequent research question, **RQ1d** targets to identify and summarize criteria emerging from the prevalent deficiencies and motivated by user perspectives.

As the foundation for successfully detecting stress has been laid, the investigation of techniques to reduce stress in users was focalized. Facing a rich corpus of related literature that reports on the effectiveness of particular applications aiming to mitigate or control stress, the present work derived from previous work which modality could be suitable to give feedback for notifying users about stress responding to research question **RQ2a**. Given that the audio and visual channels are already used for communication technology and fall short regarding unobtrusiveness and privacy preservation, the tactile stimulation has been primarily explored. Hereby, the design of an effective feedback stimlus **RQ2b** was in the focus of research leading to the consequential question in how far feedback about the stress state is desirable from an user perspective.

Based on the results from these investigations it could be shown that notifying about stress is potentially counterproductive for reducing stress. Hence the scope of the last block of research questions is on the manipulation of stressors. By researching the effects of three different potential interventions antagonizing the stressors instead of the stress perception, the research question **RQ3a** has been answered with the demonstration of how distinct aspects could be altered exemplarily. From the findings of the three research probes representing different approaches to (a) eliminate the source of stress, (b) visualize the stressor leaving the user in a passive role, and (c) require the user's action for visualizing the source of stress, it could be shown how these affect users in general **RQ3b** and in

No.	Research Question	Chapter
	I Physiological Measurement of Stress	
RQ1a	How can we physiologically measure stress responses?	
RQ1b	How do physiological and subjective data indicating stress correlate and what are the implications?	3,4
RQ1c	What are the constraints of physiological sensing devices for interactive systems?	
RQ1d	How can we identify and summarize criteria representing the differences among physiology measuring hardware?	
	II Exploring Tactile Feedback for Stress States	
RQ2a	Which feedback modality suits best for notifying about stress states?	5
RQ2b	How to design an effective feedback stimulus?	
	III Designing Interventions to Reduce Stress	
RQ3a	What are means to manipulate stressors from an interaction design perspective?	6
RQ3b	What are the effects on the user when manipulating the stressor?	
RQ3c	What are the effects on self-reflection when manipulating stressors?	
RQ3d	How to include privacy when designing stress-aware interfaces?	

Table 1.1: Summary of research questions addressed in this thesis.

particular with respect to their ability to self-reflect **RQ3c**. Since the deployment of each of these scenarios covered a different aspect of privacy, the final research questions **RQ3d** addressed the challenge of including privacy in stress-aware interfaces what has become a particular issue considering the increasing distrust among technology users.

1.2 Research Methodology

For answering the research questions, the research conducted was carried out applying different methods. For the work presented in Chapter 3 a user study had been conducted in the laboratory. Hereby quantitative data has been assessed and inferential statistical methods have been applied to reveal significant differences in the subjective and physiological data. To investigate the limitations and derive criteria for physiology-aware systems (see Chapter 4), a large corpus of qualitative data has been collected through in-depth semi-structured expert interviews being analyzed applying an inductive thematic analysis [28] followed by a top-down analysis. While the subsequently presented two user studies described in Chapter 5 combined the assessment of quantitative data with obtaining qualitative user feedback being conducted in the laboratory setting, the three works evaluating the manipulation of stressors (see Chapter 6) relied on the in-the-wild approach. Applying a variety of research methods, such as focus groups (see Section 6.2.1), prototypical deployments (see Section 6.3.4), and a diary study (see Section 6.4.2) for the comparative exploration of the three explained approaches, considerable findings could be retrieved answering the relevant corresponding research questions.

As part of my research, different kinds of prototypes had been built or respectively implemented as a result of collaborations with colleagues or projects of undergraduate students. Accordingly, for the exploration of tactile feedback in relation to stress, three different hardware prototypes providing vibrotactile, pressures-based, and thermal feedback were built. While these low-fidelity prototypes were designed to serve their purposes and function without any inconveniences for the users, the implementation of the mobile phone and tablet applications targeted to represent high-fidelity prototype of a usable application. By providing them with fully functional features, users got genuine impressions, particularly when deploying them in real-world scenarios. Following the recommendations of human-centered design for interactive systems [75], the environmental settings where the interactive systems are intended to be used in have been taken into account. By focusing on the exploration of how the dedicated application could be made more usable, the prototypes have been designed by multi-disciplinary teams given the distinct research backgrounds, and users were actively involved, by, for example contextual inquiry in focus groups. The specifically addressed human-centered design approach has been applied, since hereby the human needs and demands are considered in the first place, while its emphasis lays on the validation of the design prototypical implementation. Accordingly, the user's actions are proposing the interaction space for the intersection between the human, the technology, as well as social factors, as illustrated in Figure 1.1.

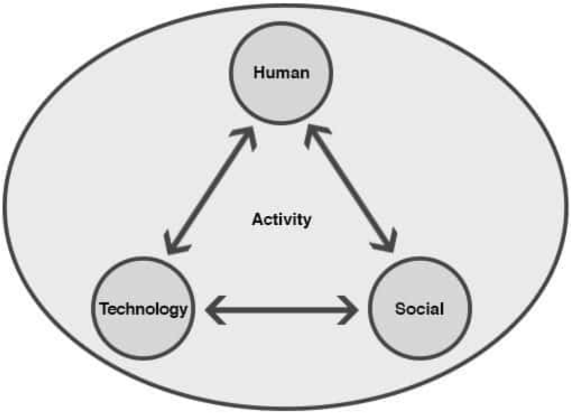


Figure 1.1: Illustration of the human-centered design perspective adapted from Winograd & Woods [284].

1.3 Research Context

The scientific work presented in this thesis has been carried out between 2016 and 2019 mostly at the University of Stuttgart in the Human-Computer Interaction group. During this period several collaborations with inspiring researchers from different fields took place and resulted in the presented research.

University of Stuttgart Among other work that has contributed to my scientific understanding and broadened my scope [139, 144, 206], the evaluation of tactile stimulation as a stress reliever [131, 205] has been carried out with the colleagues from the Human-Interaction Group at the University of Stuttgart. In this context, the collaboration with the Socio-Cognitive Systems group from the University of Stuttgart has resulted into valuable research combining interdisciplinary approaches, such as [240] and ongoing projects that are currently under submission.

Within several collaborations [204, 236] with Stefan Schneegaß, who shifted from a committed senior colleague to a Professor having his own Human-Computer Interaction Group at the University of Duisburg-Essen by now, I looked beyond the horizon of my own research and learned a lot regarding the supervision of larger user studies and students too.

Research Projects The research was partly carried out in the context of the following research projects: DFG funded cluster of Excellence Simulation Technologies (EXC2075), BMWi funded project motionEAP, EU Horizon 2020 ERC project AMPLIFY (683008), and the DFG funded collaborative research center SFB-TRR 161.

External Collaborations Together with Katrin Hänsel from the School of Electronic Engineering and Computer Science, Queen Mary University of London the reliability of physiological-sensing devices have been investigated in an extensive laboratory study [98]. Moreover, this collaboration built the starting point for valuable discussions on the limitations of such hardware and stimulated the idea to inquire user groups on their decision criteria when using physiology-aware technology.

Given the close connection to the Media Informatics Group at the Ludwig-Maximilians University in Munich, together with Mariam Hassib, Florian Alt and Christina Schneegaß I gained valuable discussion partners who supported me throughout my research and were happy to contribute to several joint publications.

Lastly, the collaboration with the Machine Learning and Data Analytics lab from the Friedrich-Alexander University in Erlangen-Nürnberg and particularly with Markus Wirth should be mentioned. Coming from different disciplines we together approached the challenge to implement a stressful task in virtual reality, which finally resulted into a successful publication [208].

As part of my research I had the possibility to visit Finland. During the UBI Summer School 2016 hosted by the Center for Ubiquitous Computing at the University of Oulu, I encountered Aku Visuri from the Center for Ubiquitous Computing and Olli Korhonen from the Empirical Software Engineering in Software, Systems and Services Group, both located at the University of Oulu and although we had distinct PhD topics and research interests, we managed to pool our expertise resulting in publications [207, 266]. My second stay took place at the end of my PhD career, when I visited the User Experience Design team of Jonna Häkkilä, as part of the Industrial Design Program at the University of Lapland. Not exclusively the openess to any crazy study, but rather the exactness with which my colleague Emmi Harjuniemi supported me throughout my visit made the collaboration unique.

1.4 Thesis Outline

This thesis consists of seven chapters supplemented by the bibliography and the appendix. While the first chapter introduces the topic and research questions, Chapter 2 provides the theoretical foundation and relevant work on stress, the four subsequent chapters present the results of ten research activities ranging from focus groups through experiments and real-world deployments up to user and diary studies. The concluding Chapter 7 presents the main findings of this work and answers the research questions ending with an outlook how future work could extend the presented research. Answering the ten research questions, this thesis is structured as follows:

Chapter 1 - Introduction The introductory chapter describes the relevance and the resulting motivation for researching stress responses in relation to digital technologies. Further, the research questions will be introduced and the research approach and methodology, as well as the research context in which this work has been carried out will be explained.

Chapter 2 - Background on Stress Responses In the second chapter, the theoretical background including theories and relevant work on previous and recent research embracing stress responses will be presented. Hereby, I focus on the theoretical foundations of stress, further addressing how stress responses can be detected using different measures. The sections lay the groundwork for my research and provide an overview over the most relevant work in the domain of research on the sources of stress.

Chapter 3 - Implications of Physiology-Aware Systems for Design Since reliable measurability is a basic prerequisite for developing and designing stress mitigating interventions, the laboratory study presented in Chapter 3 demonstrates that stress responses are physiologically measurable, which is validated by subjective data.

Chapter 4 - Clustering Limitations of Physiology-Aware Systems The fourth chapter is based on the findings considering implications of physiology-aware systems and thus, describes the investigation of their limitations from two different stakeholder perspectives. As a result the "*Design Space for Physiology-Aware Systems*" emerged, which summarizes and visualizes potential strengths and weaknesses of such hardware.

Chapter 5 - Exploring Tactile Feedback in Relation to Stress Envisioning the exploration of suitable stress mitigating techniques, the fifth chapter consists of two user studies which consecutively examine tactile feedback as a notifier about stress. While the first evaluation reveals that vibrotactile and pressure-based stimulation affect the users stress level similarly, thermal feedback is being regarded more favorable from a user perspective.

Chapter 6 - Manipulating Stressors in Interactive Systems Based on the findings from the first intervention exploration, I demonstrate how stressors can be manipulated in three different ways evaluating another approach to mitigate stress in Chapter 6. By examining the effects, particularly on the user him- or herself when (a) eliminating, (b) visualizing, and (c) self-adjusting sources of stress, I reveal that the pure preoccupation supports critical self-reflection even mutual consideration.

Chapter 7 - Conclusion and Outlook In the final chapter, a summary of the main research contributions will be provided by addressing the role of reflection and awareness. By answering the research questions posed in the beginning, the set of overall and concrete design recommendations will be introduced. In this context, the importance of privacy preservation is being highlighted when phrasing what should be considered when designing interactive systems. After acknowledging the limitations of the presented work, and outlook regarding future research challenges will be given.

Chapter 2

Background on Stress Responses

At large, there exist no single universally accepted definition of the complex construct of stress [97]. However, those situations that cause stress in organisms, and particularly humans have been examined in detailed and thus, in relevant literature there is a consensus on what such stressors can be, namely environmental conditions like extreme temperatures or noises, biological states like sleep deprivation, and cognitive events like time or social pressure, as well as prolonged work [23]. In contrast to these very precise examples of stressors, others referred broadly to stress as the "*imposition or perception of environmental or physical change, either negative [...] or positive [...], elicits a spectrum of physiologic changes that can be construed as adaptive to the organism*" [114] (p.78). Given the ambiguity among the various definitions of stress, this thesis relies on the common understanding of stressors as "*events that [...], challenge or threaten the wellbeing of an organism, increase its arousal or activation level, and deplete its resources*" [23] (p.23).

In the course of research on affective responses, it has been found that a particular state works in a different manner than emotions. Through his experiments Selye [242] could show that individuals react differently when being exposed to aversive stimuli and consequently, called the resulting response pattern *stress*. In the following, three representatives of the most relevant theories on stress will be presented, embracing the General Adaption Syndrom (GAS) [242], the concept of homeostasis [34], as well as the *Transactional Model of Stress and Coping* [152]. While the first two theories look at stress from a physiological perspective, the latter model introduces a more holistic view understanding stress as the interplay between an individual and the environment. The close connection between the perception of stressors and a change in body signals has been observed by various researchers at the beginning of the 20th century. However, the term "*fight-or-flight reaction*" has been introduced by Cannon [34] referring to the release of the hormone adrenaline which triggers the physiological responses described under "Physiological Measures" in Section 2.3.

2.1 Theoretical Foundation of Stress

One milestone in the field of stress research represents the work of Hans Selye. By his work "The Stress of Life" [242], he provided an extensive investigation of stress and was one of the first who mentioned the term *stress* and *stressors* before it became popular in the following years. Observing the biological changes in animal experiments as consequences of presented stimuli, his definition refers to so-called "*non-specifically caused changes*" [242] (p.64) which elicit a certain response pattern that he called stress. Following on this, Selye splitted the resulting process in three different phases including the the *alarm reaction*, what is visible in the individual's physiological signals; the *stage of resistance* showing signs that the individual has adapted to the particular stressful stimulus; finally the *stage of exhaustion* indicates that the individual cannot resist the aversive effect of the stressor and thus, the *alarm reaction*s' symptoms might repeat until the organism finally dies. This gradual biological program was named "*General Adaption Syndrom*" and corresponds to the concept of "*homeostasis*" [34], initially mentioned by Claude Bernard as the "*milieu intérieur*". By homeostasis it is meant the balanced state that is maintained and which faces continuous encounters with intrinsic or extrinsic challenging factors, such as stressors [42]. Since the each organism intuitively aims to be in this equilibrium, it is required to adapt to environmental changes permanently. This effort increases when the organism has to deal with stress because the recovery to go back into a homeostatic state might take enormous resources depending on the stressor's disturbing impact.

The fact that stressors can lead to positive or negative outcomes was addressed by Selye [244] first and will be explained when referring to the concept of "*eustress*" and "*distress*" in the following paragraph. Following on from this distinction,

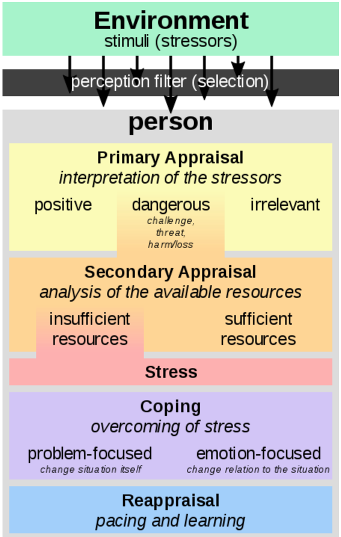


Figure 2.1: Illustration of the Transactional Model of Stress and Coping [152] adapted from Guttmann [a].

---

[a] ByPhilippGuttmann-Ownwork,CCBY-SA4.0,https://commons.wikimedia.org/w/index.php?curid=45616588

the so-called *Challenge-Hindrance Occupational Stress Model* by Cavanaugh et al. [36] assumes that stressors can be either perceived as challenges which trigger exceeding performances, or as hindrances that distract individuals from achieving their goals. In which of these two clusters a stressor is being allocated, depends on the individual's appraisal. In this context, the *Transactional Model of Stress and Coping* by Lazarus and Folkmann [152] serves as a foundation for Cavanaugh et al.'s model. This approach refers to psychological stress and focuses on the interdependence of appraisal and coping as depicted in Figure 2.1. Hereby, appraisal plays a key role since according to their assumption, individuals decide on their future actions based on their evaluation whether the perceived stimulus is threatening, challenging, or harmful and consequently requires coping. This first appraisal follows the second appraisal phase referring to the individual's estimation whether one's own resources and so-called *coping potential* are sufficient. Lastly, the individual approaches the final phase having to choose a suitable coping strategy, which is either problem-focused by adjusting

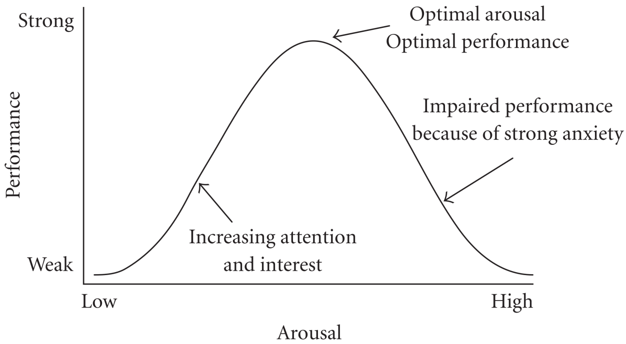


Figure 2.2: Simplified and commonly used illustration of the Yerkes-Dodson law [288] in the Hebbian version adapted from Diamond et al. [56]. Hereby, the lack of a negative impact on when performing simple task is neglected and therefore, simplifies the process.

the stressful stimulus or emotion-focused by re-considering one's situation. As an alternative to the coping, a reappraisal can be applied.

Eustress and Distress and the Yerkes-Dodson Law As part of Selye's research, he conceptualize the terms "*eustress*" and "*distress*" referring to the different positive and negative stress responses. Building upon his experiments, he inferred that the response to a stressor can either be perceived as an agreeable feeling, as it is for example, when athletes experience a high arousal shortly before a competition stars. On the other hand, whether an individual perceives as a stressor as disagreeable is also dependent on preliminary experiences and thus, external factors [243]. Accordingly, the distinction between these states is important, since as long as the individual is in control of the sensed stress, it's homeostasis is challenged which is still perceived to be exciting. In contrast, when the situations gets out of control or unpleasant feelings arise, the positive eustress can turn into an averse experience, so-called distress.

In this context, the Yerkes-Dodson law [288] is notable, since it describes with the illustration of an inverted U-shaped curve how an individual's performance relates to it's perceived arousal. As can be seen in Figure 2.2, the performance is best, when the individual is neither under- or over-aroused and thus, maintains a equilibrium between both states.

2.2 Stressors

Stressors can be divided into three categories including environmental, biological, and cognitive stressors. They differ in their appearance, meaning that the first refers to changes in the environment, such as extreme temperatures; by biological events it is meant that, for example an individual experiences sleep deprivation; and lastly, the cognitive component targets process like working or the experience of averse emotions, which require cognitive processing [23]. Moreover, Herman and Cullinan [114] distinguish between "*systematic*" and "*processive*" stressors. While the former refers to situations that represent an actual physiological threat, for example injuries or dehydration, the latter describes those circumstances that are perceived as potentially threatening. This distinction has been adapted from experiments with animals, but particularly processive stressors can be observed in humans and require the ability to process information using the limbic system in the forebrain. Those can include severe life events, such as loss of home or close relatives, but further anxiety while speaking in front of others or even cognitively demanding situations that are associated with performance evaluations, for instance mental arithmetic tasks or the Stroop color-word inference test can be regarded as processive stressors [23].

2.3 Measuring Stress Responses

Stress can be assessed differently relying on physiological, subjective, hormone, and behavioral measured as will be explained subsequently.

Physiological Measures Based on reactions of the Sympathetic Nervous System (SNS)- a part of the Autonomous Nervous System (ANS), for example sweat glands are activated, the heart beat increases, and heart beat intervals decrease [53, 252]. As one of the mostly noticable states, humans feel responses to stressors. While a stressor can be external, such as noise or also someone shouting at oneself, the internal stressor emerges from an individual's inside. As a consequence stress is perceived and this again is signified by body signals being triggered by the SNS, such sweating or gasping. Since the human body is a complicated and continuously working system, these changes can be picked up by physiological sensors to make predictions about stress or affective states. Thus, stress can be measured in different ways.

For example Electroencephalography (EEG), Electromyography (EMG), Electrooculography (EOG), Skin Temperature (ST), Electrodermal activity

(EDA), Heart Rate (HR), Heart Rate Variability (HRV), and Blood Volume Pulse (BVP) can be recorded to observe the physiological reactions towards stress. While the gold standard for heart rate assessment is the Electrocardiography (ECG), modern approaches try to use optical sensors to derive the heart's beating rate based on the blood flow under the skin - a technology called Photoplethysmography (PPG).

Subjective Measures Before non-invasive sensing was technologically feasible, research often relied on subjective self-reporting techniques. In practice, scales, such as the Relative Stress Scale [89], Perceived Stress Questionnaire [46], and the Stress State Scale [172] have been applied. Further, the non-verbal assessment of affective states, such as the Self-Assessment Manikin Scale (SAM) [26], serves as a valid tool to capture information about person's valence, arousal, and dominance.

While psychologists still often stick to either observations or self-reports, physiological sensing as well as measuring certain hormone levels, such as cortisol or noradrenalin has become a popular measurement alternative because physiological measurements deliver reliable results under specific conditions [11, 90, 119]. Moreover, a combination of these measurement techniques have been used following a multidimensional approach. Prominent research frequently uses self-reports as a comparative instrument when obtaining hormone levels or recording physiological signals [5, 128].

Hormone Measures Another method to assess stress is to measure the amount of stress-corresponding hormones such as cortisol and adrenaline in saliva or blood [1] also being released when the SNS is activated. Particularly in biomedical and neuroscience disciplines, many researchers investigate the hormone level in stressful situations to gain important insights. Since cortisol can be assessed in saliva, it is even a suitable method to determine stress levels in infants and toddlers non-invasively [138]. Measuring hormone levels is however not straight forward, as it is also dependent on other contextual factors, such as caffein intake. Moreover, hormonal changes trigger an increase in heart rate and a change of the heart's beating patterns which is why they are often often applied equivalently to each other.

Behavioral Measures Stress can be further seen in an individual's behavior as a consequence of the neuro-physiological processes leading to stress states [41]. Hereby, external observers can record the behavior among others to, for example, identify anomalies in sleeping or eating habits, as well as extensive alcohol or cigarette consumption relating those to acute stress.

2.4 Summary and Research Approach

In summary, the presented background on stress responses provides an overview on the construct of stress and highlights the disunity regarding an universal definition; besides, it has been clarified on which understanding of stressors this thesis is built upon. Moreover, the most relevant concepts in the field stress research including the explanation of the difference between eustress and distress have been addressed. Further, the key element of a stressor is described, before the four main measures assessing stress are briefly characterized.

Facing these different measures it has become obvious that stress can be sensed using various measures and means. Thus, related literature suggests to understand stress as construct being shaped by relations or so-called "transactions" embracing an individual and its environmental factors [151]. In reference to the *Transactional Model of Stress and Coping* by Lazarus and Folkmann [152] (see Section 2.1 for a detailed explanation) this thesis places its interventions at the layer of primary appraisal allowing to interpret stressors. While perceiving stressors as sources of stress, the presented approach of manipulating stressors leads to reflecting about those and consequently facilitates to cope or even reappraise with such sources of stress. Thus, the present work shows how intervention techniques can be used to circumvent the secondary appraisal layer which leads to stress, but rather supports the individual to either manipulate the stressful stimulus, re-consider one's view of the stressor, or to interpret stressors differently jumping directly to the final phase.

Chapter 3

Implications of Physiology-Aware Systems for Design

The connection between the mind and the body has been explored for a long time. In the beginning the vital functionalities have been investigated mainly in animal experiments envisioning to reveal hidden signs of affective responses. Research in medicine, physiology, psychology, and related disciplines have come a long way in the past years since more light was shed on the relevance of physiology-based data in connection to stress responses. Particularly the work by Hans Selye [242] exploring the nature of stress and Walter Cannon's [34] concept of the "*homeostasis*" have marked a central point in stress research and thus, have contributed immensely to the investigation of stress throughout the past century.

While performing his research, Selye, among others, recognized through his experiments the significant role of stress hormones activating the Hypothalamic-Pituitary-Adrenal Axis (HPA Axis) system [244]. This complex apparatus controls stress responses and triggers specific body reactions, such as digestion, the immune system, and emotional responses. Accordingly, the importance of the ANS has been revealed. Among its three subordinated systems, the enteric, the parasympathetic, and the sympathetic, the two latter ones act as antagonists (please refer Section 2.3 for a more detailed description). Knowing how these systems work together and reflect the human stress response towards threats, or more generic stressors, has paved the way for the domain of affective computing. When Rosalind Picard in her book "Affective Computing" [197] first mentioned her vision how human-computer interaction could be revolutionized if computers would sense and react according to its user's affective states, the challenge of integrating emotion-aware interactive systems was born. As a fundamental prerequisite for the implementation of such user-understanding technology, the recognition of affective states must be reliably [11, 292].

Consequently, this chapter focuses on the examination of physiological data in conjunction with its validation through subjectively assessed data and what implications referring to **RQ1b** and constraints this has for the design of stress-aware interactive systems answering **RQ1c**. As a foundation for the understanding of relevance of physiological in human-computer interaction research, the presented literature will illustrate first how stress responses can be measured physiologically and thereby respond to **RQ1a**.

This chapter is based on the following publication:

- K. Hänsel, R. Poguntke, H. Haddadi, A. Alomainy, and A. Schmidt. What to put on the user: Sensing technologies for studies and physiology aware systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 145:1–145:14, New York, NY, USA, 2018. ACM

3.1 Related Work

The related work embraces the topic of stress recognition with the help of physiological parameters. Subsequently, the data collection and processing of physiological signals is briefly summarized highlighting variations in signal quality which lead to differences among sensing hardware.

Physiological Parameters Signifying Stress As already described in subsection 2.3 there are physiological parameters that reflect if the human body is aroused or not. Among other studies [81, 148, 181], Quesada et al. [215] found that their participants responded to stress inducements with a an increase in their arousal valence using the SAM visual rating scale [26]. Winsky-Sommerer, Boutrel, and de Lecea [285] proposed that the stress-triggered biochemical reaction in cortical structures affects the arousal and further the amygdala, being responsible for anxiety. Consequently, arousal is understood as a physiological reaction towards stressors. Among the mentioned signal types heart rate in conjunction with heart rate variability is regarded to deliver the most reliable indication of stress, since prior work indicated that it correlates significantly with the emotional state [57, 224]. It has been used as a stress indicator in a variety of disciplines such as medicine [95], psychology [71, 238], and HCI [175]. However, it is prone to physical activity and cumbersome to record from an user perspective [120]. More recent approaches use optical sensors, PPG for inferring heart beats from the blood flow beneath the skin [90]. This technology bears the advantage that sensors can be placed at various positions at the body, such as earlobes, fingertips or, as commonly used in fitness trackers and smartwatches, at the wrist. This has been used for example in monitoring th heart rate under physical movement [294]. Apart from it's sufficient reliability, another advantage of the ECG signal is its richness of features, since it portraits the behaviour of the heart in detail. The signal cannot only be used to extract features such as heart rate, but also the variation of inter-beat intervalls - namely HRV. The use of HRV features for detecting stress is grounded in the effect of the sympathetic and parasympathetic nervous system on the beat beat patterns [248]. An increased variation in time between consecutive heart beats (also called inter-beat or R-R intervals) is hereby related to "cheerfullness and calmness" [80]. On the contrary, a low heart rate variability also has been associated with an incresed mortality risk, diseases [260], and decreased emotional-regulation [273] and stress [57]. Accordingly, various measures consider the data in time- or frequency domain. The most commonly used are the Root Mean Square of Successive Differences (rMSSD), the Standard Deviation of Normal Sinus Beats (SDNN), and the High Frequency/Low Frequency Ratio (HF/LF Ratio) [259]. A further stress related feature is EDA, which is mostly referred to as the activation of sweat glands; additional names include Galvanic Skin Response (GSR) or skin conductivity [54]. To prepare for the potential increased physical extortion due to a 'fight or flight' reaction, a stress response is accompanied with an increase in sweat production. This increased moisture of the skin can be easily picked up by sending a small current through the skin and measuring the resistance. EDA can be found in prior work as an indicator for cognitive load [246], stress [116] and also as a "predictor of emotional responses to stressful life events" [182]. With the increase in sweat production, there also come changes in the surface temperature of the skin due to the evaporative cooling; skin temperature showed a good prediction ability indicating stress through a drop in surface skin temperature [129, 265].

Reflecting on the accuracy of physiological signals serving as stress markers, it is still arguable in how single measures provide enough data richness to claim sufficiently reliable results [5, 120, 128, 224]. It has been broadly discussed that physiological parameters, and particularly heart activity are prone to confounding variables, such as caffeine intake or overall physical fitness. Therefore, some researchers suggested the comprehensive assessment of stress, such as Kye et al. [147] who developed a 'Multimodal Data Collection Framework for Mental Stress Monitoring' using four different sensors (Empatica E4, LG watch style, Zephyr Bioharness, IP camera) due to the challenge of getting sufficiently reliable data. To sum up, for the successful implementation of physiology-based stress-aware systems, the data validation and related implications for its usage must be considered in the design.

Data Collection and Processing of Physiological Signals When speaking of physiological data collection, various aspects have to be considered. For example, the richness of the sensors used is decisive when planning a research project. For researchers and consumers alike, the way how their data can be assessed also plays an important role. While many fitness trackers provide an additional smartphone application to monitor the activity i.e. in daily or weekly visualizations, experimenters are more concerned with the given preprocessing in many sensing devices, such as Fitbit[2]. Particularly the choice of sampling rates for assessing distinct parameters can differ according to its application scenario. While for measuring body movements in humans the activity is contained within sampling rates below 20 Hertz (Hz) [8, 17, 124, 127, 203], previous research took frequencies between 7 Hz and 50 Hz, but also 200 Hz for recognizing activities of daily living using accelerometers [132]. Since the question of which frequency to take is often raised in empirical studies, researchers should be aware of the trade-off between high sampling rates resulting in more data points allowing better interpretation of the data, but requiring more computational effort affecting, e.g. the size or costs for hardware on the other hand. While exploring the question which frequencies influence classifiers, Khusainov et al. [132] found that sampling frequencies between 10 Hz and 20 Hz improve activities of daily living classification, which was confirmed by Bouten et al. [24].

Another relevant issue when elaborating on physiological sensing, is the data processing part. It can be understood as a process comprising different sub-processes which can differ among distinct sensing devices. Often the following exemplary methodology is being followed when dealing with physiological data. In the pre-processing phase data is filtered and cleaned for

[2] `fitbit.com`

example with the help of Wiener filters. Also the Principle Component Analysis have been ran to remove irrelevant features from data recordings, particularly to model stress based on EEG data [55]. Another noise reduction technique that is commonly used for EEG signals alike, is the Independent Component Analysis [118]. The data segmentation aims to extract and classify features by using different techniques, e.g. the Fast Fourier Transformations. In practice, Sharma and Gedeon [249] applied this practice or the Wavelet transformations to physiological signals. Then, features need to be selected; for this the domain of the data set and the targeted discriminatory features should be known. The latter becomes even more important for the feature selection depending on the "discriminatory qualities of the features" [132]. Finally, events can be classified. This has been done in a vast amount of studies that have been performed to classify or model stress, features extraction techniques. For example, Han et al. [96] found that the combination of Random Forest and Support Vector Machine algorithms provides the best performance regarding the accuracy of differentiating between three different states (no stress, moderate stress and high stress). Facing the variety of analysis methods, still the question remains on how the data being used for this computations is preprocessed by the sensing technology. Likewise, the data resolution obviously has an impact on the analysis technique and therefore needs to be considered as a decision criteria when choosing a recording device, what will be discussed more detailed in Chapter 4.

A core requirement when aiming to design stress mitigating applications, is to know when users are stressed. Since from a biochemical perspective stress is triggered by arousal state [285], self-assessed arousal ratings were used as a ground truth measurement in the following study. The work presented in this chapter serves as a foundation for enabling the implementation of stress-aware interventions in interactive systems and further holds a proof-of-concept for the assumptions made that have been mentioned in prior work as for example referenced above.

3.2 Measuring Physiological Responses towards Stress

This chapter provides two main contributions: on the one hand, it is shown that physiological data and subjectively assessed self-reports correlate significantly what can be understood as a proof-of-concept that stress can be detected using physiological sensing. On the other hand, the results reveal considerable differences mong the sensing devices. Taking the measurability of stress facilitates the design of stress-aware interfaces as a foundation for the presented exploration of interventions for interactive systems, more importantly is to be aware of the detected inconsistency of the devices due to its implications for building stress-aware interactive systems.

3.2.1 User Study

In the user study different psychological measures sensed by three distinct sensing technologies have been recorded. As a ground truth the subjectively perceived stress level was assessed. In the following, there will be described the study design, the independent and dependent variables, the apparatus used, as well as the sample and procedure.

Study Design Each participant was asked to take part in each of our four conditions lasting 20 minutes in total (5 minutes per condition). In this within-subject design the sequence of conditions was partly randomized following the Latin square distribution but started with the non-stressful- neither mentally, not physically- baseline task. Another constrain was the alternation of mentally and physically stressful tasks to prevent users from exceeding their physical limits. Three independent variables were used, namely physical activity and mental stress which consisted of two levels each; the triad was supplemented by four different sensing devices recording physiological data throughout the distinct mental and physical strain. The factorial design resulted in four overall conditions which have been performed by all participants. The setup of conditions was inspired by Sun et al. [257] and is illustrated in Figure 3.1.

Variables As independent variables the sensing devices were altered and the amount of physical activity, as well as the level of stress was varied as described subsequently. The physiology-recording hardware was chosen according to their representational function. Hence, the Apple Watch, as a popular smartwatch with fitness capabilities in form of physical activity and heart rate tracking and validated on its performance regarding heart rate errors and correlation with the gold standard device [60, 251, 268] was taken. The second consumer device, the Microsoft Band 2 fitness tracker also relies on optical heart rate recording and is one of the few consumer wearables incorporating skin conductance and skin temperature sensors. To take up another sensing technology, the Polar H7 chest strap as an exemplary device has been used in the study. Being worn by many atheletes and in sports settings due to its convincing accuracy [82], it further bears the advantage to share sensing data via Bluetooth and send it a mobile phone. Lastly, the Nexus-10 MK2 by Mind Media, a laboratory measurement instrument has been used. Mostly targeted for biofeedback applications and psychological research it provides the medical standard and was taken as a reference measure.

Physical Activity To observe the devices' performance in a realistic scenario, the amount of physical activity was designed to be another independent variable. The participants were asked to either walk on a treadmill in their own, physiologically demanding pace resulting in the *walking* condition or to remain stationary on a comfortable chair accordingly being in the *stationary* condition.

Stress Amount The amount of stress perceived by the participants was varied too. For this the participants were asked to perform mental arithmetic tasks (MAT) resulting in the *stressful* condition and, to relax while listening to meditation music being in the *non-stressful* condition.

Sensing Devices Two wrist-based consumer wearables with optical heart rate technologies (the Apple Watch Series 2 and the Microsoft Band 2[3]), one chest strap heart rate monitor (Polar H7 chest strap[4]), and a laboratory measurement tool with ECG adhesive electrodes, namely the Nexus 10 kit[5] were used simultaneously to record elctrodermal activity, skin temperature, and heart rate activity.

Physiological Data From the aforementioned measurement devices heart rate, electrodermal activity, and skin temperature was recorded. There is a lot of previous work providing evidence that these physiological parameters have been shown to be reliable stress indicators [3, 129, 246].

Choice of Physiological Stress-indicating Measures Several studies used physiological measures, i.e. heart rate, electrodermal activity, and skin temperature to detect stress and showed correlation with subjective stress responses [231]. Moreover, the combination of these measures has proved to be a reliable indicator in, e.g. psychology [3, 7], for the development of a non-invasive real-time stress tracking system [159], in a real-world driving tasks to determine the driver's stress level [107], or for non-invasive stress detection in HCI [12].

---

3 https://www.microsoft.com/en-us/band

4 https://support.polar.com/e_manuals/H7_Heart_Rate_Sensor/Polar_H7_Heart_Rate_Sensor_accessory_manual_English__.pdf

5 www.mindmedia.info/CMS2014/en/products/systems/nexus-10-mkii

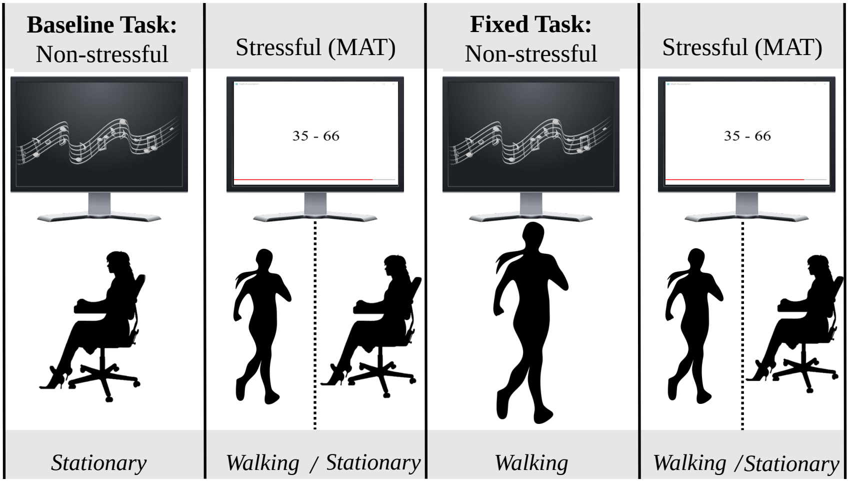


Figure 3.1: Study design depicting the different tasks, as well as the sequence of trials. Participants were asked to perform four trials each, comprising a baseline task in the beginning and a given task between the two alternating trials. For the second and the last task, they were asked to perform mathematical operations while either walking or sitting calmly, alternated according to an counterbalanced order.

Self-reported Arousal, Valence, and Dominance To collect data on the participants' stress level, they were asked to fill in the widely-used SAM [26]. This self-report measurement tool allows the non-verbal assessment of current affective state, respectively valence (pleasure), arousal and dominance, through pictures. A 9-point rating scale for each dimension according to Bradley's and Lang's original work [26] was utilized. Hereby the participants were instructed to place an 'x' on any of the five figures or between two figures.

This first classical arousal model, similar to Russell's Circumplex Model of Affect [227], captures tension as one dimension and hence, does not allow differentiations. Thayer [260] suggested that arousal can be further splitted into energetic arousal (ranging from wide-awake to tired) and tense arousal (nervous to calm). This fine graded characterization was taken up and Schimmack's and Grob's [234] recommendation to add two additional questions was followed for the study: a 5-point self-rating Likert-item for each dimension assessing tension and wakefulness [123, 222] was added.

Apparatus The stimulus material consisted of Mental Arithmetic Tasks (MAT) which had been proven to elicit stress in participants [18] and also influence physiological measures [104, 160, 245, 262]. To ensure that the physiologically observable stress was also subjectively perceived by the participants, a visual countdown and feedback (both visual and auditory) was added. Setz et al. [246] showed that this is a valid method to further increase stress. Feedback was always given after the participant's response. While for correct answers the screen turned green and the word "Correct" was displayed, false answers or time outs were signified with either "False" or "Time out" being shown on a red screen and a buzz sound. For each of the calculations, addition and subtraction of two-digit numbers ranging from 0-100 and including negative solutions, the participants were given six seconds each. A timer representing the time left for each task at the bottom of the display was provided. According to the study design, the relaxing task consisted of listening to meditation music[6]. These tasks were presented on a 60-inch display being positioned on a desk in front of the participants. The study setup was inspired by Vlemincx et al. [267]. Regarding the performance of physical activity, participants were asked to walk for five minutes on a treadmill (model: ProFitness Sierra motorized) for the walking task.

Participants and Procedure In total, 24 participants for the study were acquired via university mailing lists, leaflets and personal recruitment campaigns. For the final data analysis, one pilot participant who had volunteered for a test trial of our laboratory setup and two other participants due to technical problems during the data collection were excluded . Hence, 21 participants with a mean age of 28.9 years ($SD = 4.5$) remained; among them were eight females and 13 males. Before the recruitment, strict exclusion criteria were set up, namely no participants being taken who had been diagnosed with any heart conditions, mental illnesses or learning disabilities. Participants were also asked to guarantee that they did not suffer from alcohol and/or drug addiction. Before the experiment started, they were asked to refrain three hours from caffeine. For compensation each participant received 15 British pounds. Prior to the study, the participants were introduced to the experiment environment and briefly informed about the study background as well as the sensor placement on the body. They were asked to sign the consent form and to fill in an initial assessment consisting of demographic questions, self-reported fitness level, and smoking behavior as inquired in Weitkunat et al. [271].

After they were given a short introduction to the treadmill and the mental arithmetic task was explained, the participants were asked to put on the chest-worn

[6] As meditation music the song number 14 from the album '72 Ambient Meditations' was used

ECG sensors (Nexus 10 ECG with pre-gelled, disposable electrodes) and the Polar H7 chest belt. Hereby a sheet with visual material from the manufacturers and instructions on the correct sensor placement was provided to ensure proper sensor fit was provided. Those illustrations where discussed with the participants, before they proceeded. The sensor was attached with medical tape on the forearm because it should be comparable to the values provided by the Microsoft Band and therefore was attached close to it's location. Instructions were provided how to place the Polar H7 chest belt around the chest and close the clasp after moistening the plastic parts at the inside of the strap. Participants were also instructed to already put the cables for the skin temperature and skin conductance through the sleeve of their right arm, to prepare for the following steps. The participants were given privacy behind a screen to perform these actions. The correct placement of the sensors and the correctness of the ECG signal was later checked on the experimenter's computer. After the chest sensors were placed, the experimenter placed the skin conductance sensors on the fingers and attached the skin temperature sensor on the participant's forearm using medical tape. Next, the participants were helped to place the wrist-worn devices (Apple Watch and Microsoft Band 2). Before the study started, the correct data transmission and validity of the signals for all sensors was initially checked. Each procedure starting with the recording of the baseline. In this condition all participants were asked to remain seated for five minutes listening to meditation music via wireless headphones. Then, each participant was assigned the following conditions in counterbalanced order, alternating between *walking* and *stationary* while mental arithmetic tasks should be performed. Between the two stress-inducing trials, there was a fixed task requiring to walk while listening to meditation music via wireless headphones. This sequence of conditions is illustrated in Figure 3.1. To assess our dependent variables, the SAM questionnaire including single-items on wake/tense arousal and perceived stressfulness of the task after each trial (including the baseline) was administered. The entire procedure took approximately 1.5 hours in total.

3.2.2 Results

For the data analysis, a period of four minutes per each condition was taken. Thus, the first 50 and last 10 seconds were excluded due to novelty effects and, respectively because the recording did not exactly stop when the task ended but continued for a few seconds. Furthermore, the Microsoft Band's provided skin

[6] `https://support.polar.com` and `https://www.mindmedia.com/en/products/sensors/`

	Stationary		Walking	
	Non-Stressful Mean (SD)	Stressful Mean (SD)	Non-Stressful Mean (SD)	Stressful Mean (SD)
Arousal	-2.76 (1.14)	0 (0)	-1.76 (1.64)	0.81 (1.00)
Tense Arousal	0.71 (1.01)	1.52 (1.17)	0.86 (0.73)	1.71 (1.19)
Wake Arousal	2.29 (1.06)	3.53 (0.93)	3.05 (0.97)	3.86 (0.79)
Stress	0.24 (0.54)	2.33 (1.39)	0.86 (0.96)	2.57 (1.4)
Valence	1.19 (1.47)	0.95 (1.69)	0.9 (1.14)	1.19 (1.63)
Dominance	0.48 (1.60)	-0.38 (2.31)	0.67 (1.32)	0.0 (2.32)

Table 3.1: Results (means and standard deviations in parantheses) for the subjectively assessed measures divided into stationary and walking conditions.

resistance (R) measures (*kohms*) were converted aiming to use the same unit as the skin conductivity (G) provided by the Nexus device ($micro - mho$). For this the following formula had been applied: $G = \frac{1}{R} * 1000$.

To reveal differences among the subjectively assessed data, arousal, stress and valence scores assessed using the SAM scale for both conditions, the *stressful* and *non-stressful* were compared. A Friedman test and a post-hoc Wilcoxon Signed-Rank tests for the subjective data comparing the *stressful* and *non-stressful* within the same physical activity level, namely *walking* and *stationary* were performed. The Friedman test revealed that there are significant differences among the four conditions for arousal ($\chi^2 = 39.440, p = .000$), perceived stress ($\chi^2 = 44.305, p = .000$), wake arousal ($\chi^2 = 33.307, p = .000$), and tense arousal ($\chi^2 = 14.646, p = .002$). There were no significant differences for dominance ($\chi^2 = 2.733, p = .435$) and valence ($\chi^2 = 1.672, p = .643$). A post-hoc Wilcoxon Signed-Rank test indicates that arousal, wake arousal, tense arousal, and perceived stress were statistically significantly higher in the *stressful* condition than in the *non-stressful* condition when the participant experienced *walking* and *stationary* activities alike. The descriptive statistics of the subjective measures are summarized in Table 3.1 according to the non-stressful and stressful conditions.

Additionally, the physiological data recorded with the different measurement hardware during the stressful and non-stressful conditions had been analyzed by performing two Wilcoxon Signed-Rank tests for both conditions under the same physical activity, namely walking and stationary. For the two planned comparisons, a Bonferroni correction on $\alpha = 0.05$ was applied, which resulted in $\alpha/2 = .025$. While the Nexus reference device showed a significant increase in heart rate in the stressful condition while participants were relaxed ($Z = -2.381$, $p = 0.017$), this change was not detected by any of the other heart rate monitors. Both the Nexus and Microsoft Band revealed differences in EDA while participants were undergoing the stationary activity ($Z = -3.285$, $p = 0.001$ and $Z = -3.058$, $p = 0.002$). Again in the stationary condition, a change in skin temperature was solely registered by the Nexus device ($Z = -2.416$, $p = 0.016$). Thus, only the Nexus kit was able to detect changes in HR under stationary activity. Variations in EDA during the stationary condition were noticed by the Nexus and the Microsoft Band, while skin temperature changes were again only visible in the Nexus data.

3.3 Discussion

From the findings of the user study, it was confirmed that physiological parameters can serve as indicators for stress responses which is in line with the observations in previous studies [66, 292]. Moreover, the subjectively assessed arousal, wake arousal, tense arousal, and perceived stress were statistically significantly higher in the *stressful* condition than in the *non-stressful* condition what corresponded to the physiological values obtained in both conditions. Hereby again, the results support prior findings that subjective data correlates with physiologically detectable stress markers [107, 175]. What marked the initial point for further research, is the observation of crucial differences in the reliability of the used sensing technologies. The Nexus reference device was the only measurement tool showing a significant increase in heart rate activity, as well as in skin temperature during the *stressful* condition. All other devices performed poorly when they had to record accurate data under physical activity circumstances. Additionally, only the Microsoft Band further than the Nexus device revealed differences in EDA under stress.

Facing such differences among the used measurement devices, it becomes obvious that there are huge shortcomings for the single technologies. Not being aware of a technology's own advantages and disadvantages when choosing one for a research project can lead to fatal measurement errors. However, this issue is not entirely new. There have been multiple evaluative studies on the validity of physical activity trackers, or monitors as it is sometimes referred to [20, 58, 62, 68, 143, 223, 256]. For example Nelson et al. [186] compared them for "specific activity types" such as three sedentary, four household, and four ambulatory/exercise activities. Nam et al. [184] analyzed the activity monitors' validity in "semi-structured activities", namely steps taken, sleep and energy expenditure . Often researchers also focus on specific consumer groups. For example, whereas Reid et al. [218] and Farnell et al. [65] investigated the Fitbit performance in female adults, Chiauzzi et al. [39] as well as Floegel et al. [73] focused on consumers with physical disabilities. Further, there can be studies found where homemade hardware is compared. For example, Patterson et al. [195] validated their electronic accelerometer-based physical activity device (Actigraph) in a comparative study among different physical and mental stress, but they did not use a variety of physiological parameters and sensing technologies, which is why it is difficult to interrelate the results. Khusainov et al. [132] had a more holistic method and mentioned different monitoring device platforms and sensor types in their survey paper. The only comparative approach they present was about effect of the sampling frequency during the data collection process and how it can be combined best with machine learning. They further provide an overview on existing research studies categorized by their activities of daily living, e.g. selfcare, posture, and communication and accordingly they included 24 different sensor types used for research activities, among them twenty wearable ones, in their survey [132]. Another comparative study was conducted by Adams et al. [1]. In an in-the-wild study they collated sensor data from their participants' mobile phones, and data recorded by an Affectiva Q sensor [209], as well as subjectively assessed data on their stress level measured via Experience Sampling Method (ESM). As a result, they derived different contexts in which each method could be applied best but concluded that the Affectiva Q sensor lacks robustness. They found that both, the self-reported values and EDA data assessed by the Affectiva Q, correspond to each other depending on the threshold set for the EDA signals. Nevertheless the subjective measures showed the best reliability to assess low and moderate stress. Nevertheless, none of the studies mentioned compares wearable sensors different in their sensing technologies extensively under different psychological and physiological conditions and derives explicit properties and dimensions reflecting differences.

The lack of evaluative studies also becomes obvious in related topics such as the comparison of different computational techniques to model stress. One of the few projects which realized this was "Stress Recognition Using Non-invasive Technology" by Zhai and Barreto [293]. They used normalized data as input values to compare the Naive Bayes classifier, Decision Tree classifier, and Support Vector Machines against each other aiming to be trained such, that unknown affective states could be predicted. The latter one turned out to have the best accuracy rate of 90.1% followed by the Decision Tree classifier (88.02%) and the Naive Bayes classifier with only 78.65% accurate recognitions. Also Tudor-Locke and Myers [264] point out in that there is a lack of methodological considerations for quantifying physiological activity- not only in research and Fletcher, Poh and Eydgahi [72] highlight the "*lack of mobile health data standards*" (p.5) and the "*validation and performance standards*" (p.5) referring to missing comparability and insufficient measures for performance. Moreover, Hao and Foster [101] present a large review on physiological responses recording wireless body networks focusing on their applications, technologies used, and the associated challenges in measuring physiological signals. Among these, there are named reliability, portability, privacy and security, and energy-aware communication [101, 219] which exist as crucial factors for each physiology sensing technology.

Against this background, in Chapter 4 I present a detailed investigation of the shortcomings in sensing hardware addressed in the present chapter.

Implications and Constraints for the Design of Interactive Systems
The differences between the used sensing devices are important to consider before choosing such a device. Since this depends also on the context of usage, it should be clear if the interactive system which relies on the physiologically obtained values is going to be used in mobile or stationary contexts. If the user is free while interacting with the system, measurements should be double-checked using subjective data because most sensing devices, particularly consumer ware is affected by movements. Instead of prompting the user with questionnaires, shorter self-assessments following the example of the ESM [47, 109] could be incorporated in the interaction process with the system. Apart from the mobility, the perceived pleasantness while "wearing" the sensing device is important for designing an interactive system. Since most applications are aiming to fulfill the requirements of a user-centered design, the attachment of cumbersome electrodes is undesirable in practice. When speaking of the comfort to apply physiological sensing technology, the final aim of the application or system should be considered ultimately. If an interactive technology is able to successfully reduce the perceived stress in users, then it's practitioners are surely willing to accept a less comfortable sensing device. In conclusion, the benefit of an efficient technique would make specific shortcomings more tolerable for users. Therefore, when in the design of such techniques the overall aim including possible deficiencies should be considered to be able to estimate it's worth from an user-perspective.

3.4 Chapter Summary and Conclusion

In the results from the user study it was shown that stress responses are visible in physiological parameters, namely heart rate values, electrodermal activity, and skin temperature. Since subjective data was also collected from each participant during the study, it indicated that physiological data reflects the stress level users reported. In summary, participants perceived higher arousal, tense and wake during the trials where stress applying mental arithmetic tasks was induced. Likewise, they rated these conditions to be more stressful as indicated by ratings on the perceived stress. What the user study further revealed, was that physiological sensing hardware underlies some specific constraints. From the comparison of sensor data under different amounts of physical activity, changes in the physiological values were observed. The Nexus device was the only sensing hardware that was able to detect even fine-grained changes in the physiological data assessed.

In this chapter the physiological measurement of stress responses have been examined answering three main research questions. First, the fundamentals of how physiological markers are related to stress have been explained on the basis of related work answering **RQ1a**. Correspondingly, the performed user study has shown that heart rate activity, electrodermal activity, and skin temperature significantly correlate with the perceived stress, what confirms the validity of measures used in prior work. Further, this observation did not only contribute to answering **RQ1a** but also to **RQ1b** signifying that subjectively assessed data correlate with physiological values obtained by sensing hardware. The findings from the user study further strengthened the assumption that for the subjective self-reports on stress can be found physiological correlates. In this context, the feasibility of non-obtrusive stress measurements can be appreciated. It allows researchers to gain data on their participants stress level without self-assessments. Additionally, stress measurement can take place apart from laboratories but also in field studies using mobile sensing hardware. Following on this, the performed research also revealed constraints of physiological sensing devices for the usage in interactive systems. The results revealed that not each device was able to detect such changes referring to **RQ1c**. Hereby, particularly the accuracy of the recordings, as well as the mobility and the comfort of attachment turned out to be important criteria that need to be considered when implementing stress-aware applications in interactive systems.

Consequently, the choice of sensing hardware and further the context of sensing were found to be crucial for the detection of changes in physiological data.

Knowing this, stress reducing applications incorporated in interactive systems should consider the hardware reliability, as well as the confounding variables when obtaining stress data with physiological sensing hardware. As suggested in the implications mentioned previously, designers should check the users' stress level while using their applications preferably with assessing subjective data given the high inaccuracy in mobile sensing contexts. By answering the named research questions, the foundation is laid for investigating suitable techniques incorporated in interactive technologies to reduce stress. Nevertheless, the implications stated can have a severe impact on the realization of stress mitigating interactive systems. Therefore, the following chapter will concentrate on the investigation of the revealed constraints and consequently a design space will be introduced summarizing considerable decision criteria for sensing hardware choices from the researcher and end-consumer perspective answering **RQ1d**.

Chapter 4

Clustering Limitations of Physiology-Aware Systems

In Chapter 3 it has been pointed out how physiological measurements can be used to detect stress based on their correlation with subjectively assessed data. Moreover, the limitations in physiological measurement hardware have been revealed resulting into certain constraints in such measurements. Based on these findings implications for the design of interactive systems have been derived addressing the challenge of choosing suitable physiological sensing devices. This chapter presents a detailed investigation of the potential shortcomings in physiological measurement devices from a practitioner perspective. Hereby, consumers and researchers, being named experts in the following, were inquired and involved in the development of the *Design Space for Physiological Measurement Tools*. This tool provides a summary of the distinct aspect being worth to consider when designing interactive systems that incorporate physiological sensing.

Throughout the past decades sensing physiological signals has evolved from a high-end technological challenge into a standard functionality of every-day consumer ware. The global mobile communications analyst CCS Insight published a "Global Wearable Forecast" illustrating the development of wearables from 2015 to 2019. They predicted that the sales number for wearables would 20% per year until 2024. The market would be estimated with $29 billions given

243 million unit sales that could be reached by 2022 [37]. They further observed that around 25 millions of kids' watches were sold in 2017 on the Chinese market what is said to be outstanding given that privacy concerns and regulations are more severe particularly in Europe. According to a survey by the market research platform Statista conducted in 2018, 43% of the respondents admitted that they are "somewhatlikely" or even "very likely" to purchase a wearable, such as the Fitbit or a smartwatch [48]. Despite such impressive stats promising the ongoing success of physiology sensing hardware, increased distrust among its users could be observed. Especially, the target group of the elderly has difficulties to adapt the new technology also due to the distrust regarding the system and problems to accept such new technology if its functioning is unknown [32]. What has been suggested by prior work [15, 188], was confirmed by Rupp et al. [225, 226] who found that the two factors self-determined motivation and technological trust are significantly affecting the users' willingness to use a fitness tracker continuously. In practical terms, this means that doubting the benefits of fitness tracker usage will lead to abandoning such technologies because Kononova et al. [142] could show in their work that maintenance strongly relates to perceiving the tracker usage as a longitudinal personal benefit. Nevertheless the recording of physiological signals is still a valuable method to detect physiological changes and trace them back to stress states. Since not only consumers rely on such data, but also researchers use fitness trackers for their studies [105, 122, 145], the suitability of such tools is important. The relevance of this issue certainly lies in the awareness of the given limitations physiological sensing bears. Some researchers [122, 202] even report that they reviewed different hardware before deciding upon a suitable sensing technology or device. This behavior clearly shows that there is need for a comprehensible and concise summary of predominant decision criteria of which consumers and researchers alike would benefit.

As it has been pointed out in the previous chapter, there are certain constraints when obtaining physiological data under different conditions. In this chapter, I show how consumers and researchers perceived these limitations and in how far they affect the user experience of both groups contributing to answer what implications such constraints have responding to **RQ1c**. Following up on the findings from prior work [98], I demonstrate how relevant criteria for physiology sensing systems have been identified in two user target groups. As a finding from this investigation, the "Design Space for Physiology-Aware Systems" is being present and visualized. Consequently, I answer **RQ1d** how we can identify and summarize criteria representing the differences among physiology measuring hardware. Being aware of the restrictions in different hardware is crucial from a design and development perspective when aiming for a reliable detection of stress in the user. I regard the conscious decision for or against a specific measurement tool as the prerequisite for the potential success of an application which is able to mitigate stress. In case the detection part of stress fails, such an application will fail too and consequently the trust in technology will decrease.

This chapter is based on the following publication:

- R. Poguntke, K. Hänsel, H. Haddadi, A. Alomainy, and A. Schmidt. Developing the Design Space for Physiological Measurement Tools: A Theoretical Foundation of Decision Criteria. (*to be published*)

4.1 Related Work

Prior work has been mainly focused on exploring the consumer perspective often with focus on usage patterns or efficiency. In contrast, we hardly know on what criteria consumers base their decisions to use fitness trackers incorporating physiological sensing. Even less explored is the decision process of researchers when choosing suitable hardware for their research projects.

User Perspectives on Sensing Wearable Usage With the rise of wearable sensing devices in form of fitness trackers or smartwatches, huge companies threw various physiological sensing hardware on the market promising consumers to increase health and wellbeing by monitoring physical activity. Nevertheless 50% of the users abandon their fitness trackers after one year [153], which has been extensively portrait in prior work [50]. To overcome this problem Canhoto and Arp [33] suggested to promote adoption and increase "device portability" and "resilience" to drive long-term usage. Respectively, there has been previous work on the capturing of consumer feedback on affective devices and on technological innovations in general as Prahalad and Ramaswamy [213] discuss. The information about consumers, their activities and usage contexts, being referred to as "customer intelligence" [49] has been proposed to be used to gain more insight in purchase criteria [31, 45, 84]. To bridge the gap between consumers and research examining user needs, Schuurman, Mahr and De Marez [239] suggest to apply the so-called "Lead User-concept" to increase user engagement as a driver for innovation. From a research perspective, there has been taken into account the user perspective recently. For example, Fritz et al. [77] investigated in a long-term study how 30 wearable owners use and

benefit from their fitness trackers, like Spiel et al. [254] did; both contributing to the understanding of the persuasive mechanism activity tracker manufacturers' often aim at. Moreover specific user groups had been interviewed, as Tholander and Nylander [261] did with athletes to learn more about their benefit from smartwatches used for sports. Moreover, various studies focus on the target group of elderly using health monitoring devices [32, 142]. In a first exploration of user needs regarding affective wearables, Hassib et al. [106] conducted an online survey on acquiring, sharing and receiving physiological data with 109 participants. They found that users preferred to get information from all types of data and mostly wanted to share the information with closely related people, i.e. partners and families.
While this work focused on the effects of sharing affective and physiological data neglecting the pragmatic requirements and challenges when dealing with physiological sensing hardware, Niess and Woźniak [188] had a distinct focus. They inquired users on the motivation and the related goals for using fitness trackers against the background of the decreasing popularity of activity recording wearables. Additionally, their participants reported that they would like to have a better understanding of the data provided by their device and more importantly, that they lack confidence in these data. With respect to the design of physiological sensing wearables prior work has put a lot of effort into the exploration of user preferences regarding wearable wellness devices [149, 217] and its integration in design. For example, Kim et al. [133] proposed a development process that comprises three stages to promote user-centered design and Zhou, Ji and Jiao [296] aimed to mass customize devices according to consumer needs by using affective and cognitive design.
In contrast to the manufacturers interests to investigate the consumers' usage patterns and opinion by market research methods, there has been no exploration on the requirements set by researchers who work and apply physiological sensing devices. The work by Nasir and Yurder [185] represented a starting point by collecting consumers' and physicians' feedback about wearable health technologies in the overall context of technology acceptance through an online survey. While Márquez Segura et al. [170] focused in their work on designing wearables to be socially engaging by asking design experts, i.e. larp designers providing interesting observations for the HCI community, the present work contributes to the understanding of consumers' and researchers' needs from a pragmatic point of view resulting in the seven dimensions of the *Design Space for Physiological Measurement Tools* and its sub-dimensions.

4.2 Developing the Design Space

In the following the development process of the *Design Space for Wearable Physiological Measurement Tools* will be explained in detailed. Hereby, the method and interview protocol applied in the user study will be explained. In the subsequently presented results the genuine experiences on wearable sensing devices will be illustrated, as well as the evaluation of previously obtained criteria.

4.2.1 Qualitative Evaluation

In semi-structured in-depth interviews with ten participants qualities for physiology sensing hardware relying on the expert and the consumer perspectives were deduced. By "experts" I refer to researchers who have extensive experience with physiological sensing hardware (see Section 4.2.1 for a detailed description of their expertise).

Method and Data Analysis Two groups of stakeholders, namely five consumers and five experts using with wearable sensing devices were interviewed. Therefore, ten individual semi-structured interviews comprising 242.29 minutes of qualitative data in total with an average length of 24.23 minutes each ($SD = 5.07$) were conducted. The interview length for both groups was comparable, experts ($M = 24.3$, $SD = 2,52$) and consumers ($M = 24.17, SD = 6,53$). For the analysis, two different methods resulting in the hybrid approach as recommended by Blandford, Furniss and Makri [19] were combined. Since two aims were pursued, namely collecting genuine experiences, opinions and concerns and evaluating the formerly found aspects, the data was analyzed twice with two coders performing the same analysis and frequently checked with each other to achieve reliability. An inductive thematic analysis [25] to identify themes emerging from the qualitative data was applied first. Second, statements were assigned to predefined codes following the deductive thematic analysis approach.

After having transcribed each interview, two separate coders marked all relevant statements using the qualitative data analysis software ATLAS.ti[7]. Guided by the six phases suggested by Braun and Clarke [28] shaping the data-driven, or inductive approach, the statements were then coded and searched for themes. After the themes had been compared, those that did not match for the two coders were discussed extensively until a consensus was reached resulting in five final

[7] https://atlasti.com/

themes characterizing the interviewees' thoughts, opinions and concerns. In contrast to the bottom-up approach used for the qualitative data on genuine experiences with physiological sensing hardware, a top-down analysis was performed subsequently meaning that there existed a predefined set of terms consisting of the five initial aspects identified in previous work (cf. [98]). These were namely *Comfort of Attachment*, *Mobility*, *Data Richness*, *Data Accessibility*, *Data Reliability*. Those five terms were used as codes for the interviewees' statements and new, not fitting aspects, were labeled with new open codes. After this first step, the statements' represented by codes were discussed and new dimensions or sub-dimensions from the new added codes were derived representing distinct notions of the labels or adding new perspectives. Through this deductive method, the existing dimensions (cf. [98]) were reviewed and refined, providing a contextual foundation for the *Design Space for Physiological Measurement Tools*.

Interview Protocol The interview question set consisted of five questions for the expert group and one more referring to a use case scenario for the consumer group being divided into three different phases as depicted in Figure 4.1. Hereby, the target was to retrieve as much information about personal experiences, criteria for picking a particular sensing device, and opinions on certain aspects as possible. For a coarse structure, the semi-structured interview was clustered into two introductory questions, opinions on devices' properties and resulting criteria reflecting the expectations for physiological sensing devices. While the first opening question was the same for both groups referring to *"personal experiences (good and bad) with wearable devices"*, the second question slightly differed. By this I wanted to know from the consumers which criteria play an important role to them when choosing a device, e.g. when they want to buy a new one. Likewise, the experts were asked secondly according to which criteria they pick sensing devices for their studies and research projects. For the consumers it was proceeded with an use case question asking them to *"imagine to take part in a long-term research experiment requiring to wear a wearable device sensing heart beat rate, etc. all day long"*; wanting to know *"what properties of the sensing device would be important to consider for deciding to take part and what would be possible deal breakers"*. Subsequently, both groups, consumers and experts, were presented five preliminary derived dimensions being introduced in [98]. Then they were asked to explain each dimension to figure out what their understanding is and which properties and aspects they associate with the given terms. In case they were not sure about the meaning, an example of how it could be understood ws given. After each dimension had been discussed, it was inquired *"which of these aspects played an important role and why"* - for the consumers when they considered buying a sensing device and for the experts when choosing

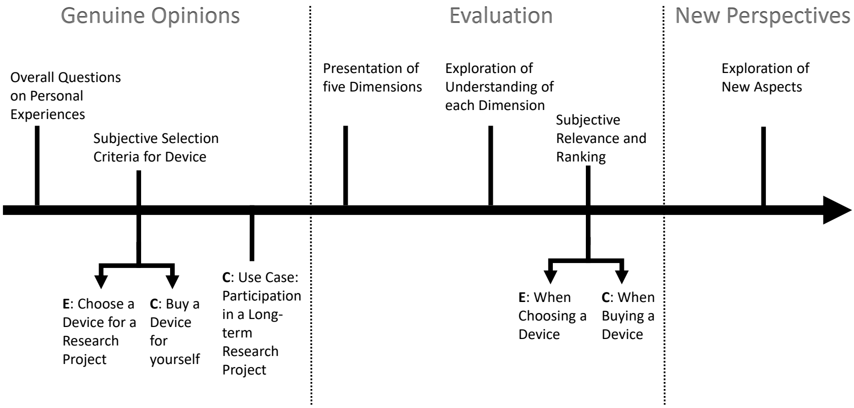


Figure 4.1: The entire interview protocol for both interview groups comprising the three following phases: Collecting genuine opinions, evaluating the formerly derived dimensions, and exploring new perspectives and aspects.

their apparatus for a research project. Accordingly, the interviewees were asked to rank the discussed dimensions based on their subjectively perceived relevance. Finally, a question was posed on whether there are *"any other important aspects or criteria for sensing devices, which we have not covered"* before the participants were thanked for their participation.

Participants and Procedure The consumer sample consisted of five interviewees ($M = 43.80$, $SD = 21.48$ years) including one male participant. The consumers were acquired via personal contacts and made sure that each of them had been owned a wearable fitness tracker for a minimum of six months at least. For participants *C1* and *C5* it has been the second device. Another selection criteria for the consumer sample had been that they wear it daily. This was also to ensure that they had sufficient usage experience and could comment on the whole spectrum of usage scenarios they experience in their everyday lives. Their usage behavior did not differ much, when asked for their motivation to own a physiological sensing activity tracker, all interviewees stated that they wanted to track their activity and particularly their step count. Three of them use it additionally to monitor their running or general fitness activities. The five inquired experts ($M = 30.60$, $SD = 2.61$ years), among them two females, were also acquired via personal contacts and were selected according to (a) their quantitative experience in their research field given in years, (b) their qualitative experience referring to the variety of sensor types they had worked with, and (c) a minimum of peer-reviewed publications involving physiological sensing. Thus, each expert holds or is currently finishing his or her PhD and has been working with a minimum of three sensor types for at least four years comprising several publications ($M = 15.6$).

After the interviewees were informed about the study purpose without revealing too much information on the content of the questions to avoid priming effects, they were provided with a consent form summarizing all information and informing about their right to withdraw the participation at any time. All interviews were conducted individually via Skype and audio recorded given the participants' consents. After having started with demographic questions, e.g. age, gender, and occupation, two different protocols (see cf., Section 4.2.1) were followed: one for the expert group and one for the consumer group. The entire procedure had been approved by the Ethic Commission of our institution.

4.2.2 Results

In the following the results from the qualitative study will be presented. They are subdivided into three parts following the interview protocol and describing experiences with wearable sensing devices, as well as deriving dimensions for the *Design Space for Wearable Physiological Measurement Tools*.

Genuine Experiences on Wearable Sensing Devices Initially the interviewees were asked about their experiences with physiology sensing wearables and devices. Whilst reflecting upon their subjectively perceived advantages and disadvantages, they often named aspects that had been summarized in [98]. As can be obtained from Figure 4.2, each researcher mentioned almost all aspects that had been observed previously. Likewise, all dimensions were also referenced by the interviewed consumers (cf., Figure 4.2). Consequently, *Comfort of Attachment* and *Mobility* represented the consumers' standpoint and the other three dimensions *Data Richness*, *Data Accessibility*, and *Data Reliability* reflected more the experts' criteria as indicated by the results from Figure 4.2.

Consumers When the consumers were inquired about their experiences with their individual wearables, e.g. fitness trackers all consumers mentioned *Comfort of Attachment* without naming it explicitly. While some referred

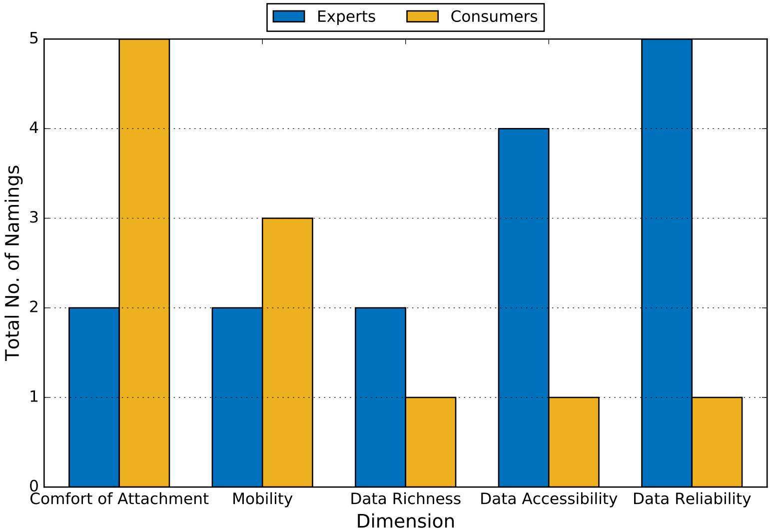


Figure 4.2: The total number of namings over all interviewees for the respective dimensions before it has been presented to them. As can be seen, *Comfort of Attachment* was named by all consumers and *Mobility* by the majority, while only one consumer mentioned *Data Richness*, *Data Accessibility*, and *Data Reliability*. Those dimensions that referenced data aspects were in contrast referred to by the interviewed experts showing that *Data Reliability* was mentioned by all experts and *Data Accessibility* by almost each.

to the size of their device as *"it should not be a huge, heavy chunk on the arm"* (*C1*) or *"too thick and big"* (*C4*), others emphasized *"the unobtrusiveness"* (*C2*) by stating that *"this model is convenient for [her] because [she] can wear it all day, I also do not have to take it off when I take a shower or whatever."* (*C1*). The latter statement also addresses the second dimension *Mobility* implying that the wearer is mobile while using the device. Thus, the consumers revealed that, e.g. taking a shower (*C1,C5*) means being mobile to them. Likewise, the battery life was important because as *C1* stated: *"I do not have to charge it everyday, so that I can use it continuously"*. Another aspect mentioned by *C1* and *C5* was summarized under *Robustness against External Factors*, since it refers to wearing it whatever activity the user does. This also means to *"not have to take it off when [I] take a shower"* (*C1*).

For the triad of data tackling aspects, each of them was mentioned by only one consumer each. The *Data Richness* meaning functions and variety of sensors was named by *C1* when weighing the price/quality ratio for such a device and contemplating the qualities of her fitness tracker. *Data Accessibility* seemed to occupy *C5* only. She explained that she picked her device considering the smartphone application quality and usability since she found many freely available applications *"badly translated or too confusing"* (*C5*) so that it had been an important criteria for her. Likewise, *C5* stated in the interview that she found *"inaccuracy [for her former device] annoying"* because arm movements had been also counted as steps. In summary, the consumers had a stronger focus on the two dimensions *Comfort of Attachment* and *Mobility*, while neglecting the other three dimensions slightly when choosing a sensing wearable.

Experts Only two of the interviewed experts thought of the dimension *Comfort of Attachment*. *E3* said that she considered *"if [the device] would be restrictive for the participants in their daily lives"*, researcher *E2* linked her decision on whether to use a device to the question *"How wearable is the device?"* because she *"was looking into concepts that work in the real world for a normal consumer, and not something that would need a lot of wires"*. The latter statement also implies the aspect of *Mobility* just as stated by the consumers. *E3* also mentioned battery life referring to this aspect as the consumers had done. *E2* enumerated multiple questions that accounted for *Data Richness*, for example when asking *"What data type"* or *"In what form"* she gets the data additionally speaking about the distinction between aggregated, filtered or raw data. *E3* named two other sub-aspects again matching the broader topic of *Data Richness*, namely the sampling rate and the question *"Which data do we get from the device?"* explaining that she picks devices for research projects also based on what kind of data she needs and respectively, which sensors she wants to use. *Data Accessibility* has been mentioned by four experts being occupied with *"How easy can I access the data"* (*E3*) and the related questions whether there needs to be implemented something, e.g. an API to be able to synchronize the device or access the data (*E3*); or a special software is needed, e.g. an SDK is given (*E2*). Moreover, *E5* explained that for him an issue was that the data needed to be *"read out in real-time 'on the fly' while the participants were wearing"* the device and that this had been difficult because of missing serial interfaces or lacking possibilities to transfer data via Wifi or Bluetooth, as stated by *E2* too. The most frequently mentioned criteria for all of interviewed experts was *Data Reliability*. *E4* stated that *"for scientific use, the most important feature for me is getting a very accurate value, with a high time resolution"*.

Almost all of our researchers talked about *"not very good experiences"* (*E4*) with wearables they had used or decided not to use because of *"noise in their data"* (*E1, E5*) or *"it was very unreliable"* (*E2*) being visible in the data, when, e.g. *"they contradict another sensor that measures the same or [...] if you move the hand, it becomes visible in the measurements without any changes"* (*E4*). Also *E4* was bother by the fact that *"most of the consumer version sensors, are either averaging a lot of making cases and that is not as accurate as you get from clinical sensors"*. It can be observed in Figure 4.2 that the triad of data linked dimensions, namely *Data Richness*, *Data Accessibility*, and *Data Reliability* was more important for researchers than for consumers when they had been asked about their overall experiences with sensing technologies.

Evaluating the Design Space Following the interview protocol, the interviewees were asked to comment on each of the five dimensions identified in previous work (cf. [98]) aiming to first capture their individual understanding of what these dimensions refer to and what properties or characteristics they linked to them. These dimensions have not been validated in the introducing study, but emerged from the authors' observations and experiences. For this, the present work aimed to validate those from a consumer and researcher perspective.

In most cases both interview groups briefly named the dimension and then described how the respective dimension is represented in the real-world context to them. It was found that the initial dimensions could be split into sub-dimensions deduced from different aspects stated by the interviewees. In Table 4.1, there is an overview depicted, showing the five initial dimension and what sub-dimensions had been derived from the interviewees' statements. The mapping to the initial dimensions was done on the basis of the interviews and extensive discussions between the coders as described in Section 4.2.1. All dimensions are listed in the first row and all sub-dimensions being deduced from the following qualitative results are listed in the first column. The interviewees' statements are mapped to the respective dimension and specific sub-dimension. To highlight the most important results, the following color coding is followed: green cells signify the most important overlaps where at least two interviewees from each group (consumers or experts) mentioned this aspect; yellow cells refer to this criteria that

	Dimension					
Sub-Dimension	Comfort of Attachment	Mobility	D. Richness	D. Accessibility	D. Reliability	New
Wearing Comfort	*C1, C2, C3, C4, C5 / E1, E3, E5*					
Connectivity				*E2, E4*		
D. Accuracy					*C1, C2, C3, C4, C5 / E3, E4,E5*	
D. Access. Effort				*C1, C2, C3, C4, C5 / E1, E3*		
D. Format				*E2,E5*		
D. Transm. Effort				*C5 / E1, E2, E3, E5*		
Degr. of Preprocessing			*E2, E3, E5*			
Degr. of Resolution			*E2, E3, E4, E5*			
Degr. of Restrictiveness	*C1, C2, C3, C4, C5 / E3, E4,E5*					
Degr. of Testedness						*E3, E5*
Ease of Setup		*E4,E5*				
Invasiveness	*E3, E5*					
Phys. Properties	*C1, C3, C4, C5 / E2*					
Robustness		*C1, C3, C5*				
Sensor Variety			*C1, C2, C3, C5 / E2, E3, E4, E5*			
Software Reliability					*C1, C2 / E1, E2, E3, E4,E5*	
Trustworthiness						*E3, E5*
Unobtrusiveness	*C2, C3, C4 / E3, E4*					

Table 4.1: All (sub-)dimensions shown in yellow and respectively blue, if they have been mentioned by at least two consumers or respectively two experts; green indicated that at least two of each group named it.

had been mentioned by two or more consumers; finally the blue cells highlight what two or more experts had been stating in the interviews.

Consumers Consumers referred to the dimension of *Comfort of Attachment* by naming four sub-dimensions. *C1*, *C3* and *C4* emphasized the *Wearing Comfort*. *C1* said that she is happy with her device *"because it is small and very comfortable to wear"*. *C4* explained that it is important *"that it fits well"*. Moreover she mentioned the *Physical Properties* as another characteristic arguing that her device feels *"not like a chunk on the leg"*. A closely linked aspect was considered by *C4*, who described this dimension as *"one does not get stuck with it"* or that it should not impede the wearer (*C5*). In connection with this, *Unobtrusiveness* was also being mentioned by *C2*, *C3*, and *C4* saying, e.g. that *"[it] should be unobtrusive"* (*C4*) and that they *"would not wear it [if] it is too obtrusive"* (*C3*). Further, *C2* and *C5* referred to the physical quality and handiness of the device, such as different adjustment options for the quick-release catch and *"setting the size of the bracelet"* which is summarized under the *Wearing Comfort*. Another involved criteria was the *Degree of Restrictiveness* as *C5* added.

For the consumers, the next dimension *Mobility* was closely linked to the previously discussed issue, which is why *C2* also stated that it is as important as *Comfort of Attachment* since he *"wears it daily and during all activities"*. On the contrary *C3* stated that *"mobility is important but it does not have the highest priority. It is more important that it is comfortable to wear so that I can wear it anytime"*. Likewise she clarified that *Mobility* for her means to be able to take the device *"anywhere I go anytime, e.g. under the shower"*. What her statement also implied was the sub-dimension *Wirelessness* being also mentioned by *C4*.

With respect to *Data Richness*, all consumers preferred to have more data meaning that they appreciated different kinds of sensors providing, e.g. *"pulse or calorie count"* (*C2*) or *"jogging GPS tracks"* (*C5*). Hence, *Data Richness* was subdivided into *Sensor Variety* inter alia.

C5 was the only consumer who had an explicit requirement with respect to *Data Accessibility*. She explained that for her a *"clear criteria"* had been that activity tracking *"software is often not available for Ubuntu"* operating systems which is what she uses herself. As a consequence she looked for a device that provided access to the data through a smartphone application since this is *"independent from platforms [operating systems]"*. All other consumers reported to access their data through a smartphone application (*C1,C3,C4*) or directly on the device's display as *C2* does. In general, they judged the effort to access their data as relatively low resulting into the sub-dimension *Data Accessibility Effort*.

The final dimension *Data Reliability* played an important role for all consumers. So, most of them thought primarily of *Data Accuracy* (*C1*). Whereas *C1* and *C3* double checked their devices' results on activity tracking with a separate smartphone application and *C5* even compared the results of two different wearable fitness trackers. *C4* stated that she *"can make trade-offs"* regarding the reliability because she would then have to compare it but the values *"approximately fit"* according to her estimation. Facing the fact that also *C2* noticed irregularities in the *Data Accuracy*, the results suggest that all consumers were aware that their fitness trackers did not work 100% accurately.

Experts Regarding the *Comfort of Attachment* interviewee *E1* mentioned that *"the usage of the system can be a burden to the user"* in case they would have to attach it by themselves thinking of complicated electrode placements or sensor installations. *E3* was more concrete in her explanation and mentioned as requirements that the *"sensor is comfortable to wear; unobtrusive no matter where you wear it, it should not impede,[...] no restrictions in what you are doing usually"*. Thus, she named the aspects that were phrased as the new sub-dimensions *Wearing Comfort*, *Unobtrusiveness* and *Degree of Restrictiveness*. Moreover, the experts referred to *Invasiveness* which is linked to the utilized sensing technology referring to potential invasive consequences for the wearer. *E3* took exemplarily *"skin irritations"* as an important criteria, while *E5* stated that invasive methods would first, lead to loosing participants and second, cause problems with the ethic committee of their institution.

When mentioning *Mobility E1* said that *"For the mobility, we wouldn't want the wearers to be stretched to a wire, I guess we would go with something that is wireless and allows mobility"*. Moreover *E3* associated *Mobility* with the *Degree of Restrictiveness*, whereas *E4* and *E5* thought of the *Ease of Setup* with regard to *Mobility* considering *"How easy are [the devices] to transport? How easy is the setup? How easy is the carrying?"* (*E5*).

From a research perspective *E1* argued that *"we always prefer the devices that have the most amount of sensors in the least amount of space"* but again referring to the study context explaining that hereby the *"data richness is not really an important or essential factor"*. A device being equipped with *"multiple types of data"* (*E2*) and *"how many different signals are recorded"* (*E4*) was summed up in the sub-dimension *Sensor Variety*. Additionally the sub-dimension *Degree of Resolution* was pointed out by *E2* stating that *"high resolution is important"* strengthened by *E4* listing *"precision or time resolution"* and *E3* mentioned exemplarily the *"sampling rate"* as a metric. Based on *Data Richness E5* initiated the new sub-dimension of *Degree of Preprocessing* referring to the opportunity to get *"not only one measurement parameter but more derived measurement parameters"* giving the example of heart beat values and the derived heart rate variability.

With respect to *Data Accessibility E1* clearly stated that he would *"prefer devices which give open access to the data"* emphasizing that it could be a problem for researchers if they would need to develop their own software to access the recorded data because this is time-consuming as he reported from his own experience. Further he referred to *Data Accessibility Effort* as also did *E4* by mentioning *"how easy it would be to get the data off the sensor"*, *Data Transmission Effort* and *Software Reliability* by emphasizing that it is appreciated if *"software is reliable and you can get the data whenever you want; also you do not have to develop your own algorithm for Bluetooth transmission of the data"*. *E3* agreed with this and also considered *"How do I get the data? Do I have to code something or not? Does it work out-of-the-box or do I have to synchronize?"*. For *E5* also the *Data Format*, e.g. if *"one gets the data in a format that one understands [or a commonly accepted], e.g. a CSV file"* belonged the dimension of *Data Accessibility*. The sub-dimension *Connectivity* was further mentioned by *E2* and *E4*.

Referring to *Data Reliability E1* said *"this is the most important subject that you have here on this list because if you are developing machine learning algorithms which would work with the data given by sensors, then machine learning algorithms would learn also some noise in the sensory data"*. Hereby he highlights the severe consequences unreliable data can have, particularly for researchers. Likewise *E3* and *E5* highlight that *"the data we get should measure what we want and do that accurately"* (*E5*) referring to *Data Accuracy*. *E3* and *E4* further mention the reliability regarding the recording what is summarized in the new sub-dimension *Software Reliability*.

New Dimension: Trustworthiness The sub-dimension *Degree of Testedness* was mentioned by *E3* who said that *"some sort of confidence in the long-term existence of the company"* is needed. And *E5* explained that his team had positive experiences with the manufacturer's support which had been important for the project success. Building up upon personal experiences the new dimension of *Trustworthiness* mentioned by the researchers exclusively was added.

Dimensions	**M (SD) of rankings** Consumers	Experts
Comfort of Attachment	2.2 (1.1)	2.4 (1.7)
Mobility	2.4 (1.5)	3.0 (1.6)
Data Richness	4.6 (0.5)	4.4 (0.5)
Data Accessibility	3.0 (1.6)	3.2 (1.3)
Data Reliability	2.8 (1.3)	1.6 (0.9)

Table 4.2: Rankings of the five previously found dimensions according to consumers' and experts' subjective relevance (means and standard deviations). Higher values signify that this dimension has been ranked lower while low values represent better ranks.

Ranking of Dimensions Besides their explanations and associations regarding the five aspects *Comfort of Attachment*, *Mobility*, *Data Richness*, *Data Accessibility*, and *Data Reliability*, participants were asked to rank them in order according to their subjectively perceived relevance - a lower value indicates a higher importance. As can be obtained from Table 4.2 consumers ranked *Comfort of Attachment* and *Mobility* highest with an average rank of 2.2 and 2.4, followed by *Data Reliability* on rank 2.8 in average. For *Data Accessibility* each rank was given once to this dimension resulting in an average rank of 3.0 with the highest standard deviation of 1.6. The dimension *Data Richness* was ranked on 4.6 in average by the consumers with the lowest standard deviation (0.5). For the experts' rankings *Data Reliability* was given the highest priority with an average rank of 1.6 ($SD = 0.9$) followed by *Comfort of Attachment* on rank 2.4 in average ($SD = 1.7$). Again each rank was given once to the dimension of *Mobility* ($M = 3.0$, $SD = 1.6$) and almost the same pattern can be observed for *Data Accessibility* ($M = 3.2$, $SD = 1.3$). Least importantly priorized was ($M = 4.4$, $SD = 0.5$). Remarkably, *E1* and *E2* could not decide whether *Comfort of Attachment* or *Mobility* was more relevant for them. While *E1* put both second, *E2* considered these two dimensions most relevant.

Derived Themes Describing Opinions, Thoughts, and Concerns During the interviews all participants commented on many different aspects and expressed their thoughts on the usage of wearable sensing devices including their concerns. Besides the criteria included in the *Design Space for Physiological Measurement Tools* as dimensions and sub-dimensions, the following five themes reflecting the interviewees' perspectives have been identified: *Convenience Factors Social Acceptability*, *Financial Costs*, *Study Context*, and *Privacy*. Due to their high subjectivity a parameterization and representation using two extremes, contrary to the other named dimensions included in the *Design Space for Physiological Measurment Tools*, is hardly possible and moreover difficult to measure objectively. Therefore, the following five aspects have been labeled as themes allowing valuable insights that capture the "soft" considerations before deciding for a device.

Convenience Factors What has been also revealed through the inquiries were so-called *Convenience Factors* reflecting properties that were meaningful to practitioners but which were not crucial for the purchase decision. For example, the requirement of an *Unconcerned Handling* meaning that the users does not have to take care of the device is something that is "nice-to-have" but cannot be accounted as "must-have". For the consumers the benefit mainly is in the circumstance that *"you [do] not have to constantly think think of it"* (*C1*), also referring to not having to *"take it off when I take a shower"* (*C1*) or like for *C2* to feel no impedement while playing an instrument. The *Unconcerned Handling* further manifests in the *Power Supply*, since *C1* emphasizes the importance of not having it *"to charge everyday"*, also mentioned by *C2* and *C3*, who compared it to a smartphone that needs to be charged daily. Likewise, expert *E2* commented that the *"battery life"* is important for her as a researcher and *E3* confirms this by listing this as one of the properties she considers before conducting a study. Referring to the maintenance effort, the issue of the *Ease of Operation* saying that it should be *"easy to start and stop and only limited interaction is needed"* (*E3*) was discussed. Moreover, for consumers this dimension represents an important criteria, and *C4* perceives it as the most important, if the device *"is easy to handle and operate"*. Since as *E2* stated *"it matters if [the device] gives the expected outcomes, and for example how many times it disconnects"* and also like *E4* points out that *"it is important for me that when I start using the tool for sensing, I know when it has started [...] I need to know if it is actually working"*, *Feedback* as another meaningful aspect has been identified. In the interviews important hints on the usefulness of a feedback function, e.g. the device should give *"a warning that there is an issue with the data transmission"* (*E2*) were gained. For example, the practitioner would be notified if the *"data was corrupt"* (*E2*). Interviewee *E5* further explains that *"if I have a device that is not recording on-the-fly and that I let record the data all day long, then I do not want to realize in the afternoon that it stopped recording after half an hour"*. Apart from these thoughts, the *Availability* of the devices pointed out by *E5* was one of the criteria when choosing a device. Because he considered that, e.g. *"some eyetrackers are not available in large numbers [...] or it is not possible to ship it to Europe."*. *E2* brought up another aspect saying that *"mobility would mean connectivity to other devices [...] it connects to my phone"* referring to *Connectivity*. This particular property corresponds to the "nice-to-have" property of *"Wirelessness"* being desired by *E2*.

Social Acceptability *E2* was the only expert who talked about *Social Acceptability*. She said in the context of *Comfort of Attachment*, that it *"is not simply that the sensor is comfortable to wear, but also that consumers are fine with wearing it in public"* referring to *Social Acceptability*. Asking her a bit more detailed what she would expect to be socially acceptable wearable, she gave an example answering that *"wrist bands are acceptable, rings that look odd are acceptable because they are small. Eye glasses are starting to be a bit acceptable, but they are a little bit freaky in public. Anything head mounted is not publicly accepted. Only if you wear it within a setup, like a conference or other people are wearing it."*. By this, she mentioned many important factors, such as the size and the context, but also the aspect of *Fashion*. Accordingly, interviewee *C3* mentioned the aspect of embarrassment when you wear a sensing technology stating that she does not want *"to look embarrassing"*.

Financial Costs For the *financial costs* expert *E1* admitted that the *"prices should be also considered"* telling from his own experience that the financial opportunities depend on the project, what was also confirmed by *E2* who linked the role of financial expenses on sensing technologies to the study context saying *"price is important in the project where I need to conduct a field study and I would need a lot of devices"*. Likewise *E3* and *E5* agreed that having a certain budget also sets constraints and *"some things [devices] are not affordable"* according to *E5*. Moreover, the price for a activity tracker was subject to considerations for the consumers. While *C5* mentioned the *"price/quality ratio"* as a criteria, *C1* stated that she *"would generally contemplate the price. [...] The price should be reasonable."*.

Study Context As mentioned against the background of *Financial Costs*, the *Study Context* has a huge impact on how the study needs to be conducted for researchers. Throughout the interviews various examples were found describing that the study context influenced either the budget distribution, as *E1* explains saying that *"either the project can cover the expenses or the user has to buy them"*, or the prioritization regarding the relevance of certain criteria like the following statement of *E2* points out: *"If I am planning to do something that requires a field study, where people need to use the device for two or more weeks, then the Mobility and Comfort of Attachment come as number one. [...].Then Data Reliability comes as number two, then Data Accessibility and Data Richness. If it is something that I am looking into like what I can extract from the data on a lab setup level, then of course, the Data Reliability is number one and the Data Richness comes as number two, Data Accessibility number three, Comfort is number four and Mobility is last, because of the lab setting."*. These view was shared by all experts except for *E4*, who did not mention this in the interview.

Privacy The aspect of *Privacy* was mentioned by only one customer. *C1* stated that she *"would not necessarily want to be monitored how regularly and when [she] move[s]"* reflecting her concern that activities trackers could be misused to monitor people. Likewise, from the researchers only *E2* mentioned that the design space lacks the dimension of *Privacy* without giving additional explanations.

4.2.3 Structure of the Design Space

The *Design Space for Wearable Physiological Measurement Tools* has been developed on the basis of ten semi-structured in-dpeth interviews with experts and consumers. In the following the structure of the design space will be presented in detail. All six dimensions including their 17 respective sub-dimensions will be explained and put in the context of physiology-sensing device usage.

Comfort of Attachments The *Comfort of Attachment* is mainly important from the user perspective and summarizes physical and social discomforts associated with wearing a device. This dimension consists of four sub-dimensions referring to the underlying aspects being mentioned by the interviewees. *Wearing Comfort* is one of these; this sub-dimension refers to the pure agreeableness that is experienced, or not experienced when the sensing device is being attached to the body. Accordingly, there might be restrictions imposed on the wearer which is represented in the sub-dimension *Degree of Restrictiveness*. For example, if the wearer wants to play an instrument which requires the hand that the device is attached to for playing, the sensing device could hamper the instrument and restrict the musician while playing. In such a case the *Physical Properties* of a

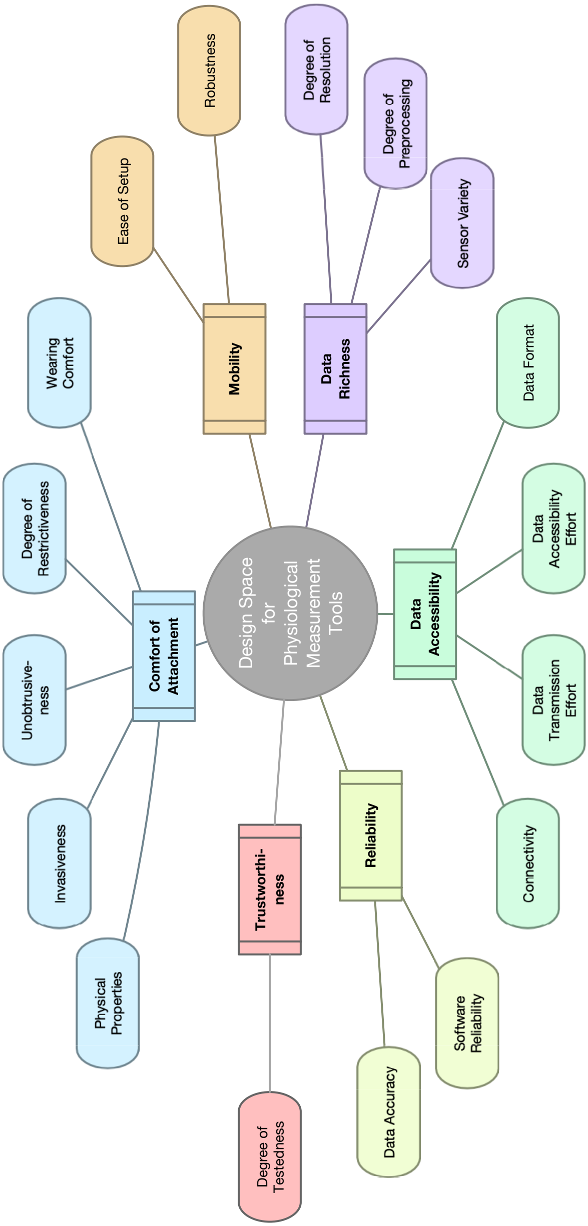


Figure 4.3: The *Design Space for Physiological Measurement Tools* in a mind-map-like visualization; including the six dimensions and its 17 sub-dimensions sorted by colors.

device, e.g. form factor, such as its color, shape, size, etc influence the perception and respectively the comfort. Another crucial factor is, if the attachment bears *Invasiveness* or even worse, has harmful consequences for it's wearer. While some sensing technologies might be cumbersome to wear, e.g. EEG hardware or chest straps, others might be less invasive when being put on easily accessible body locations, i.e. the wrist. *Unobtrusiveness* of a device was further found to be another selling point. In case the device communicates privacy-sensitive data, the wearer might not want to show that he or she is using such a device and therefore the criteria of an unobtrusive wearing is important.

Mobility The ability of being mobile when wearing the physiological sensing device has been summed up in this dimension. Another significant issue for researchers is the *Ease of Setup* of a particular device. In some application scenarios, the assessment of physiological signals might require an enormous effort to setup the hardware or software. Thus, this sub-dimension reflects a researcher's perspective and neglects the consumer, since it has only high relevance when high-resolution or expensive equipment is used. Another sub-dimension affecting the perceived *Mobility* is *Robustness*; this refers to the resistance against external, and particularly environmental factors, such as water or heat.

Data Richness When it comes to *Data Richness*, two different factors were considered, the *Sensor Variety* referring to the amount of sensors being provided by a device and the quantity of the sensor data. The later is being taken up by the following two sub-dimensions. While the *Degree of Resolution* means, for example the sampling frequency provided by the device, the *Degree of Preprocessing* comprises the amount of, e.g. filters that is used to remove noise in the data or artifacts before the data can be accessed.

Data Accessibility For the dimension *Data Accessibility* three main factors were distinguished. Before the data can be accessed, there might be further equipment needed and therefore the sub-dimension *Connectivity* means the ability to connect the physiological sensing device to other technologies, e.g. a laptop. In the second step after the data recording, there is often a transmission of data required. In practice, data has to be transferred from a server or a sensing device, which can range from very low to extremely high *Data Transmission Effort*. As a final step often before accessing the data, there might be a special manufacturer's software needed or a Software Developer Kit (SDK) needs to be installed resulting in a variation of the *Data Accessibility Effort* to get granted access to the recorded physiological signals. An additional substantial factor for accessing data is the *Data Format* that is being used for providing the physiological signals. Since some products follow their own proprietary formats, others rely on a standard, such as Comma-Separated Values (CSV) files. Whether a standardized format is chosen can have severe consequences for the researcher wanting to access the data with the lowest effort as possible, what makes this issue a considerable sub-dimension.

Reliability What has been initially introduced as the dimension *Data Reliability* within previous work (cf., [98]) and found throughout the interviews, was simplified into *Reliability* to address a broader spectrum. By this aspect originally it has been referred to the newly added sub-dimension *Data Accuracy* meaning the degree to which the user of the data can be sure that no false values have been recorded. An additional understanding refers to the *Software Reliability* being another sub-dimension describing the susceptibility to errors.

Trustworthiness The dimension of *Trustworthiness* emerged from consumers', and particularly researchers' considerations regarding the *Degree of Testedness* affecting their decision criteria. If the device's manufacturer was offering a huge amount of the device and had tested the device for certain criteria, e.g. reliability or invasiveness beforehand, this was clearly a confidence-building argument for purchasing such a device.

4.3 Discussion

In the following, the reasoning of the *Design Space for Physiological Measurement Tools* based on the interviewees' statements supported by related literature will be presented. Moreover, the deduced themes will be discussed and finally limitations of the conducted study will be acknowledged.

Reasoning the Design Space from the Qualitative Results In general, it has been observed that the two dimensions *Comfort of Attachment* and *Mobility* had been strongly emphasized in the consumers' statements. As can be seen in the quotes, the quality to be able to wear an affective device *"all day long"* (*C1*) so that wearers *"do not realize it"* (*C3*), is what makes it *"unobtrusive"* (*C2*). *E3* put the *Unobtrusiveness* also as a requirement when thinking of the *Comfort of Attachment*. Moreover, there were more aspects being referred to by the consumers implying that *"due to too much movement the wristband comes loose"* (*C1*) or *"the quick-release catch does not close correctly sometimes and opens too easily"* (*C2*), what affected the *Wearing Comfort*. This sub-dimension is significantly determined by the *Physical Properties* of a physiological sensing device. For example, *C1*, *C3*, *C4*, and *C5* explicitly said that they do not want a *"huge, heavy chunk on the arm"* (*C1*) also referring to the shape and size, e.g. being *"too bulky"* (*C3*). Additionally, Lazar et al. [150] report that users perceived their devices often as "uncomfortable" and "obtrusive" what is in line with the participants' statements and the fact that the dimension *Comfort of Attachment* was either ranked on the second position (mean rank of 2.2) as can be obtained from Table 4.2. Bodine and Gemperle [21] investigated the interdependence of a device's functionality and its perceived comfort to wear it. They found that the degree of functionality had an important impact on how the *Wearing Comfort* was evaluated. They concluded from their results that not only specific body locations were chosen over others for certain functionalities, but also the functionalities as such must be appropriate to their perceived discomfort. Speaking of functionalities, Hernandez et al. [115] investigated wearable devices' suitability for employing the ESM revealing that one crucial factor influencing user experience and ESM's outcome is screen size, what is recognized as *Physical Properties*. Closely related to this issue, *Social Acceptability* was mentioned by *E2* and also by *C3* saying that one does not want *"to look embarrassing"* with a device. Interestingly, within the context of *Social Acceptability*, *Fashion* played an important role, particularly for women when choosing a device. *C3* admitted that: *"For everyday life it is amazing but for something chic, I would not wear it"*. Likewise *C4* picked the bracelet according to the occasion saying: *"I have a second bracelet that looks a little bit more chic"*. The issue of not having *"the looks"*, as *C1* calls it, has been considered by prior work too [156]. Further, Hui-Wen Chuah [43] considered smartwatches as a way out of the dilemma of not following fashion goals in activities trackers and shaped the term "fashionolgy" referring to smartwatches that combine technology and fashion. Many different factors, such as *Fashion* and aesthetics, but also moral codes and the context of use have an impact on the perception of *Social Acceptability* as has been explored by Norene Kelly within the development of the WEAR scale measuring the *Social Acceptability* of a wearable device [130]. Accordingly, this aspect has been discussed as a theme and was not included in the *Design Space for Physiological Measurement Tools*.

The dimension *Mobility* is closely linked to the *Comfort of Attachment*, and thus it often implies aspects relevant for this dimension. *E4* exemplarily discussed the sub-dimension *Unobtrusiveness* controversially saying that *"the sensor should not* *affect the behavior of the participant"* referring to being mobile and the *Degree of Restrictiveness*.

In this context *Wirelessnes* was mentioned as a significant advantage enabling the wearer to be mobile as has been pointed out in prior work under the term "portability" [101, 219]. When talking about large hardware, the researchers *E4* and *E5* further paid attention at the *Ease of Setup* asking questions like *"How easy are [the devices] to transport? How easy is the setup? How easy is the carrying?"* (*E5*). Like Townsend et al. [263] call it, the "deployment of new technology" (p.1) is a major challenge and therefore important to be considered because it is directly linked to the *Mobility* of the wearer. For example, highly sensitive sensing devices, such as the Nexus kit often do not only require more time to be set up, but also allow only limited physical activity when *Reliability* should still be given for the measured data.

A different criteria was introduced by *C5* who specified the dimension *Robustness* by explaining that waterproofness (*C1*) is an important criteria when choosing a device and considering how mobile oneself is. Hand in hand with *Robustness* goes, what was identified as *Unconcerned Handling*. *C2* mentioned battery life as one thing he does not want to take care of. The *Power Supply*, was also mentioned in previous work as a crucial challenge for physiological sensing systems [101, 263]. In a user study conducted by Lazar et al. [150] participants were asked why they stopped using their smart devices after a short period of time. They found that apart from perceiving the recorded data to be useless, consumers had difficulties to maintain the devices due to its high frequency of charging or making sure that the bluetooth connection was enabled, etc. All these factors influence consumers' user experiences, which is highly dependent on the ease of usage. This identified theme also underlies subjective feelings and the question of how annoying the frequent charging of a device is perceived. The suitability in this regard must be answered individually for different user groups and user needs, the newly added dimension *Ease of Operation* can be measured with the help of questionnaire, such as the System Usability Scale [230] or the Usability Metric for User Experience [70]. According to Townsend et al. [263] the "ease of use", how they call this dimension, is one of the advantages that a wireless tool holds over other systems using cables. Among these preferable properties is also the "reduced user discomfort" and the "enhanced mobility" [263] which had also been proven to reflect the consumer's perspective best, since both dimensions, *Comfort of Attachment* and *Mobility*, had been enumerated by each consumer or at least the majority (see cf., Figure 4.2). The other three dimensions, namely *Data Richness*, *Data Accessibility*, and *Data Reliability*, associated with the recorded data, seem to be more important for researchers. This can be obtained from Figure

4.2 because these aspects were enumerated by one consumer only each and but by almost each expert.

Accordingly, *E2* named several questions like *"What data type do I get? In what form do I get the data? Is it aggregated or completely raw data from the sensor? Is it interpreted somehow? Is it filtered data?"* what also shows the complexity of the dimension of *Data Richness* being the reason for why it has been split into the three sub-dimensions *Degree of Resolution*, *Degree of Preprocessing*, and *Variety of Sensors*. The latter aspect was added because the amount of sensors also plays a role for the consumer grounded in *C1*'s statement that she uses *"only a very small part of [her] watch"* and *C2* wished for some more data or quantified parameters, such as *"pulse or calorie counter"*, while *C3* sums it up as follows *"How many data do I get from my device"*.

The question *"In what form do I get the data?"* raised by *E2* tackles another important sub-dimension being grouped under the dimension *Data Accessibility*, namely the *Data Format*. As this dimension shows, there are a couple of factors that can be crucial for the ease of conducting a research project. Moreover the automatic synchronization of a device as mentioned by *E3* or the presentation of the data, e.g. in a smartphone application (*C4*) or on a small display integrated in the fitness tracker (*C2*) has to be considered for when choosing a device. This dimension *Data Accessibility Effort* refers to the amount of effort that it takes to access the data. Correspondingly, interviewee *E3* explained that they wanted to read out data in real time for a study pointing towards another important sub-dimension, what has been called *Data Transmission Effort*. Finally, the *Connectivity* of a device completing the sub-dimensions of *Data Accessibility* has been paid attention to due to *E2* explaining that the ability to easily connect the physiological sensing technology to other devices, e.g. a PC or a smartphone can facilitate the realization of projects by far, when there is no special soft- or hardware needed.

From a researcher's perspective, *Accuracy* was also mapped to *Data Richness* by *E4*. Despite that accurate data is more likely to be recorded, if *Data Richness* is given, *Accuracy* was assigned to *Reliability* as a sub-dimension. Because from an overall view the data has to be reliable- no matter how many sensors are available and how fine-grained the resolution is. This grouping was strengthened by the interviewees, who said that *"data that do not represent the reality and that you cannot rely on"* cannot be reliable (*E5*). And for *E1* other criteria are put in the background *"if [the device] wasn't reliable then that is a problem"*. The challenge of granting *Reliability* for physiological sensing systems has also been mentioned in prior work discussing developments and advances in these technologies [101, 263, 219] without providing researchers or consumers with feedback on the crucial issue of *Reliability*. Also the consumers mentioned this issue, as *C1* who admits that she *"can not be sure one hundred percent, whether I slept so long or whether I was really always awake when the bracelet shows I was awake"*. The findings from a qualitative user study conducted by Niess and Woźniak [188] on wearable owner's motivation and goals are in line with the observation that users lack confidence in their fitness tracker's data. Since the information provided by activity trackers on sleep quality influence their owners' perceptions, the reaction of *C1* to manually enter sleep quality information in their devices due to insufficient reliability can be explained with the users' urges to be in control of their goals as found by Niess and Woźniak [188].

Related to the topic of trust is also the *Trustworthiness* which describes the confidence in a company offering a device. Since it can be crucial in the decision process purchasing physiological sensing hardware to know if a company can be trusted, researchers perceived this aspect as important. *E5* stated the problem that *"often are these things [devices] announced for a long time but then they are not published in the end [...] are not available in large numbers, only some people get them then, or it is not possible to ship it to Europe"*, while *E3* explained her criteria that she needs to have some sort of confidence in the long-term existence of the company because it had happened to her that when a company was bankrupt *"the support was stopped, servers were down, and so the sensors became useless"*.

For the *Design Space for Physiological Measurement Tools* the dimension *Trustworthiness* has been taken up. By mentioning this dimension, the interviewees expressed a certain lack of confidence in an advertised device which may promise fancy functionalities, although it has not been sufficiently tested before, which is summed up in the *Degree of Testedness*.

Concerns beyond Decision Criteria However, researchers and consumers also expressed deviated opinions on *Privacy* related to the usage of wearables, which has been identified as another challenge for physiological sensing systems in prior work [101, 219]. For example, *C1* stated that she does not *"want to be monitored how regularly and when I move"*. This statement clearly refers to the storage of the data that is recorded by the sensing device and is representative for the findings from Raij et al. [216]. In their study, they asked 66 participants about their opinions on the disclosure of private information including, for example physiological data. They found that people who had a stake in the data, meaning their own data had been reviewed by themselves, had a better understanding of the privacy threats and perceived it more threatening than those people who had no stake in it. One of their key findings was that physiological and, particularly information on stress and conversation excerpts were prone to evoke the most concern regarding privacy in users. The question of how to deal with personal data relating to health and wellbeing being recorded by wearables, is also becoming more relevant at the workplace since companies encourage employees to promote wellbeing [87]. A user study on the willingness to share physical activity at the workplace and hereby disclose private data, showed that sharing health related at the workplace is being viewed sceptically due to concerns that private and work activities could be mixed [88]. In this context, Gorm and Shklovski [88] further propose to rethink the handling and advertisement of revealing privacy sensitive data. Addressing this problem Reynolds and Picard [221] proposed an ethical contract, which turned out to be perceived as a beneficial mean in dealing with and respecting privacy for affective data. An distinct perspective on potential improvements for wearable devices provided Marakhimov and Joo [168]. They published an interesting user study on the link between consumer concerns, e.g. privacy or health related, initiated coping processes, and the infusion of health care wearable devices. They revealed that problem-focused coping strategies are strongly related to extended use. Further, their study gave support for the assumptions that owners abandon their health care wearables only after few time [35] because the owners consider these new technologies often to be menacing for their health and wellbeing. This phenomenon arises from privacy concerns, as their study indicates.

As another considerable aspect the *Ease of Operation* has been identified as being meaningful for participants in terms of a *Convenience Factor*. Regarding their operation, wearables often require too much effort, which Norman [189] declares to be a killer for the ease of use. This again affects the user experience quality negatively and corresponds to consumers' experiences. *C4* emphasized that it would be a deal breaker *"if [she] had to set/operate so many things, e.g. many buttons"*. Comparing the GUIs of modern wearables to Fogg's user experience guidelines derived from his "Persuasive Technology" [74], they often suffer from a lack of comfortable user experience because they do not provide an experience to the user and rather fulfill their function and therefore hampers familiarization. The fact that the familiarity with a device represents a considerable detail and is linked to the *Ease of Operation* becomes obvious in Hernandez et al. [115] study indicating that this dimension has an impact on the user experience and thus the suitability of a wearable device for employing the ESM.

Limitations Throughout the interviews overlapping aspects were found and hence could be linked to not one design space dimension exclusively. Given the subjectivity that is always part of qualitative data, I acknowledge that the assignment of the distinct sub-dimensions is disputable for some cases. Further interviewing all participants remotely could have had an effect on the results too. Likewise, the sample size of five representatives for each stakeholder group does not seem sufficient. However, the present work does not aim for a comprehensive assessment, but rather represents an exemplary research method to validate and summarize valuable criteria shaping the decision process for purchasing and respectively using a wearable sensing device.

4.4 Design Recommendations for Wearable Physiology Sensing Devices

Derived from the presented research on physiology-aware systems performed in the present chapter comprising the identification of quality criteria from practitioners' perspectives, five design recommendations to be considered in future physiology sensing devices will be introduced in the following. These *Design Recommendations for Wearable Physiology Sensing Devices* contribute to the state-of-art research shaping the design and development of coming generations of physiology-aware technology.

Support the Quantified Self Approach Users provide large amounts of data, i.e. physical activity in terms of steps or physiological measures, such as heart beats, information about their sleep quality, weight, age, or calorie intake. In the consumer interviews it has been observed that they appreciated to be able to monitor their data. For example *C4* stated *"I like that I can see how many steps I made throughout the day, and also my sleep rhythm: when am I awake, when am I in my deep sleep phase and also control my weekly progress"*. This representative statement shows that users are keen to get a comprehensive overview of their activities referring to the so-called quantified self approach. This is also in line with findings from Packer et al. [193] stating that users value to monitor and annotate their activity data as strengthened by prior work [99]. Likewise, researchers can seize their curiosity in their own data to acquire participants. As *E3* said the individual monitoring represents *"a benefit for the participants if they wear the device. For example for the Fitbit, they see the analysis and so they are more interested to participate; they find it exciting"*. Thus, designers should support and even improve possibilities for users to monitor their data to also increase their interest in participating in scientific studies and *"learn something about yourself"* in return, as *C1* pointed out. The fact that learning is a strong motivation for users also to use self-tracking tools has been found in an extensive investigation by Choe et al. [40].

Increase Transparency of Data Recording Mechanisms According to almost all of the interviewed consumers, users noticed variations in what their devices were displaying. So, *C3* said that her device *"is accurate within approximately 100 meters- well this is pretty much okay. I'm satisfied with this"*. Also *C2* pointed out that the accuracy is not super good because *"the step count is higher after I moved my arms extensively"* referring to arm movements that are counted as steps. Consequently *C2* wished for *"an improved reliability"*. These statements highlight the fact that consumers often lack knowledge about how their devices work from a technical point of view. Although they *"do not exactly know how reliable [the device] is."* (*C4*), they notice that it is not working super accurately becoming obvious when *C1* admits that she *"can not be sure one hundred percent, whether I slept so long or whether I was really always awake when the bracelet shows I was awake"*. Consequently, it is necessary to support the user understanding regarding data recording mechanisms. Correspondingly, underlying processes determining the count and respectively the presentation of data should be made more transparent, so that the recordings are easier to follow. This is emphasized by *C5*'s questions, asking *"how will you know from the description how accurate/reliable a device is"*. This user need can be also found in the results from Niess and Woźniak [188], since 76% of their participants wished for a better understanding of their fitness trackers. By increasing the transparency of data recording mechanisms, this would not only strengthened the users' confidence in their devices, but also researchers would benefit from not including inaccurate data into their analyses. An exemplary solution could be to include simple and quick prompts for the user from time to time whether the tracked step count is correct or due to many arm movements in a short period, similar to the "Intelligent Recommender System Logging Stress" Visuri, Poguntke and Kuosmanen [266] proposed. Further, the user interface design can be used to visualize how certain the recorded data is. For example, color-based metaphorical visualizations could provide an indication regarding the reliability of the data due to difficult environmental conditions, such as very high or low temperatures.

Reduce Maintenance Effort The effort to put into operation and to initialize a wearable sensing device can differ enormously. Moreover, for users and researchers alike this is an important quality feature when dealing with devices. Whereas for researchers it can increase the complexity of a study when, e.g. the software needs to be synchronized so that data recordings can be accessed, for users the maintenance can be overburdening if it is too cumbersome. Like *C4* said *"if I had to set/operate so many things e.g. many buttons"*, it would be a deal-breaker for her. This observation is also made by *E3* who explains the important criteria for her *"that it [the device] is easy to start and stop and only limited interaction [with the device] is needed"* because this also facilitates the handling effort for researchers and further reduces the danger of making mistakes for both, researchers and consumers.

Provide Device-Status Feedback Another function that has been requested by both interview groups was that the wearable sensing device should provide feedback in case it is not working properly. For example, *E2* said that *"sometimes you do a long session, and then you realize a lot of data was corrupt, so it [the device] sensed incorrectly"* referring to what *E5* is also reporting: *"I do not want to realize in the afternoon that it stopped recording after half an hour"*. Consequently, *E2* thinks that *"it would be good if the device gives a warning that there is an issue with the data transmission"*. While from a research perspective it could have severe consequences, if the entire data recording is useless in case errors in the recording are noticed too lately, malfunctions are also annoying for users. As *C1* states: *"I once had the problem that the watch did not work for a while"*. Therefore, it would be beneficial if e.g. visual cues provide feedback instantly in case the device has stopped recording data. But also for customers such a function could yield advantages because according to Canhoto and Arp [33] to feedback the ability to record data is a crucial factor for adopting the device and therefore prevent abandoning the fitness tracker early. This requirements is not new and has been recommended by Jakob Nielsen [187] making the "*Visibility of systems status*" the first usability heuristics for user interface design. This emphasizes the crucial importance direct system feedback has.

Increase Resistance against Environmental Factors This final recommendation was inspired by the consumers experiences on wearing their wearable also while showering and when going to bed. For example, *C1* explained that *"this model is convenient for me because I can wear it all day. I also do not have to take it off when I take a shower or whatever"*. Interviewee *C5* mentioned that she likes that her device is water-proof because that does not restrict her. This highlights that consumers prefer a high resistance against environmental circumstances, such as heavy exposure to water or sunlight for their devices. Likewise, increasing the resistance of a device supports reliable recordings. As *C1* points out: *"This way [by wearing the device also under the shower], I also really track entirely and do not have to take it off and then put it on again and thus half of the information that might be important is lost"*. Accordingly, researchers as *E3* mentioned that the device should bear *"no restrictions in what you are doing usually"* to make them *"work in the real world for a normal user"* like *E2* pointed out. This requirement is in line with the findings from Canhoto and Arp [33] who identified the ability to *"use their wearables anytime and anywhere, and capture data consistently"* (cf., P.31,32) as a key component for establishing sustainable usage of a device. Thus, the design principles referring to the form factor should be designed in a way, such that users do not feel restricted because this also additionally allows unadulterated data acquisition for researchers representing real world usage patterns.

4.5 Chapter Summary and Conclusion

In this chapter, the results from ten expert interviews, five experienced researchers in affective computing and five physiology sensing devices consumers classified as "intensive users", have been reported. It has been described what methodological approach I applied for the interview analysis, as well as the question set targeting both, previous findings and new aspects plus open questions. Finally, from the qualitative results there have been derived six main dimensions complemented by 17 sub-dimensions summarized in Figure 4.3 forming the *Design Space for Physiological Measurement Tools*.

This chapter has been built upon the main problem identified in Chapter 3 showing that physiological sensing bears some restrictions when being used, e.g. in mobile settings. By interviewing two distinct groups of practitioners of physiological sensing technology, these constraints have been examined from different perspectives validating what has been introduced in prior work [98]. Thus, **RQ1c** has been answered to in detail by asking researchers and consumers about their experiences with physiological sensing devices in respect to the mentioned disadvantages or limitations. Referring to **RQ1d** I have presented the qualitative data analysis based on the inductive thematic analysis approach, which has been applied to identify representative dimensions summarizing important criteria. As the main contribution to target the research gap of providing an overview on the distinct criteria ans potential issues involved in the decision of using physiological sensing technologies, the *Design Space for Physiological Measurement Tools* has been developed. By this tool the great potential referring to its versatility and utility, as well as the limitations physiological measurements can have, are visible as one glance. Being aware of such restrictions and reflecting consciously on the use case, the user group, and the design goals of interactive systems incorporating physiology sensing, is a fundamental prerequisite for their success. Thus, the detailed investigation of physiological measurement capabilities and imperfections answered throughout the research questions **RQ1a**, **RQ1b**, **RQ1c**, and **RQ1d** has been made subject to extensive research reported on in Chapter 3 and Chapter 4. This work builds the foundation for the subsequently presented investigation of how to design stress-aware interactive systems.

Chapter 5

Exploring Tactile Feedback in Relation to Stress

In the past two chapters, the physiological nature of stress responses has been addressed. As has been discussed, knowing when someone is stressed can be understood as a prerequisite for designing interventions helping to reduce stress. As one potential solution to help coping with stress responses, the presentation of feedback, and particularly visual feedback has been explored [100, 173, 290]. In this context, related work often has studied the efficiency of the application being designed provide upon stress. In practice, Matthews et al. [173] found that presenting a leading feedback not being controlled by the participants helps users to relax more effectively than real-time feedback. Likewise, the combination of different sensory modalities has been also examined [276]. Accordingly, Klamet, Matthies, and Minge [140] provided five different feedback types, namely three tactile types and one visual and auditory each, at distinct body locations. From their findings they concluded that tactile feedback is the most promising modality to be further explored. Focusing on the tactile sense is also reasonable from a practical point of view because both, the visual and the auditory sense, are needed to navigate which does not require much mental effort, but still limits the possibilities to use them appropriately for providing fine-grained feedback on stress states. Moreover, recent technology makes use of these channels providing visual or auditory feedback. For example, smartphone users can switch between being notified about calls and messages through blinking LEDs, beeping tones, or vibration. Nevertheless, tactile feedback bears several advantages over its visual and auditory neighboring senses: First, it preserves the privacy of its' users which is a particular benefit when thinking of feedback about stress states. Additionally, the skin- as the largest sensing organ [141] -provides multiple locations where tactile feedback can be applied. And finally, also users with visual or auditory impairments can rely on tactile feedback since it works independently from other sensory channels. All these reasons provide a solid explanation why previous work has also explored widely tactile stimulation as an opportunity to provide unobtrusive feedback. Hence, pressure-based as well as thermal stimuli's suitability for interacting with mobile devices has been explored previously [67, 275].

Consequently, the present chapter will provide an overview over previous work focusing on the investigation of feedback employing different modalities, namely the visual, auditory, and tactile stimuli aiming to answer which feedback modality suits best for notifying about stress state referring to **RQ2a**. In contrast to prior studies, mostly investigating the effectiveness of feedback providing applications, my research focus is the communication of stress states in dependency of the tactile perception experienced by the users. Thus, the subsequently two presented works explore the characteristics of a suitable stimulus aiming to answer **RQ2b**, how an effective feedback stimulus could be designed. For this, I conducted two user studies. While the first study explores the effect of pressure-based stimulation in comparison to vibro-tactile feedback, in the second work I focused on thermal feedback- a less familiar variant of tactile stimulation -which has been subject to prior research [94, 280]. As part of the latter study, I additionally collected qualitative data on the desirability of stress feedback. The findings form both user studies provide interesting insights on the suitability of tactile feedback for notifying users about their stress states.

This chapter is based on the following publication:

- R. Kettner, P. Bader, T. Kosch, S. Schneegaß, and A. Schmidt. Towards pressure-based feedback for non-stressful tactile notifications. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, pages 89:1–89:8, New York, NY, USA, 2017. ACM
- R. Poguntke, J. Ilk, A. Schmidt, and Y. Abdelrahman. Designing thermal feedback for notifying users about stress. In *13th EAI International Conference on Pervasive Computing Technologies for Healthcare - Demos and Posters*, PervasiveHealth 2019. EAI, 6 2019

5.1 Related Work

Tactile feedback has been subject to extensive research given it's main advantage of unobtrusiveness. In the following, I will given on overview over related work focusing on the exploration of tactile stimulation patterns, and particularly thermal feedback against the background of investigating the suitability of such feedback.

Tactile Stimulation For consumer devices, for example smartphones or wrist-worn wearables vibrotactile feedback have been used extensively to convey information or just draw the attention to something. Interestingly, vibration has been found to have a pain relaxation enhancing function [163, 192]. Besides, tactile stimulation can take various forms, such as tapping, dragging, squeezing, and twisting [183, 255]. Moreover, pressure can be applied also as an input source. Wilson, Brewster, and Halvey [277, 278] for instance, found that mobility has a significant negative effect on controlling pressure. Moreover, Feng et al. [67] explored how pressure input could be used to interact with mobile phones. The advantage of pressure is that it can be perceived easily by the peripheral receptors [171]. Additionally, pressure-based feedback yields advantages such as unobtrusiveness [210]. Therefore, Alvina et al. [2] explored spatiotemporal vibrotactile patterns on different body parts and confirmed it's recognizability, while others focused more on its functionality as an alternative feedback mechanism [211, 270, 295]. There are also other approaches using Electrical Muscle Stimulation (EMS) [237] or changes in temperature [253] addressing the thermal perception. Particularly the latter approach seems promising, since thermal stimulation patterns are able to gain the user's attention quickly but also unobtrusively.

Applications of Thermal Feedback A specific form of tactile stimulation is thermal feedback. Hereby, either cold or warm stimuli are applied. Humans can perceive those stimuli as long as they do not lie within their individual neutral zone that comprises a temperature range between 6°C and 8°C [126], which does not allow to sense temperature changes. Further, the perception is experienced differently among body locations depending, for example on the density of hair or skin thickness. Often hairless and thin-skinned parts, such as the palm or lower side of the wrist are chosen for the presentation of thermal stimuli since there the sensation is more likely to be perceived. However, more and more research is concentrating on the exploration of suitable locations to apply thermal feedback [140]. But not exclusively the body part where the stimulus is being presented affects it's perception, also other factors, namely the ambient temperature and in the humidity in the environment [93], as well as the influence of fabrics used in clothing [94]. A possible field of usage for thermal stimulation applications, was also proposed by Wilson et al. [274, 275], who demonstrated how visual and audio media content, e.g. in form of widgets, could be thermally augmented. In the context of mobile interaction, Wilson et al. [281] evaluated the suitability of different thermal stimuli in stationary and mobile settings to determine a suitable feedback designed for its application in mobile devices. Thinking of the emotional connotation colors have, Wilson et al. [282] could show in a combination of an online questionnaire and a lab study, that warm temperatures are subconsciously linked to insecure webpages. The fact that distinct temperature ranges can be mapped to different affective meanings [280, 276] was also exploited in a study by Wilson, Davidson, and Brewster [279] who found that the interpretation of thermal stimuli, for example in social media activity leads to conclusions on the social interaction style and even personality traits. Given the enormous potential tactile feedback, and particularly thermal stimulation has, the present work builds upon prior research [140, 167, 281] and explores how an effective feedback stimulus could be designed to provide suitable feedback on stress states.

5.2 Exploring Pressure-based Tactile Stimulation

As pointed out in Section 5.1, prior research has been occupied investigating novel application scenarios for pressure-based feedback. In contrast, the present work focuses on the exploration of eefects on the user when receiving such stimulation. Monitoring the stress level using physiological sensors,the aim was to test whether pressure-based tactile feedback could be used as a potential notifier about stress states. For this, two user studies have been carried out: First, a preliminary evaluation of tactile feedback stimuli (vibrotactile and pressure-based) has been performed; Second, a user study on the physiologically and subjectively observable effects of pressure-based feedback under stress has been conducted.

5.2.1 Tactile Stimulating Prototypes

Two prototypes of wrist-worn wearables (see Figure 5.1) have been built. First, a wristband capable of providing vibrotactile feedback as known from fitness trackers and smartwatches has been developed. Second, a wristband with a novel type of tactile feedback leveraging pressure-based feedback similar to the work of Pohl et al. [210] has been built. Both wristbands had the formfactor of watch-straps (approx. width: 2.5 cm; length: 30 cm). Jeans fabric was used on the outside and an elastic fabric coated the inside. Each wristband was filled with a bicycle tube, which was cut to the right length. Finally, each wristband was vulcanized on both ends and connected to an Adafruit Metro Mini 328 to trigger the feedback.

Pressure-based Stimulation Wristband The pressure-based wristband fills the tube with air which in return applies pressure to the user's wrist. To infuse the pressure wristband with air, a pressure pump and valve from a disassembled AEG BMG 5611 blood pressure meter[8] were used. Both were attached to the bicycle tube inside the wristband. Accordingly, feedback was applied by filling the wristband with air.

Vibrotactile Stimulation Wristband The vibrotactile wristband contains ten shaftless vibration motors. To maximize wearing comfort, each vibration motor was attached to a small 3D printed case with a slight curvature towards the wrist. The motor cases were loosely connected via threads to remain flexible and keep them at a constant distance of 20 millimeters. The vibration motor assembly was inserted into the bicycle tube, so users could not directly feel the motors on their skin but perceive the tactile stimuli through the thin fabric layer.

[8] www.etv.de/products/en/Health-Care/Blood-pressure-gauge/AEG-BMG-5611.html

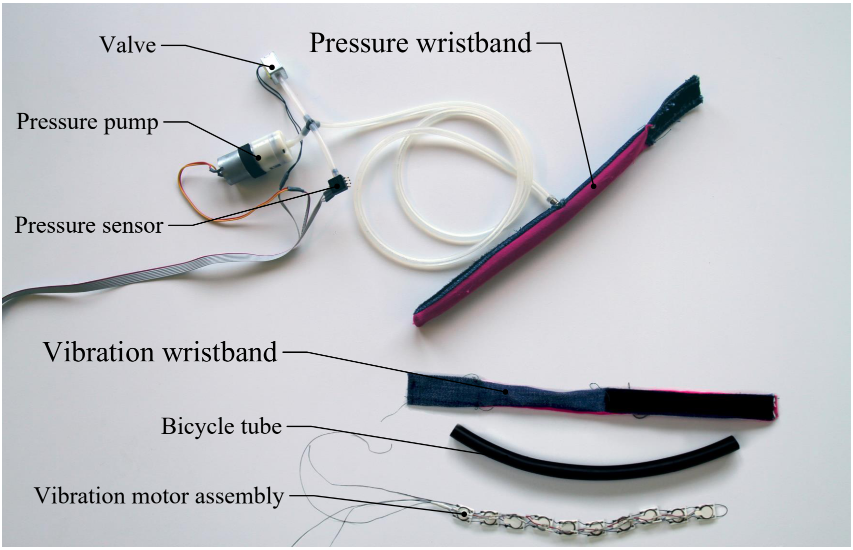


Figure 5.1: Assembled pressure-providing prototype (top) depicted with pump, valve and pressure sensor. Partly assembled vibration-providing prototype (bottom-right) with bicycle tube (black) and vibration motor assembly (white) shown separately.

Feedback Types The prototype-Arduino connection allowed to precisely control when and how much pressure or vibration was applied. Two signals patterns similar to physiological signals of humans were repeated periodically. The first signal was derived from the pulse. A single impulse is given in every time frame. In contrast, the second signal was derived from the human heart beat. It consisted of two consecutive short stimuli. The rate of repetitions was determined by the user's resting heart rate (i.e., number of heart beats per minute), which was recorded before the signal was applied.

5.2.2 Preliminary Evaluation of Tactile Feedback

Prior work outlines that even minor adjustments in frequency, intensity, and the feedback pattern itself can lead to different feedback perceptions [29]. Therefore, a preliminary study exploring comfortable feedback patterns for both, vibrotactile

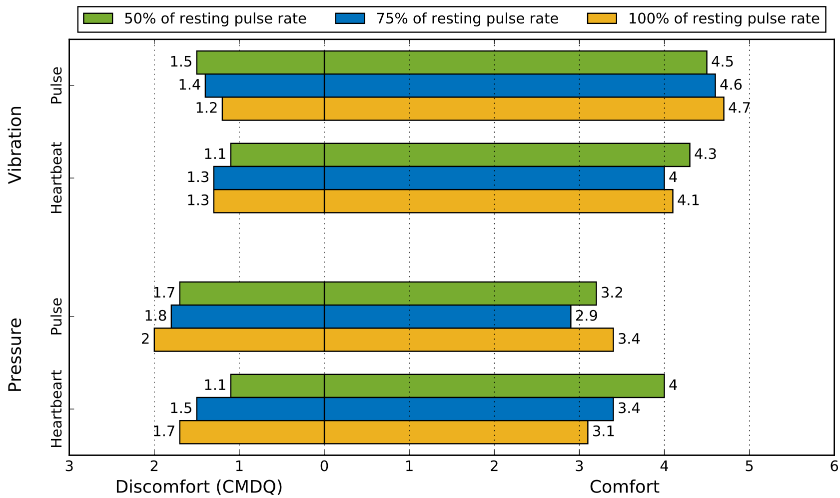


Figure 5.2: Mean values on the perceived comfort and discomfort using the Cornell Musculoskeletal Discomfort Questionnaire [108]. The heartbeat pattern with 50% frequency has been perceived most comfortable for pressure. For vibration the simple pattern has been perceived slightly more comfortable. For both, pressure and vibration, the heartbeat pattern with 50% frequency has been perceived least discomfortable.

and pressure-based stimulation was conducted to determine the most comfortable stimuli being later used in the main study.

Measures Determining the most suitable feedback pattern, discomfort using two items depicted in Figure 5.2 from the Cornell Musculoskeletal Discomfort Questionnaire (CMDQ) [108] was measured. Furthermore, the participants were asked how comfortable they perceived the tactile stimulation on a Likert item scale ranging from 1 (not comfortable at all) to 6 (very comfortable).

Participants and Procedure 10 participants aged between 22 and 42 ($M = 29.2$, $SD = 6.4$) were recruited via personal acquisition. In a within-subject design two feedback methods (pressure and vibration), two feedback patterns (pulse and heartbeat) and three feedback repetition rates were used as independent variables balanced according to Latin Square. The rates were calculated by taking 50%, 75%, and 100% of the resting heart rate which was measured with a MPXV5050GP pulse monitor for each participant before presenting the different patterns. After attaching the prototypes to participants' left hands, the 12 different tactile feedback stimuli for 10 seconds intermittent by five seconds break were then applied.

5.2.3 Results

The results depicted in Figure 5.2 show that the heartbeat pattern at 50% resting heart rate had been perceived most comfortable for pressure ($M = 4$). For vibration also this pattern had been rated slightly more comfortable ($M = 4.3$) however not as comfortably as the pulse pattern at 100% resting heart rate. For both, pressure and vibration, the heartbeat pattern at 50% resting heart rate had been perceived least discomfortable ($M = 1.1$ each) (see Figure 5.2). Due to its high subjectivity in the ratings, the aim was to find a compromise between 'most comfortable' and 'less discomfortable', hence the latter tactile stimulation pattern was chosen as the final pattern being applied in the main study.

5.2.4 User Study

In the main study, the effect of different feedback methods, namely pressure-based feedback compared to vibrotactile feedback as well as to no feedback, which served as our control condition, on the user's stress level has been investigated. Thereby, three different stress levels (i.e, easy, medium, and difficult) have been elicited. Physiological and subjective data have been assessed as described in the following.

Measures and Stress Elicitation Task As subjective measures, the Short Stress State Questionnaire (SSSQ) [110, 111] has been used. The SSSQ contains two different parts: First, 24 items that assess the stress level during a baseline period; Second, another 24 items measure the perceived stress asking about the participant's feeling while a task has been performed.

EDA rate which indicates the activation of sweat glands related to activation in the sympathetic nervous system has been recorded during the study. An increase in the EDA indicates an increased stress level as has been shown in prior work [79]. To investigate the influence of tactile feedback on the participants' stress level, a verbal MAT [18] had been used. In this task, participants count verbally backwards in steps of seven which proved effective in previous research [104, 160, 162, 245, 262].

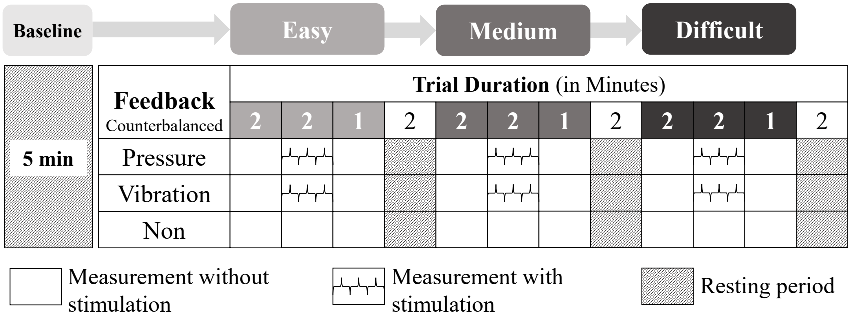


Figure 5.3: The study design consisted of three trials for each participant, ranging from easy over medium to difficult. For each difficulty level all three feedback methods were applied in one trial each and in counterbalanced order.

For the MAT, three different ranges of difficulty to elicit ascending stress levels had been chosen, presenting random numbers: easy (two-digit numbers), medium (three-digit numbers), and difficult (four-digit numbers). From the given starting number the participants needed to count backwards in steps of seven.

Participants and Procedure 14 participants (6 female, 8 male) aged between 20 and 30 ($M = 25.4$, $SD = 3.3$) have been recruited via university mailing lists. Upon arrival in the lab, the study purpose was explained and all participants filled in an informed consent form as well as demographic questions. Afterwards, two sensors were attached to the participants' index and middle finger tips recording EDA, for which a Mindmedia Nexus biofeedback kit 4 [9] was used.

Feedback was provided to the participants by showing the current score on a 17" display. For each correct answer, the score was highlighted in green color and increased by 10 points, whereas the score was highlighted in red color and was penalized by not increasing in case of a mistake. Three different sessions were performed with an increasing the difficulty level from easy over medium to difficult. It was deliberately decided to not counterbalance the difficulty levels' order for preventing carry-over effects.

Each session for a specific stress level consisted of three different trials (see Figure 5.3 inspired by the study design by [175, 267]. In each trial, one feedback method (pressure, vibration, no feedback) was applied in counterbalanced order according to a Latin Square. The study took about 90 minutes including a

[9] http://www.mindmedia.info/CMS2014/en/products/systems/nexus-4

Feedback	Difficulty Level		
	Easy	Medium	Difficult
Pressure	-11.1	-16.6	-14.7
Vibration	-10.9	-15.0	- 15.7
No Feedback	-11.5	-13.6	- 17.0

Table 5.1: Mean results for the Short Stress State Questionnaire showing the individual change score among all participants divided into the three stress levels easy, medium, and difficult. Since positive values indicate stress, the results suggest that approximately an equal amount of stress has been subjectively perceived in easy conditions. Whereas for the stress levels medium and difficult, no intervention and for the latter pressure have been rated most stressful.

5-minute baseline trial at the beginning and nine 5-minute trials (3 feedback types $\times$ 3 stress levels) intermittent by 2-minute breaks. During the study, a 60 seconds-countdown was shown on the display. After 60 seconds, the participant was told a new number to continue with the MAT until the trial was over.

5.2.5 Results

In the following, the results of the main study embracing subjective and physiological data are described.

Subjective Data The results for the SSSQ [110] are depicted in Table 5.1. They show that during the easy trial there was hardly any difference in the subjective stress ratings. In the medium level, participants rated the condition where no feedback was applied as most stressful. Pressure-based feedback was perceived slightly more stressful in the difficult level.

Physiological Data The EDA values of the 2-minutes stimulation phase were averaged for each participant and thereby calculated one mean value for each participant. The results show that EDA increases during the stimulation phase compared to the baseline. An overview of this increase is depicted in Figure 5.4. Comparing the rises of the different feedback patterns among all participants, it can be observed that providing no feedback at all has the lowest deviation from the initial baseline measurement ($M = 11\%$ easy, $M = 15.9\%$ difficult) followed by pressure feedback ($M = 13.5\%$ easy, $M = 17.4\%$ difficult). Vibrotactile feedback showed the highest deviation from the baseline and therefore the highest increase

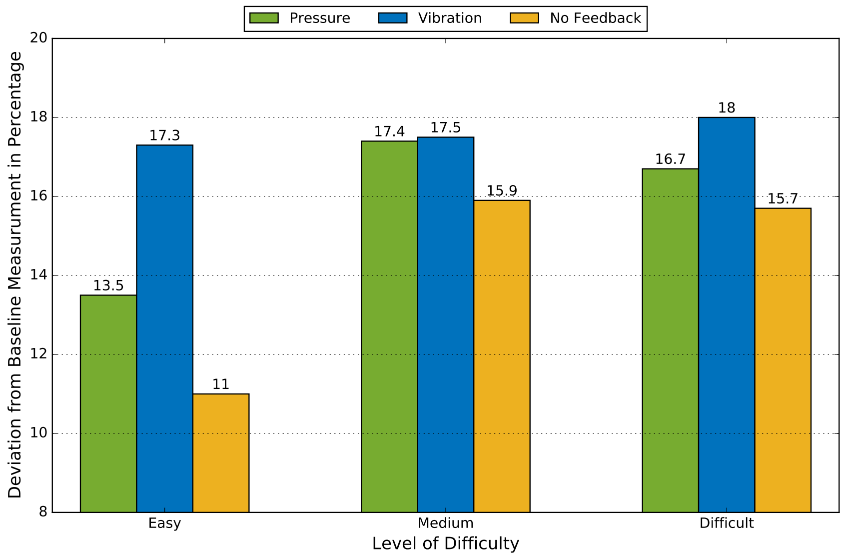


Figure 5.4: The deviations from the baseline measurements during the stimulation phase according to the three distinct levels of difficulty. High values indicate increasing electrodermal activity which signifies increased stress.

in EDA ($M = 17.3\%$ easy, $M = 18\%$ difficult). A two-factor ANOVA did not reveal statistically significant differences for the tested tactile stimulation and for the difficulty level on EDA.

Looking at the medium and high stress level, vibrotactile and pressure-based feedback slightly increases the EDA compared to not providing feedback. In contrast, the vibrotactile feedback elicits the greatest increase of EDA for the low stress condition which was induced by the difficulty level 'easy' compared to the other feedback patterns. Hence, it becomes obvious that under low mental stress the effect of vibrotactile is greater considering the low EDA value for no feedback in the same condition. Over all three conditions it can be seen that stress is almost constantly high when vibrotactile feedback is provided. In contrast, pressure-based feedback leads to lower EDA values when the user is put under low mental stress.

5.2.6 Study Conclusion

In the presented study the effect of pressure-based feedback in comparison to vibrotactile on the user's stress level has been investigated. For this physiological and subjective data have been recorded from 14 participants. The results indicate that vibrotactile feedback has a greater impact on the user's stress level, particularly in low-stress situations, compared to pressure-based feedback. This has been signified by the electrodermal activity data on the perceived stress. To sum up, the evaluation of the effects of pressure-based feedback as a potential stimulation for notiying users about stress showed to be only slightly better than vibrotactle stimulation being already familiar with from smartphone or wearable notifications.

5.3 Thermal Feedback for Notifying Users about Stress

As has been pointed out in the beginning of this chapter, tactile stimulation pattern are regarded to be a promising alternative compared to visual or auditory feedback. Hence, previous research has been occupied investigating pressure-based and vibrotactile stimuli for different use cases [93, 275]. Despite their various advantages, they also bear some limitations, particularly when it comes to unobtrusiveness. Given the sounds these tactile feedback methods make when being applied in practice, potential notifications about the user's stress state could be also noticed by bystanders. Additionally, the results from the previously presented work Section 5.2 indicate that both patterns, vibrotactile and pressure-based, affect the user's stress level. For this, in the present study I examine the suitability of thermal stimulation due to its novelty for average-users. Targeting to answer **RQ2b**, there will be tested different temperatures, rates of change, and body locations for designing an effective feedback stimulus which is able to notify users about their stress state.

5.3.1 User Study

The aim of the user study was to explore thermal feedback for tactile stress notifications according to user preferences. For this, five *temperature levels*, three

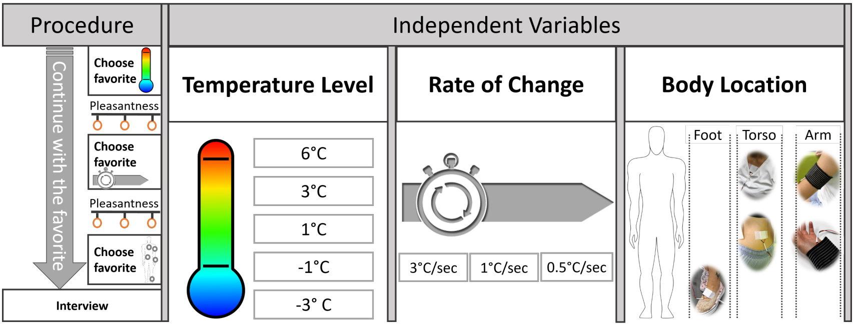


Figure 5.5: Study design illustrating the procedure and different measures used to evaluate the suitability of thermal stimuli.

rates of change, and five distinct *body locations* have been compared against each.

Study Design and Variables Evaluating different *temperature levels*, *rates of change*, and *body locations*, these were the independent variables. Five temperature levels (-3°C, -1°C, 1°C, 3°C, 6°C) and three rates of change (± 1°C/s, ± 3°C/s, ± 0.5°C/s) were compared and rated before five body locations (wrist, upper arm, lower back, upper chest, foot arch) have been tested with the preferred stimuli. All of them were presented to the participants in randomized order but following the same sequence, namely *temperature levels* first, then *rates of change*, and lastly *body locations* (cf., Figure 5.5). A within-subject design was applied. While pleasantness for the temperature and rate of change should be rated, perceived discomfort and interference had been measured for the body location evaluation. Hence, these were the dependent variables for evaluating specifically the preference of thermal feedback stimuli.

Hardware Prototype A hardware prototype (see Figure 5.6) was built. It provides thermal feedback by conducting heat through a Peliter element (TEC1-12706) of 40mm × 40mm × 3.9mm (cf. Figure 5.64). The Peliter element was connected to a L298N motor controller (cf. Figure 5.62) to control the voltage and current direction, hence control the rate of temperature change and values. The motor controller was powered by a DC power supply at constant voltage of 15V. The motor was controlled by an Arduino UNO. The Arduino UNO was connected to a laptop, which ran a controlling software program written in C++. This program adjusted the temperature value of the Peltier element.

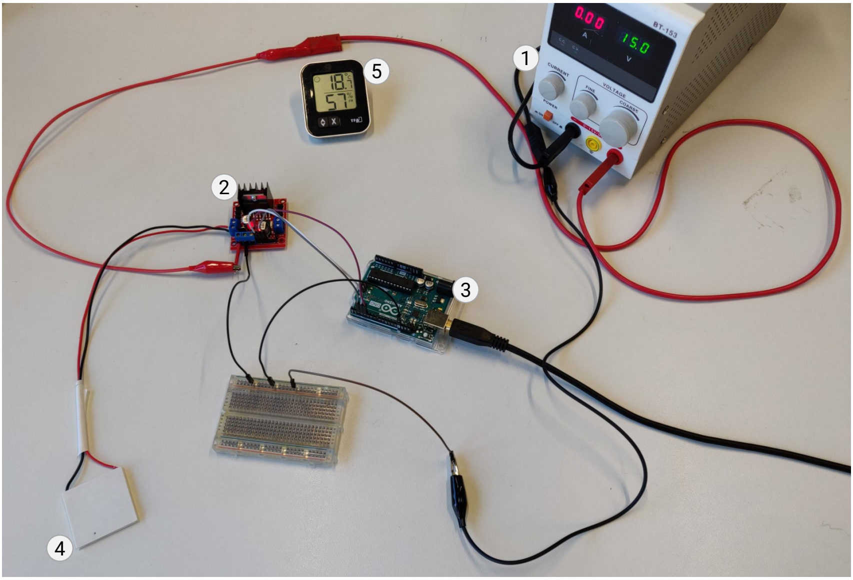


Figure 5.6: Hardware prototype providing thermal feedback consisting of DC power supply (1), motor driver (2), Arduino (3), peltier stimulator (4), thermometer, and hygrometer (5).

Questionnaires For assessing pleasantness, the participants have been asked to rate "*How pleasant would that temperature level be*" on a five-point Likert scale ranging from "Very unpleasant" (=1) to "Very pleasant" (=5). This question was introduced by the sentence "*If you imagined receiving the thermal feedback during a stressful situation*". This scale had been used for both, the *temperature level* and the *rate of change* evaluation. For comparing the discomfort and interference with the ability to work among the five body locations, two out of three items from the CMDQ [108] have been taken. The first question was "*If you experienced ache, pain, discomfort, how uncomfortable was this?*" being rated on a three-point Likert scale (1="Slightly uncomfortable", 2="Moderately uncomfortable", 3="Very uncomfortable"). The second question also contained a three-point Likert scale but provided the answers "Not at all" (=0), "Slightly interfered" (=1), and "Substantially interfered" (=2).

Qualitative Questions Additional to the questions asked to compare the feedback parameters, a short semi-structured interview was conducted after the study procedure. For this, the participants were asked to comment on whether they would perceive thermal feedback as helpful in stressful situations and what they generally thought of thermal feedback. Further, I was interested in why they had chosen the preferred *temperature level*, *rate of change*, and *body location* and if they could imagine other body locations to be appropriate for providing thermal stimulation. Lastly, they they were asked to consider and name advantages or disadvantages of thermal feedback for stress notification, particularly when compared to other feedback methods.

Participants and Procedure In total 21 participants (9 females) with a mean age of 23.4 years ($SD = 3.5$) have been recruited. They were acquired via university mailing lists and personal contacts. After the participants were explained the study purpose and given the consent form reassuring that they agreed to have their answers audio recorded, they were asked to imagine a stressful situation which elicits stress specifically for themselves. Then the Peltier element was attached and fixed with a bandage having a velcro strap while collecting demographic data. Subsequently five different *temperature levels* were presented and after each stimulus being presented the participant was asked to rate the perceived pleasantness. For assessing the favorite *rate of change*, the most liked temperature level the participant had previously rated was taken and continued with for evaluating each of the three rates of change. Then the participant was asked to attach the Peltier element to the five different *body locations* successively intermitted by the specific ratings of discomfort and interference for each body location. Hereby, again the *temperature level* and *rate of change* they had rated best before was presented at the distinct body locations evaluating the perceived discomfort and interference for each body part right after its presentation. Before the participants were thanked, a short semi-structured interview was conducted. The study took approximately 30 minutes for each participant and they were compensated with sweets.

5.3.2 Results

In the following, the evaluation results are presented leading to find the preferred thermal feedback for notifying users about stress.

Quantitative For the *temperature level* the rating suggested that cold stimuli were perceived more pleasant ($M = 3.71$, $SD = 0.98$) than the warm temperature levels ($M = 2.75$, $SD = 0.92$). While the most pleasant stimuli, -1°C was rated highest ($M = 3.95$, $SD = 0.86$), the least pleasant was also the warmest, 3°C ($M = 2.48$, $SD = 1.29$). The exact ratings for the other stimuli can be obtained from Table 5.2. Comparing the three *rates of change*, it was found that the

	Temperature Levels	**Body Locations**		
	Pleasantness	*Discomfort*	*Interference*	
6°C	2.52 (1.16)	0.33 (0.58)	0.45 (0.6)	**Wrist**
3°C	2.48 (1.29)	0.48 (0.75)	0.62 (0.59)	**Arm**
1°C	3.43 (0.98)	0.33 (0.77)	0.33 (0.69)	**Chest**
-1°C	3.95 (0.86)	0.29 (0.72)	0,48 (0.75)	**Back**
-3°C	3.48 (1.33)	0.37 (0.6)	0.47 (0.6)	**Foot**

Table 5.2: Results (means and standard deviations in parantheses) for the *temperature level* pleasantness rating and the evaluation of *body locations* according to discomfort and interference with the ability to work. Before taking the preferred one for comparing it among all five body locations, each participant received each temperature level.

lower rates were preferred. Accordingly, ± 0.5°C/s was rated highest ($M = 3.86$, $SD = 1.11$) followed by ± 1°C/s ($M = 3.57$, $SD = 0.93$) and ± 3°C/s ($M = 3.14$, $SD = 1.11$). Since the preferred stimuli have been taken individually for each participant and continued with throughout the remaining study procedure, it became obvious that five participants chose the warm temperature and the rest preferred to succeed with the cold stimuli; to be precise nine participants decided for a temperature level of -1°C. Regarding the preferred *body location* for presenting thermal feedback, participants preferred the lower back, meaning they considered it the least discomfortable ($M = 0.29$, $SD = 0.72$). The other body locations were rated similarly (cf. Table 5.2). Five participants did not feel any discomfort when having presented the thermal feedback at any of these locations. Additionally, it was assessed whether the wearers felt any interference with their ability to work when receiving the thermal feedback. While mostly no or a very light inference was reported for the locations, the upper arm was rated highest ($M = 0.62$, $SD = 0.59$) regarding such an interference and five participants particularly commented on this location when being asked about a possible interference. The exact ratings are depicted in Table 5.2.

Inferential Statistical Analysis A Friedman test was applied to test the effect of *temperature levels* on the perceived pleasantness indicated by the Likert scale ratings. A statistically significant difference ($X^2(80) = 30.88, p < 0.0001$) was found for this. When comparing the cold and warm stimuli of the same temperature level, based on Wilcoxon signed-rank test there was no significant difference ($Z = 27.5, p = 0.56$). In contrast another Wilcoxon signed-rank test between *temperature levels* -3°C and 3°C showed a significant difference ($Z = 25.5, p = 0.026$) between the perceived pleasantness of the thermal stimuli. Further, the effect of the *rate of change* on the pleasantness ratings was tested. The Friedman test suggested a statistically significant difference ($X^2(40) = 7.12, p = 0.03$). A post-hoc analysis applying a Wilcoxon signed-rank test indicated that there is no statistically significant difference in the pleasantness rating between the *rates of change* of ±0.5°C/s and ±1°C/s, nor between ±1°C/s and ±3°C/s. However, a Wilcoxon signed-rank test between the pleasantness rating for the *rates of change* of ±0.5°C/s and ±3°C/s revealed a significant difference ($Z = 17.5, p = 0.01$). A Friedman test to compare the discomfort rating on the Likert scale between the tested *body locations*, did not indicate a statistically significant difference ($p = 0.5$). Another Friedman test to compare the interference rating on the Likert scale between the different *body locations*, did not show statistically significant differences ($p = 0.82$) alike.

Qualitative Results On the general idea of being notified by thermal feedback on one's stress level, 18 out of 21 commented positively. Only two participants stated that they did not see the benefit from thermal feedback and one participant even criticized it to be unpleasant in stressful situations. When being asked particularly on their considerations about thermal feedback signaling stress, three interviewees said they would appreciate such feedback in case it is not distracting in the very moment. Regarding the question why the participants chose the *temperature level*, *rate of change*, and *body locations* as they did, 16 participants explained that they usually feel hot when being stressed and thus, the counter stimulus, namely something cold, might help them to cool down and respectively to cope better with the situation. Another interesting explanation was given by one interviewee stating that he perceived the cold stimulus as unusual in comparison to the heated devices he wears on his body, such as the smartphone and smartwatch. For choosing the *rate of change*, three participants preferred a low rate to avoid being distracted by a faster change. In contrast, five participants worried that having a less intense *temperature level* or a low *rate of change*, they would not notice the thermal feedback under stress and consequently be irritated because they expected it then. According to the preference for certain *body locations*, one third would appreciate to receive thermal feedback at the wrist. Four of them wore a smartwatch and loved the idea to integrate thermal feedback in existing wrist-worn wearables. The second favorite location being deduced from the interview results was the upper arm, as chosen by three interviewees. The upper chest and likewise the foot were preferred by one participant each. Almost half of our sample did not state any preference on the body location, while one participant explicitly disliked the lower back. As an advantage of thermal feedback, eight participants stated that such feedback preserves privacy, which becomes even more important when signaling stress. Another two mentioned the advantage that other feedback channels, namely audio or vibration are already occupied by other devices and therefore thermal feedback has a unique value making it easy to identify its source and meaning. One additional remark was referring to the quality of unobtrusiveness of thermal feedback; one participant said that it felt like more "*direct*" coming from his own body and not being induced externally.

5.3.3 Study Conclusion

By this work the general suitability and preferences of thermal feedback for the specific use case of notifying users about their stress level has been investigated. For this, three different measures, namely *temperature level*, *rate of change*, and *body location* have been evaluated and rated according to their subjectively perceived pleasantness, discomfort and the interference to work. For each of those variables distinct stimuli have been compared in a user study involving 21 participants. From both, the quantitative and qualitative results, it can be obtained that users appreciate the idea of thermal feedback as a notifier about stress given its advantages of privacy preservation and unobtrusiveness. Further, they were interested in incorporating such a stimulation pattern in existing technologies, such as smartwatches which are already used heavily. Additionally, the quantitative ratings revealed that cold stimuli, in particular -1°C with a change rate of ±0.5°C/s are preferred for applying thermal feedback. Regarding the best *body location*, the participants' feedback was ambiguous. While most interviewees liked the possibility of having familiar locations, namely the wrist, they feared it could interfere with their ability to work on the other hand. The upper back seemed to avoid this issue and support the privacy quality being mentioned as an considerable aspect.

5.4 Discussion

Suitability of Pressure-based Tactile Stimulation Throughout the two consecutively performed research experiments the suitability of tactile feedback have been explored for the dedicated purpose of providing feedback about stress states. In the first study the focus has been laid on the investigation of effects when presenting classic tactile stimulation pattern, here vibrotactile and pressure-based feedback. From the results of the electrodermal activity

values, it can be obtained that the user's stress level greatly increases as soon as vibrotactile feedback is presented. This observation is in line with findings from related work [92] showing that vibrotactile feedback had been perceived more mentally demanding. In contrast, the results suggest that pressure-based tactile stimulation is particularly useful in situations where the overall stress level is rather low. However, although the pressure-based feedback indicated lower ascents of the physiologically traceable stress level, the difference has not been shown to be significant. This finding is also supported by prior research [211] revealing that both stimulation patterns are perceived similarly when asking for their annoyance and comfort. Further, the subjective ratings of the participants' stress level (cf., Table 5.1) suggest that the individual perception of one's stress is not significantly affected by the presentation of tactile stimulation. When asking the participants informally about their opinion on the two distinct feedback patterns, they reported that the study setting as such had been perceived very stressful. Therefore, the physiologically indicated stress could be evoked by the experimental setting additional to the stress-inducing mental arithmetic task. This explains why the subjectively perceived stress level had been high in general, even when no stimulation pattern has been presented (cf., Table 5.1).
Referring to how the feedback stimulus should be designed, the preliminarily conducted study revealed that participants perceived stimuli presented at a lower rate more comfortable. In practice, 50% of the resting heart rate had been rated more comfortable ($M = 4.0$) than 75% ($M = 3.4$) and 100% ($M = 3.1$) of the resting heart rate at average. Thus, feedback notifying the user about one's stress level, should follow a slow rhythm not imitating the heart beat when wanting to consider the perceived comfort. Nevertheless, comfort is being sensed highly subjectively. As addressed in Chapter 4, it is a broad term which implies different perspectives and what makes it difficult to find generalizable guidelines. According to comfort, Pohl et al. [211] investigated the suitability of tactile stimulation with respect to a longer usage period, namely in an 1 hour-experiment. During that time of mobile usage, participants reported that they had not been "*feeling inhibited or annoyed by the device*".

Properties of a Thermal Stimulus for Design Suitable Feedback
Based on the findings from the first study, I focused on thermal sensation for notifying users about their stress level, since the pressure-based feedback seemed to not provide a huge benefit in comparison to classic vibrotactile feedback. As known from prior work [279, 281] thermal feedback perception is highly subjective and includes various variables, such as the *temperature level*, *rate of change* as well as the *body location* where the feedback is being presented. Hence, the aim of the latter work was to identify users' preferences for these main influencing factors building a basis for the design of a suitable feedback

stimulus. As can be obtained from the results, the perception of thermal is highly subjective and prone to personal preferences. For instance, one participant rated all temperature levels as unpleasant and another did vice versa. This highlights that thermal feedback is a personalized feedback channel that does not follow "one size fits all" configuration, which is in line with previous work. Further, prior work shows that thermal cues are conveying emotions [281] and are rated in dependency to context factors [85]. The findings validate the influence of the variables of thermal feedback on both the comfort and interference.
When inquiring the participants about their reasons for choosing the parameters as they did, it became obvious that there needs to be ensured a balance of feedback intensity which is not distracting but can still be noticed in a stressful situations. Accordingly, for choosing the *rate of change*, three participants preferred a low rate to avoid being distracted by a faster change. In contrast, five participants worried that having a less intense *temperature level* or a low *rate of change*, they would not notice the thermal feedback under stress and consequently be irritated because they expected it then.

Interestingly, the qualitative results are not necessarily reflecting how people rated the different stimuli in the questionnaires. Most striking is the preference on the body location. Although the mean ratings are very similar for each of them and did not show any significant difference, in the interviews, participants stated that the wrist or upper arm would be a suitable location despite the upper arm was rated to have the most discomfort ($M = 0.48$) and the highest interference with the ability to work ($M = 0.62$). In this context, the named advantage of preserving privacy as explained by eight participants, played an important role. While on the one hand users appreciated the unobtrusiveness of thermal feedback which can be also found in related work [157], on the other hand they would appreciate to have such a thermal stimulus incorporated in existing technology, such as wrist-worn devices like smartwatches. Speaking of the integration into prevalent interactive systems the specific purpose of signaling stress through thermal feedback has to be considered. Given that another two participants mentioned the advantage that other feedback channels, namely audio or vibration are already occupied by other devices and therefore thermal feedback has a unique value making it easy to identify its source and meaning, leads to the question whether integrating thermal feedback in existing wrist-worn devices would not take away this advantage due to its close proximity to already used feedback.

Despite the positive comments on the quality of unobtrusiveness of thermal feedback and one participant saying that it felt like more "*direct*" coming from his own body and not being induced externally, the efficiency of such feedback was questioned by three participants. This concern emphasizes the problem that possibly a notifier about stress could intensify the feeling of being stressed, what is also strengthened by the subjective data on the perceived stress obtained from the first user study. Accordingly, there can be seen in 5.1, that there is almost no difference in the participants' perception of their own stress level regardless if a tactile stimulation is provided or not. Additionally, the high subjectivity in what how the feedback stimulus should be designed has been shown to require individual knowledge about a person's preferences, particularly when it comes to the preferred location. For this, the exploration of the concept of notifying about stress has been shown to be an important step revealing the challenges when wanting to design such feedback. In the following chapter another approach will be investigated which neglects the subjectivity of tactile perceptions and rather relies on a more generic concept, namely to manipulate the potential source of stress or so-called stressor.

Limitations Both research studies targeted to examine the suitability of tactile stimulation, and particularly thermal feedback for notifying the user about one's stress state. Although this concept seems promising since for example biofeedback has been shown to improve effectively stress coping [258], the it is still not fully discovered in which demanding situations exactly such feedback would be appreciated taking into account the complexity of stress. While in the qualitative data obtained in the later study it has been found that participants had ambiguous opinions about the utility of such a system, a detailed investigation evaluating its effects in the wild is still missing. Therefore, the concept which has been focused on in this chapter provides a theoretical foundation for the implementation of such a stress notifier considering the stimulus design for its potential integration in an interactive system.

5.5 Chapter Summary and Conclusion

In this chapter the exploration of tactile feedback as a suitable mean to notify users about their stress state have been presented. For this, two consecutive user studies have been carried out. While in the first I focused on the investigation of the effects of familiar vibrotactile and unfamiliar pressure-based tactile stimulation, the second work arose from the findings of the first study that indicated that both tactile feedback pattern affect the users' stress level similarly, which has become obvious in the subjective data (cf., Table 5.1) However, a slightly higher distraction for vibration could be observed in the physiological data (cf., Graph 5.4). Thus, a close similarity of the feedback stimulus to existing solutions being incorporated in smartphones or wrist-worn wearables, seems to be only partially suitable. In fact, combining the findings from prior research with the results from the first user study presented in this chapter and the qualitative statements from the interviewees of the second study, the answer to **RQ2a** is that neither visual nor auditory feedback suits for notifying about stress states. As being stated by the participants, these two are already occupied by recent technology competing for the user's attention. Moreover, throughout the performed research it has been shown that thermal feedback suits better for the named purpose since it preserves privacy and it is not too similar to already used tactile stimuli which is why I focused on unfamiliar stimuli, consequently thermal feedback in the subsequent experiment. Being in line with related work that thermal feedback is a promising alternative, with the second user study I aimed to explore the characteristic a thermal stimulus should have in order to be preferred by users for notifying about stress. For this, five temperature levels, three rates of change, and five body locations have been compared against each leading to the determination of the most suitable thermal stimulus. The results indicate that mild cold stimuli, in particular -1°C with a change rate of ±0.5°C/s are perceived the most pleasant. As the preferred body part, the lower back seemed to be the least discomfortable ($M = .029$) with a neglectable potential for interference with the ability to work. These findings respond to **RQ2b** implying that in order to design an effective stimulus, mild cold stimuli with a high change rate should be considered. Particularly when picking the body location where it is going to be presented, the findings suggest that the design perspective should take into account the subjective value of privacy preservation. While some users would prefer to have such a thermal stress notification system incorporated in familiar technology, for example a smartwatch being worn at the wrist, others valued the privacy of receiving thermal feedback on their stress level unobtrusively and without bystanders being informed about it.

To conclude, as the qualitative data gathered in the latter study reveal, feedback about one's stress level is seen very ambiguous by users. While 18 interviewees considered this approach to be interesting and worth to be continued, there are also some critical aspects mentioned. In practice, such feedback could also lead to contrary effects, namely the stress level would increase. A hint that the notification about one's stress level does not affect, at least the subjectively perceived stress level was shown in the first study presented. Consequently, the following Chapter 6 explores a different approach how to apply stress mitigating techniques in the context of interactive systems.

Chapter 6

Manipulating Stressors in Interactive Systems

With the aim to incorporate stress mitigating techniques in interactive systems, in the last chapter the exploration of suitable and effective feedback for notifying users about their stress state has been in the focus of attention. As has been suggested by the findings from the two performed user studies, the utility of such a concept for real-world scenarios is questionable since users reflect upon potential ascents in stress when they are made aware of their current stress level.

Considering awareness, there are different ways of how self-awareness could be used in the context of stress reduction methods. While the last described approach relied on making users aware of their stress state as it instantly occurs or respectively develops into unpleasant state affecting their wellbeing, another concept introduced in the present chapter is based on supporting the awareness of how users cope with potential sources of stress, so-called stressors. In this context, I refer to stressors as something that elicits stress or a feeling of discomfort in the user. Consequently, the first study concentrates on smartphone notifications as a potential stressor, since it has been shown to have multiple negative effects, namely causing interruptions [200] by demanding attention [179], evoking response pressure [228, 200], and increasing the risk of developing a Fear of Missing Out (FOMO) [61, 190]. For the two subsequent studies cognitive load and the perceived feeling of being busy have been referred to as stressors. Mental stress, for example elicited by performance or cognitive demands [23, 175, 252], going hand in hand with the feeling of being busy have been shown to be closely connected to feeling stressed.

Considering subjectively identified stressors as a challenge for researchers to be manipulated, stress management interventions have been explored in prior research [30, 194, 283]. For this, different approaches, such as stress relieving mobile games, haptic stimulation or biofeedback have been tested. The main body of research has been performed within the field of "affective computing" creating considerable examples of stress-recognizing systems which are able to provide in-situ feedback on the user's state [175, 258, 290].

Although promoting self-awareness and designing stress-aware technology have been researched extensively, the role of stressors in this context has been largely neglected so far. Therefore, in this chapter, I focus on the investigation of stressors evoking and resulting into a feeling of stress. Through three studies, I investigate by which means stressors can be manipulated and how such manipulation affects the users answering **RQ3a**, **RQ3b**, and **RQ3c** (see Table 1.1). Thus, three approaches were realized to (a) eliminate the stressor, (b) make the stressor visible on a private and semi-public basis, and (c) visualize the stressor publicly and making it adjustable for the user. By embracing all three kinds of the privacy continuum, namely privately visible, semi-publicly visible, and fully publicly visible, the aspect of privacy preservation has been systematically manipulated. Hence, the last **RQ3d** has been targeted to answer, approaching the challenge of how to include privacy when designing stress-aware interfaces.

This chapter is based on the following publication:

- R. Poguntke, et al. NotiModes– An Investigation of the Effects on Smartphone Users when Manipulating Notification Delivery. (*to be published*)
- R. Poguntke, et al. Investigating Attributes for an In-situ Ambient Visualization of Cognitive Load Using Physiological Data. (*to be published*)
- J. Häkkilä, R. Poguntke, E. Harjuniemi, L. Hakala, A. Colley and A. Schmidt. BuSiNec – Studying the Effects of a Busyness Signifying Necklace in the Wild. In *Proceedings of the 2020 ACM Conference on Designing Interactive Systems*, DIS'20, New York, NY, USA, 2020. ACM

6.1 Related Work

In the following, related work characterizing major aspects of stressor research including coping strategies is summarized. Further, the role of interactivity opportunities against the background of user control is addressed as a significant aspect of the contribution presented in this chapter.

Stressor Manipulation and Coping As already pointed out in Section 2.2 giving some background knowledge about the distinction of stressors, as processive stressors there can be regarded demanding situations in general. Although the research corpus of studies that examined stress eliciting conditions and circumstances [23, 151, 242, 243] is huge, there has been conducted little work on the systematic manipulation of stressors, particularly in relation to digital technologies. However, in this context the individual's appraisal of a potentially stressful situation can be understood as a possibility to manipulate one's perception of the stressor- even though the stressor itself cannot be changed. Lazarus and Folkmann [152] introduced the *Transactional Model of Stress and Coping* to explain that the respective coping strategy depends on the individual's appraisal. This cognitive process Baum et al. [13] classify as a "transmission variable" being prone to an individual's personality. Consequently, also the coping strategy varries according to distinct factors. For example whether one has an optimistic or pessimist view [232] can influence how stress is being dealt with. Moreover, Ptacek et al. [214] could show that although sex differences do not allow predictions in what way individuals will cope with stress, women's and men's different sozialization is shaping the choice of their coping strategy considerably.

User Control in Interactive Systems The opportunities to interact with computers have been researched extensively by previous work [59, 121, 287] with the raising complexity of interactive systems. In this context, McMillan and Hwang [177] propose the three dimensional "*Measures of Perceived Interactivity*" to assess the interaction potential of media. While this tool has been primarily designed to reveal differences between web-based communication in contrast to traditional media, it shows that there exists a range of how interactive users can apply technology. In the domain of HCI the understanding of interactivity as a valuable factor has gained importance in connection to so-called "*context-aware*" computing. By this term, "*an applications's ability to detect and react to environment variables*" (p.2) [9] is being referred to. Given the fast technological advancements in the development of such applications, Barkhuus and Dey [10] examined user preferences regarding the granted degree of user control when interacting with them. Based on the categorization of three "*levels of interactivity*" how they call it, their study revealed that taking away the user's autonomy is not favored. To distinguish the three levels, they define the "*personalized*" application which allows users to customize and adjust settings according to his or her needs, the "*passive context-aware*" application which makes decisions transparent and leaves potential adjustments in the settings to the user and just proposes changes, and the "*active context-aware*" application which neglects the user and tailors the application exclusively based on the collected data [38].

Based on this distinction introduced in prior work, the present chapter also applies three different levels of user control being also noticeable in the granted interactivity opportunities. While the two performed studies on the visualization of stressors are clearly separated into putting the user into a passive role (cf., Section 6.3) and on the other hand requiring his or her action (cf., Section 6.4), the approach to eliminate stressors varies in the granted user control according to the specificly active mode. Consequently, the results of the following studies strengthen the findings from Barkhuus and Dey [10] by showing that users dislike autonomously acting systems and rather tend to accept applications that make suggestions but leave sufficient decision space to the user. These gained insights lead to the presented "*Considerations for a Stress-Aware Notification Management*" (cf., Section 6.2.6).

6.2 Eliminating Stressors

As the manipulation of stressors is being investigated in the present chapter, the first user study explores the elimination of potential stressors, in this case smartphone notifications as an opportunity to reduce stress in users. Hereby, a use case affecting the most private level has been chosen to be able to provide different nuances of user control when manipulating notification delivery. For this, the times when a notification is being delivered has been adjusted according to three different modes (cf., Section 6.2.3) realized in an Android mobile application. These modes have been developed based on prior work and the findings from a focus group targeting to reveal subjective user experiences for each mode. Since the three modes further varied in their degree of user-control another aim was to deduce which degree of user control is favored by users for managing notification delivery.

6.2.1 Evaluating Notification Delivery Modes in a Focus Group

Prior to the implementation of a smartphone application, a focus group was conducted to explore (a) how users are dealing with interruptions and (b) how their notification management could be improved avoiding interruptions. From the findings and in conjunction with prior work, three different notification delivery manipulation modes were designed.

Participants, Procedure, and Data Analysis In the focus group participated seven smartphone users (M = 21.86, SD = 2.9 years; 6 males, 1 female). All but one were students and reported to have installed several applications on their phones from the categories messenger (WhatsApp), social media (Facebook, Instagram), e-mail (Google Mail) and some unspecific applications, such as calendars or apps to manage their finances. Before demographic data was collected, the purpose of the focus group was explained and the participants were asked to sign the consent form. Given their consent, further the discussion was audio recorded. Afterwards feedback on the perceived importance and the annoyance of notifications was gathered, and respectively it was inquired in which use cases notifications were appreciated. In this context it was further asked, whether the participants experienced any interruptions or other aversive consequence when receiving notifications throughout their day. Before ideas on how to prevent such interruptions were collected, users were inquired how they predominantly dealt with annoying notifications. Then current solutions being available in smartphones nowadays were discussed and how these could be refined or improved. Finally, the participants were thanked for their time and participation. The entire session lasted approximately 70 minutes. For analyzing the qualitative data emerging from the focus group discussion, the audio recording was reviewed in addition to the notes taken during the session. After collecting similarities and differences in the participants' statements, suggestions describing commonly agreed approaches were formed.

6.2.2 Focus Group Results

Throughout the discussions within the focus group, participants agreed on the negative consequences affecting their productivity when receiving multiple notifications. Regarding the perceived interruptions three out of seven admitted to be too curious and not being able to put their phone aside saying that: "*One is distracted immediately and takes a look* [on the phone]". Another participant referred to this problem stating: "*I believe that we are a bit addicted to receiving notifications and inevitably take a look at them*". In practice, situations, when notifications have been perceived as annoying, playing mobile games but also periods when users wanted to work or study had been mentioned. Besides the annoying character of operating system notifications, all participants argued that notifications were also useful to present important information briefly at one glance to the user. They named the advantage of choosing between auditory, visual, and haptic feedback when receiving notifications. Interestingly, participants also emphasized that they had disabled showing notifications, particularly text-based messaging content on the lock screen due to privacy concerns mentioning shoulder surfing incidences.

Suggested Concepts Preventing Interruptions Reflecting upon approaches and solutions how to prevent interruptions by unwanted notification delivery, the participants named several promising ideas which have been partly already realized in premature form by, e.g. Apple. Among several interesting suggestions as, for example a "*time box*" where the user puts in the smartphone and is able to take it back after a pre-defined amount of time has elapsed, those concepts are presented subsequently, that have been either realized already in prior work, or respectively have inspired the design of the mobile application.

- **Occupation Mode** – Next to the profile picture there is displayed the current status of a person comparable to the "not available" status in Skype.
- **Context Analysis** – Based on context information, e.g. sent mails, visited websites, etc. artificial intelligence should be able to filter notifications depending on their priority. Further, location data, e.g. GPS could be used to allow notifications on the user's way to work but not while being in a meeting. A similar approach has been introduced with Apple's iOS 11 operating system called the "Do Not Disturb While Driving" feature [10]. It automatically blocks notifications when users are driving. Further this feature can be customized to send not-available messages and it can be even set to allow senders to send urgent messages right away.
- **Priority Filtering** – Text messaging including importance signalling words, such as "now", "immediately", or "emergency" should be allowed to be sent any time. Also specific contacts and even applications could be added to a filter allowing the user block notifications coming from these. Auda et al. [4] implemented a rule-based notification deferral system for

[10] https://support.apple.com/en-us/HT208090

which "*trigger*" or "*exempt*" words can be set that lead to either letting pass or blocking the specific notification.

Other existing modes, such as the "Do not disturb" or the flight mode were actively used by one person each. The participants stated that having all connections, also calls and SMS disabled was not desirable for practical usage. In comparison to such "static" modes as named by the participants, the users preferred "dynamic" modes relying on artificial intelligence to adapt to the user's habits and contexts. Hereby, a possible drawback named by participants could be that the system requires too much time to learn which might lead to uninstalling the app after a short usage period. In general, participants appreciated solutions that respect their privacy and were easy and usable in handling exceptions.

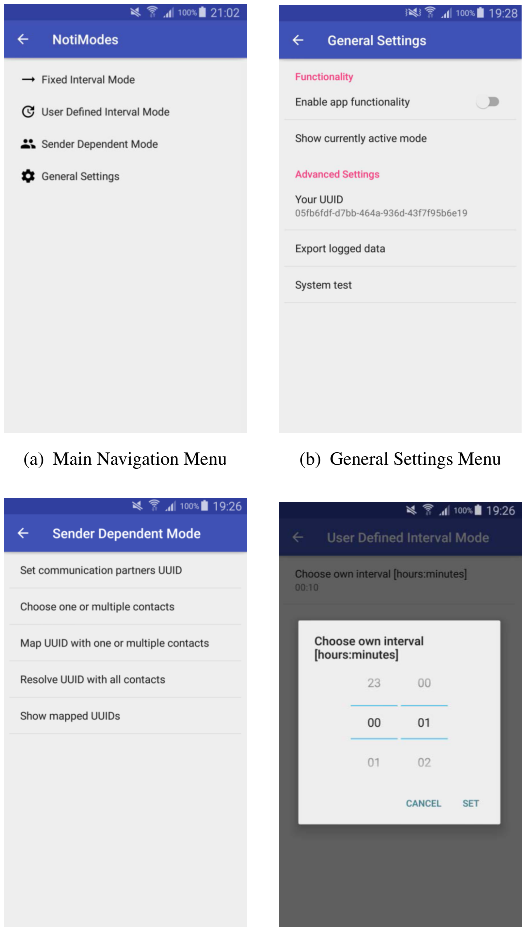
Conclusion All in all, the focus groups participants agreed on the annoyance factor and the negative consequences notifications can have. They further admitted that interruptions caused by notifications affected their task performance, which strengthens findings from prior work [6]. Although various present solutions for dealing with notifications were discussed, the users considered context-aware systems based on artificial intelligence, priority filtering, and "timeouts" for blocking notifications to be favourable. Given the ongoing research attempts to improve priority filtering [4], the focus of this user study was to manipulate notifications in terms of delayed delivery. Additionally the concept of sender responsibility for sending messages was explored. Consequently three different modes have been implemented in an Android application called "NotiModes" being evaluated in a four-week long user study in the wild.

6.2.3 Implementation of NotiModes

The application *NotiModes* was developed to work on smartphones having the Android version 5.0.1 or newer. Besides the application, a Node.js server was used to realize the notification delivery modes. In the following, the *NotiModes* application and the developed notification delivery modes are described in detail.

Android Application The application *NotiModes* uses Android's Notification Listener Service Application Programming Interface (API) [11] to access notifications. As users open *NotiModes* for the first time, they are informed about the data collection and have to approve it to be able to continue using the

[11] `developer.android.com/reference/android/service/notification/NotificationListenerService.html`



(a) Main Navigation Menu (b) General Settings Menu

(c) Sender-dependent Mode (d) User-defined Interval Mode

Figure 6.1: Views of the NotiModes Android application including the main menu, the settings overview, and the UDI and SD modes, for which users needed to configure the settings according to their needs.

application. After explicitly confirming the data collection and granting the app permission to access notifications, the main view of the application is shown (see Figure 6.1(a)). The app's main view contains a menu for the setting of the app and each notification delivery mode. As the app functionality in the "General Settings" view (see Figure 6.1(b)) is enabled, *NotiModes* activates the smartphone's silent mode. Consequently, the app blocks notifications and respectively their visual, auditory, and tactile signaling, namely LED blinking, sounds, and vibration. Only incoming calls are allowed to pass through with a short vibration. When a notification is received, the application stores it transiently locally on the smartphone and removes it from the notification bar, so that notifications cannot be seen on the lock screen too. While accessing notifications, *NotiModes* ignores "ongoing" notifications (*i.e.*, timer, downloading).

NotiModes- Delivery Modes Three different notification delivery modes (see Figure 6.1(a)) have been designed based on inspirations from the focus group results and concepts that have been researched by prior work. Thus, the participants' idea of implementing a "timeout" mode referring to providing distinct intervals which are pre-defined but can be individually set by the user has been taken up in the mode implementation. Hence, the Fixed Interval Mode (*FI Mode*) blocks notification delivery for one hour, while the User-defined Mode (*UDI Mode*) allows users to set the "not disturb" interval individually. This difference in the interval settings was supplemented by the Sender-dependent Mode (*SD Mode*) that gives the sender the possibility to know about the deferral of the message and to decide upon the urgency of the mobile message. Consequently, the three implemented modes represent distinct levels of user control and respectively user responsibility. To realize the latter mode, a server was used to support the communication between the sending and the receiving partners working on the basis of exchanged unique IDs between sender and receiver. In the following, each delivery mode is described in detail.

Fixed Interval Mode In the Fixed Interval (FI) Mode, notifications are delivered once in an hour. After the mode is activated, *NotiModes* intercepts all incoming notifications and displays them to the user at the end of a one-hour interval. During this interval, the app temporarily stores incoming all but "ongoing" notifications and removes them from smartphone's notification bar. Since *NotiModes* puts the smartphone on the silent mode, the user does not get any audio or haptic feedback. Furthermore, as a notification is received, the app creates an override notification to hide the heads up part of the posted notification and instantly removes it. In addition, *NotiModes* sets the notification light on black color. These steps ensure that the user does not receive notifications during the one-hour interval. When the interval is over, the app notifies the user about all received notifications during the last hour and removes all stored notifications. After the hourly notification delivery, a new one-hour interval starts. As soon as the user reads and explicitly removes all delivered notifications from the notification bar, *NotiModes* starts intercepting new incoming notifications again.

User Defined Interval Mode In the User-Defined Interval (UDI) Mode, the notification delivery behaves similarly to the *FI Mode* with the only difference that the user can set the length of the delay interval according to individual needs. Using the settings menu (see Figure 6.1(d)), the user can determine the interval from one minute up to 23 hours and 59 minutes. As a default the interval was set to 10 minutes but could be changed according to their individual needs anytime.

Sender Dependent Mode The Sender-Dependent (SD) Mode is based on the *FI Mode* regarding it's delaying functionality. The important difference hereby is, that the sender of a message is informed about the postponing of sent messages and can decide if the message should be delivered directly or at the end of the interval. This mode has been inspired by the work of Schmidt et al. [235] transferring the responsibility of potentially annoying messages to the sender of a message instead of the receiver. Compared to the other two modes, to be able to use *SD Mode*, *NotiModes* must be installed on the smartphones of both, the sender and receiver. Furthermore, both communication partners must have a connection to the server that is used to provide communication between them. Upon installation of *NotiModes*, a Universally Unique Identifier (UUID) is created. To enable *SD Mode*, it is important that the mapping is set up on both devices of sender and receiver, by linking the UUIDs and contact names. The mapping is done by entering the UUID of the communication partner, choosing the communication partner from the contact list of the smartphone and finally map the UUID to the contact (see Figure 6.1(c)). In case, the sender wants to send a message *NotiModes*shows a dialogue realized as a pop-up window asking about the confirmation of delivery. The sender can either force the immediate delivery of the notification or approve the delay. In the case of approving, the notification is going to be delivered after the one-hour interval is over. The UUIDs have been used to avoid storing the names of the participants on the server and consequently preserving their privacy.

Data Collection The *NotiModes* application collects information about the device (*i.e.*, device type, Android version), notification meta-data (*i.e.*, package names, timestamps for received notifications), and interaction with the application (*i.e.*, active settings). The app stores all collected information locally on the smartphones using a SQLite database server. The users can export and view all stored data using the "Export logged data" button which can be found in the

General Settings (see Figure 6.1(b)). *NotiModes* stores notifications content (*i.g.*, the title and text of notifications) temporarily and only during the deferral interval. After the delivery, the notification content is deleted from the storage. The app locally stores only the contact information of the communication partners for *SD* mode. The server was implemented for the *SD* mode, and all requests use a secure connection.

6.2.4 User Study

The aim of the user study was to evaluate the user experience and acceptance with respect to the presented notification delivery modes eliminating potential stressors. Further, subjectively perceived effects and particularly the perception of the varying user control has been assessed bearing implications for managing notification delivery.

Study Design In a longitudinal evaluation thirteen participants used all three modes manipulating notification delivery following a within-subject study design. The three modes plus a control condition, where the notification delivery was not manipulated, were presented in randomized order according to Latin square over all participants (see Figure 6.2). Each mode was used for one week and switched automatically to the other. Consequently, the total study duration was four weeks.

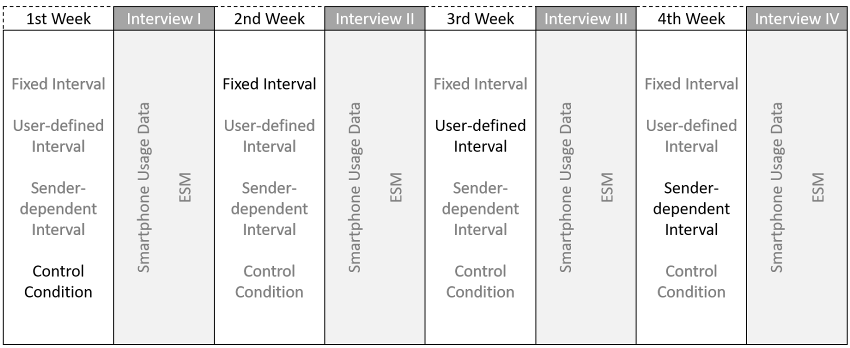


Figure 6.2: During the four week study, the participants were assigned to three modes plus an additional control condition randomly for one week each. As quantitative variables phone usage data and ESM data, as well as qualitative data via in-depth interviews had been collected throughout the whole study period.

Through semi-structured interviews after each week of mode usage including one additional control week where no notification delivery was manipulated, the users' subjective opinions and preferences were collected. As a requirement to be able to realize the *SD Mode*only pairs of participants that knew each other beforehand were recruited.

Measures In this user study, a two-fold evaluation process was applied. On the one hand, quantitative data was assessed through subjective ratings evaluating the user satisfaction during the usage of the three different notification delivery modes, as well as the logging of smartphone usage, in particular the amount of incoming notifications. On the other hand, an extensive qualitative evaluation has been performed by carrying out interviews with all participants after every week of usage, at the point of time when the delivery mode was switched. These interviews aimed to get an in-depth view on individual experiences from users when interacting with the three different notification delivery modes.

Quantitative Evaluation The three notification delivery modes served as independent variables complemented by a control condition. To investigate how the notification delay affects the user, the participants' stress level, activeness level, and happiness level have been assessed via the ESM [47, 109]. Following prior work [178] the users have been prompted between 8 am and 11 pm four times per each three-hour time window (8am-11am, 11am-2pm, 2pm-5pm, 5pm-8pm, 8pm-11pm). A reminder was sent when the ESM had not been answered for 30 minutes; there had been provided at least one ESM right after the user received the delayed notifications to ensure that assessments have not only been taken during the calming period. Each of these three dependent variables was presented as depicted in Figure 6.3. Participants had to move a slider to rate their stress, activeness, and happiness level on a 5-point scale ranging from 0 (=very stressed, sleepy, sad) to 4 (=very relaxed, active, happy).

Qualitative Evaluation It was performed one interview after the usage of each mode as well as the control mode, leading to in total four interviews per participants. The interviews lasted 9.32 minutes per session in average and were conducted either face-to-face or via Skype. All interviews were audio recorded given the participants consent and transcribed individually. In every one of the four semi-structured in-depth interview the participants were asked six questions concerning the (a) advantages of one specific mode, (b) the disadvantages of the mode, (c) the willingness to use the mode privately, (d) effects of the specific modes on the participants' daily routine, (e) the overall concept of the modes, and (f) what improvements they wished for to be willing to use such a mode in the future. For the *UDI Mode* and *SD Mode* additional questions were phrased,

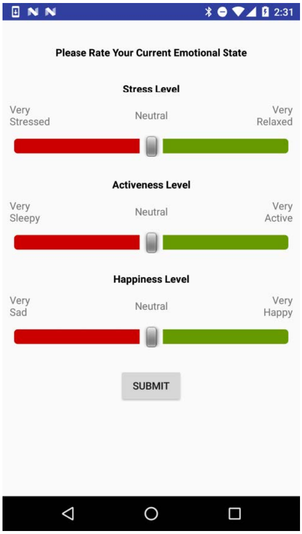


Figure 6.3: Assessing stress, activeness, and happiness ranging from 0 signifying a low level to 4 indicating a high level. The ESM prompting participants at random times per day had been used.

addressing the individual settings the users chose as well as their experiences. In the final interview, two questions were posed aiming to assess the overall concepts by comparing and discussing the different modes.

Participants and Procedure 35 participants were recruited in total of whom 18 dropped out during the entire study process. Another four participants had to be excluded due to technical issues with their phone operating system, leaving a sample size of $N = 13$. While interviews were conducted with all participants, six users among them three females ($M = 23.0$, $SD = 3.16$ years) could not send their logging files comprising the quantitative data which was the final step for each participant. For this, it will be reported on the sub-sample of seven participants ($M = 25.57$, $SD = 11.82$ years) among them four females of whom quantitative data has been received. Each participant was compensated with 40 Euros. As a requirement for participating in this user study, the participants had to be smartphone users and had to sign up for the study as a pair of two, since a chat partner was needed to test the *SD Mode*. To ensure that the participants were interacting with their smartphones frequently, they were asked to rate their usage behavior on a scale from 0 (=never) to 9 (=all the time). With an average rating of 7.15 ($SD = 1.21$) it can be referred to them as regular smartphone users.

Participant ID	UDI Changes	Min. Interval Length (Min)	Max. Interval Length (Min)
P1a	4	0	10
P1b	0	0	0
P2a	8	1	421
P2b	2	120	180
P3a	1	30	30
P3b	0	0	0
P3c	3	60	240

Table 6.1: Usage data for the *UDI Mode* per participant including the amount of changes within the *UDI Mode* as well as the minimum and maximum interval length given in minutes.

Before the user study started, all participants received an email including the study consent form and a detailed guide on how to install and use the *NotiModes* application on their device. By sending these information before the study start, it should be ensured that users were aware of the notification delivery manipulating functionality of the *NotiModes*app. After receiving the signed consent form, an email with an online link to the *NotiModes* app was sent and the participants were asked to fill in a short demographic questionnaire. Furthermore, they should rate their smartphone usage. Before using the *SD Mode*, each participant had to enter the assigned UUID for his/her study partner. After having successfully installed and started the application, the specific mode was activated and ran for one week before switching to the next notification delivery mode.

6.2.5 User Study Results

Subsequently the results of the two-fold evaluation process will be presented. Due to technical difficulties, quantitative logging data and ESM measures of six participants had to be removed from the analyzed data set. Therefore, the following subsection will describe the results of the notification logging and ESM results for seven out of 13 participants. From the logging data usage information on how users customized the interval length in the *UDI Mode* (cf., Table 6.1) could be inferred.

	Mode	Mean	Standard deviation
Happiness	Control	2.41	0.53
	FI	2.52	0.59
	UDI	2.43	0.51
	SD	2.28	0.61
Activeness	Control	1.88	0.64
	FI	1.87	0.76
	UDI	1.96	0.59
	SD	1.83	0.68
Stress	Control	2.34	0.70
	FI	2.35	0.59
	UDI	2.45	0.69
	SD	2.06	0.85

Table 6.2: Means and standard deviations of the assessed happiness, activeness, and stress levels for the each of the four conditions: control, *FI Mode*, *UDI Mode*, *SD Mode*.

Experience Sampling Method (ESM) Ratings Collecting data on the happiness, activeness, and stress level, users were prompted several random times per day. The *happiness level* has been rated highest at average during the *FI Mode*, while it has been very similar for the control and the *UDI Mode* condition. During the *SD Mode*, participants reported the lowest happiness level. For the activeness level it was observed that participants felt most active during the *UDI Mode*, whereas the other conditions showed only minor differences (see Table 6.2). During the *SD Mode* the stress level has been rated lowest, whereas during the *UDI Mode* participants gave the highest ratings at average. The exact values can be obtained from Table 6.2.

Interview Results From all participants qualitative feedback was gathered in a semi-structured interview after each week of mode usage resulting in 52 interviews (four interviews per participant). For the data analysis, an open-coding process was applied. Hereby, first all data was transcribed based on the interview recordings and imported to ATLAS.ti[12]. Two researchers then derived a set of 16 codes covering, besides others, the advantages and disadvantages of each mode, individual experiences, and suggestions for improvements. In a second phase, their individual code assignments were compared and all codes that did not match

[12] ATLAS.ti - `https://atlasti.com/`, last accessed 2019/04/05

among the two coders were discussed extensively until a consensus was reached. Through a final discussion the most important key findings emerging from the qualitative data were deduced.

Favorable User Control and User Experiences The overall preferences for each delivery manipulation mode had been assessed in both samples as part of the interviews ($N = 13$). On a scale from 1 (="I don't like it at all") to 5 (="I liked it very much") the participants were asked to rate the particularly used modes. While the *UDI Mode* was rated highest ($M = 3.15$, $SD = 1.14$) followed by the *SD Mode* ($M = 2.81$, $SD = 1.07$), the least preferred mode was the *FI Mode* ($M = 2.54$, 0.78). This ranking is supported by stated comparisons throughout the interviews. Accordingly, P11a, P4a, and P12a reasoned that they ranked the *UDI Mode* higher because of the possibility to customize the time interval. As depicted in Table 6.1, almost all participants made use of this feature and chose to adjust different notification delay intervals according to their individual needs. Concerning their preferred degree of control, users stated to prefer a combination of different modes. Having a certain interval of delay, but also defining exceptions based on certain people for communication apps (P6c, P3c) would be a favorable and useful improvement granting the user the sufficient amount of control to avoid distraction while allowing to be contactable for emergencies. Correspondingly, participants appreciated the reduced frequency of notifications (P4a) when the *UDI Mode* had been activated. They emphasized the positive impact of receiving notifications in a bundle after a certain amount of time elapsed (P4a, P6c) and found it very handy to define the time span themselves (P3a). Participants noted that they felt less distracted (P12a, P12b) and interrupted (P11a) using the *UDI Mode*, stating that they enjoyed in particular the delay in communication because "[...] there would be no need to reply right now because messages are not coming in real-time anyway" (P12a).

For the *FI Mode*, participants stated to have enjoyed some quiet time without any notification alerts (P6c, P12a), especially for specific occasions like having to study (P4b). The FI enables the user to quickly assess the notifications (P2b) from a certain time interval like one hour with a low effort. Moreover, participants felt less distracted in the FI (P12b), in particular in regards to less important notifications like those sent from the operating system or social media (P2a, P11a). Participant 12a claimed that the *FI Mode* also decreases the checking habit throughout the day.
Referring to the *SD Mode* P1a mentioned getting used to not receiving notifications in time and P12a furthermore adds, similar to the *UDI Mode*, that texting is becoming more relaxed when not expecting to receive an answer immediately.

When using the *SD Mode*, P4a highlighted the usefulness of specifying different notification patterns for messaging services and non-messaging related applications. P11a describes that the experience was positive and he/she "[...] liked [the SD mode] very much because [he/she] will get the messages right away, but the other notifications were delayed so [he/she] didn't get them all the time". P11a furthermore emphasized the handiness of the chat partners having control over the delay. Besides, having the guarantee to receive urgent messages in time, which was not possible in fixed or *UDI Mode*, was mentioned as an advantage (P1a, P3a, P3b, P3c). Lastly, P2b stated the usefulness of the *SD Mode* to reduce the number of notifications especially for chat partners who tend to write many messages which could be summed up into just one.

Despite the overall positive feedback on the used modes, participants also had some critical remarks. Since sending messages right away is the common default as pointed out by a minority of users (P2a, P12b, P12a), interviewee P4a would rather like to reverse the *SD Mode*, meaning that the user is asked whether he/she wants to delay a message. Participant 3c furthermore mentioned the additional workload *SD Mode* generates, when communicating with many people on various different channels. While appreciating the delay of notifications for non-instant services, e.g. e-mails (P4a), the participants saw disadvantages in using this mode for instant messaging notifications (P4a, P12b, P2a). They were missing the opportunity to end this mode before the previously defined time interval elapsed (P3c, P3b, P2b). P1a and P1b mentioned to be worried about not receiving urgent messages and P12a expressed concerns about the high number of notifications at the end of one interval, feeling "[...] it was a bit too much when it all came at once" (P12a). During the interviews, participants expressed various concerns about the *FI Mode*. They claimed to want to see messages from friends and family instantly (P11a, P2a, P3b) and rated the fixed time interval length as either too long to wait for answers from chat partners (P1b, P3b) and on the other hand too short to have some real quiet time (P3c). Another disadvantage was to never know whether new messages came in (P1b) and not to be able to define the time interval individually (P2b, P3c, P12a). The participants mentioned being afraid to miss something while using the *FI Mode* (P4a, P4b, P3b), especially when something important or urgent comes up (P3b, P12b). P12b and P1a stated to have missed many messages, for example while waiting for a ride home or going out to lunch with friends, with P1a reporting that "[...] this is totally depressing. I missed a lot". P2b and P3b found this mode to impede having fluent text-based messaging conversations.

With regard to the effects on the users' everyday life, less than half of the sample (P1a, P1b, P2b, P3b, P4a, P11a) reported none or only a minor influence of the

UDI Mode on their all day life. Similarly, six participants did report no or only minor influences of the *SD Mode* on their usual smartphone usage behavior (P1b, P2a, P2b, P3b, P4a, P12b). Seven out of all thirteen participants did not notice any changes in their usage behavior using the *FI Mode* mode (P1b, P2a, P3a, P4b, P6c, P11a, P12b). Two participants stated that they even checked their phone more often during this mode, especially when expecting messages (P2a, P4a). And for P4a not being able to receive messages at any time even increased the subjectively perceived stress level. Those who reported an effect, stated positive remarks for their daily routine exclusively. For example, participant P3a stated to check the phone less during the *UDI Mode* when compared to the *SD Mode*. Further the *UDI Mode* provided support to leave the phone on the side for some time (P12a) and made it easier for the users to concentrate (P3c). Moreover, P2b enjoyed not getting notifications for a certain time interval, whereas P2a positively highlighted not receiving any notifications during the night. Accordingly, two other participants (P11a, P12a) mentioned checking their phone less while having used the *FI Mode*.

6.2.6 Considerations for a Stress-Aware Notification Management

From the presented results of the users study on the elimination of stressors in form of notifications, the following four considerations have been inferred. These can be taken as suggestions how to support a notification management that fits stress-aware technology.

Allowing the Customization of Delay Intervals As reflected in the results, users preferred to set their intervals autonomously fitting their needs. Particularly, having pre-defined modes for, e.g. studying (2h), sleep (7h), or meetings (1h) was also suggested and seemed appealing since it would facilitate the application's usage.

Allowing Different Delays for Different Apps In the interviews participants said that they appreciated that system notifications had been delayed. Accordingly, they suggested to provide different pre-defined delaying intervals for specific applications clustered into groups, such as "entertainment" or "news".

Providing Filters for Classifying Notifications In connection with reviewing the sender dependent mode, some participants would have liked to have an algorithm which recognizes words and classifies them as emergency words, such as "hospital" or "doctor". Another promising approach which is in line with the suggestions of our participants to filter notifications, was introduced by Westermann, Möller, and Wechsung [272]. They found that users are likely to accept notifications if they are provided with requests including brief explanations what the application does.

Allowing Exceptions Evaluating the currently available solutions, such as the Do-not-disturb mode, the users stated that allowing exceptional solutions would have provided a huge benefit to them. While the Do-not-disturb mode is blocking every prospective disruption, they wished for a feature to define exceptions, such as contacts or even contexts when messages or calls would be able to answer while using this mode. How such a rule-based notification deferral could be applied in practice was researched by Auda et al. [4].

6.2.7 Study Conclusion

In this presented user study, a thorough investigation of how users experience and perceive different notification delivery modes varying in their degree of user control has been presented. As the gathered user feedback revealed, participants preferred a combination of different modes allowing to define specific silent intervals based on the premise of being able to allow exceptions for emergency cases. Consequently, users tend to control when and by whom they can be reached relying on technology to define exceptions. With the aim to eliminate the source of stress, namely notifications, potential effects on the user have been assessed through quantitative and qualitative data. Considering the results from the ESM ratings, the average scores have shown to be similar among the three manipulation conditions, as well as the control condition. Thus, no significant differences for the users' happiness, activeness, and stress level among these conditions could be observed. However, for the stress rating the greatest difference could be detected while the *SD Mode*had been activated. Interestingly, the standard deviation is also the highest for the stress values in this condition. Accordingly, it can be concluded that some users experiences lower stress under this particular mode, while others did not notice any changes in their mood. The findings from the 13 in-depth interviews strengthen the observation that staying absent from notifications is appreciated by users. Further, the qualitative data revealed that the participation in this user study encouraged participants to reflect upon their notification management and consequently helped them to appreciate the positive effects of silencing their phone.

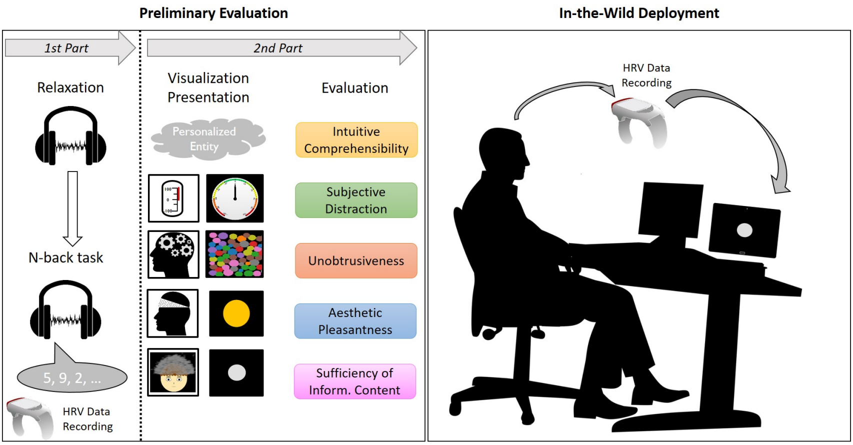


Figure 6.4: Study design divided into the two performed studies: In the preliminary evaluation participants performed the N-back task after a short relaxation period to elicit cognitive load being measured with the Empatica E4 wrist-band. The first part was followed by the randomized presentation of the nine visualizations intermitted by the five variables' assessments. For the office deployment, the first part was repeated measuring cognitive load to obtain baseline values before the ambient visualization was used for one week followed by an in-depth interview.

6.3 Visualizing Stressors with a Passive User

Another technique of manipulating stressors is to draw the user's attention to the potential source of stress. As exemplary use case, the following user study investigates the impact of an ambient visualization signifying the user's cognitive load. For this, a preliminary evaluation of visualization attributes had been carried out first to investigate how cognitive load could be visualized suitably. Based on these results an in-the-wild deployment of the ambient feedback system was exemplarily tested with two users in a semi-public setting. The study design is depicted in Figure 6.4. Following Moere's definition of an ambient visualization, data which reflects time-varying information is represented using the user's peripherical attention [180] (p.1).

Mapping Visualizations to a Design Space Inspired by prior research [64] two dimensions, namely "degree of abstraction" and "amount of information" had been taken to chart a design space covering different kinds of visualizations.

The first dimension refers to the familiarity with existing or known concepts, which is seen in the visualization objects. For example a white circle is much more abstract than a schematized illustration of a human's head. Particularly in interaction design and usability research the familiarity of objects is often applied to facilitate matching user interface components to objects being known from real world experiences. This central concept, which we took up as one dimension of our design space shaping the nine visualizations is part of the ten so-called *10 Usability Heuristics for User Interface Design* introduced by Nielsen [187].
The other dimension describes the amount of information that is provided in the visualization. For instance, if a numerical range and a color scheme are shown in the visualization or if only one parameter, such as the size varies. Again this aspect can be found in the *Usability Heuristics for User Interface Design* [187] referring to the *aesthetic and mimimalist design* by not providing redundant information. Similarly, the criteria of "information capacity" has been identified in prior work [212] to be crucial for the subjective perception of feedback representations, since with an increasing amount of information often the complexity of a visualization rises too.

Thus, these two dimensions bearing huge implications for the design are depicted in a design space consisting of two axis forming a coordinate system what resulted into four quadrants: (a) lots of information and high degree of abstraction, (b) lots of information and low degree of abstraction, (c) limited amount of information and high degree of abstraction, and (d) limited amount of information and low degree of abstraction (see Fig. 6.5). Based on this design space nine different visualization (cf., Section 6.3.2) have been designed and implemented, so that each quadrant is represented. Consecutively, the preliminary evaluation of all visualizations had been conducted with the dedicated purpose to determine the most desirable and suitable visualization for providing feedback upon the user's cognitive load.

6.3.1 Preliminary Evaluation of Ambient Visualizations' Attributes

This user study evaluated different visualizations being representatives of a design space comprising the two dimensions "degree of abstraction" and "amount of information". Collecting quantitative and qualitative data, a descriptive statistical analysis was performed. Based on the results, the subjectively perceived best visualization will be concluded and thus, deployed it in a field study to provide ubiquitous feedback on a desk worker's cognitive load.

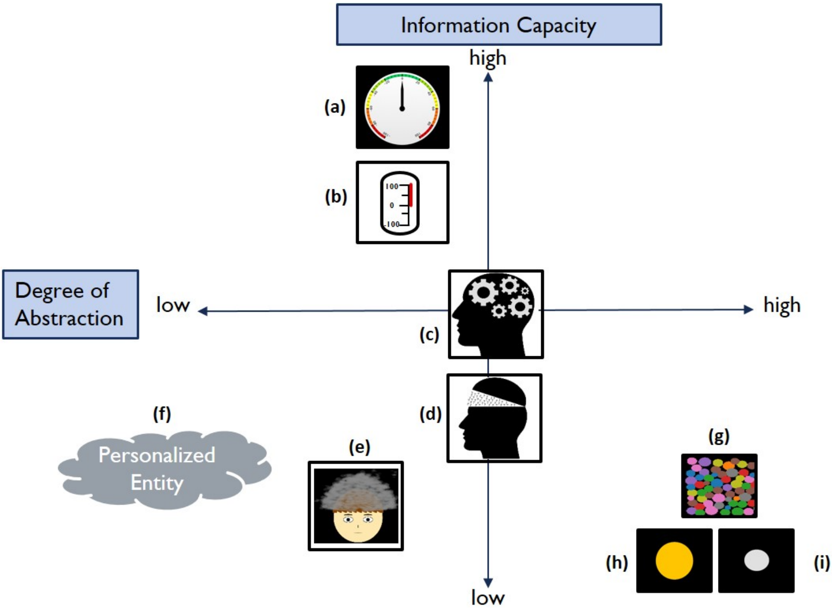


Figure 6.5: Two-dimensional design space illustrating the degree of abstraction on the x-axis and the amount of information on the y-axis. The nine following visualizations represent different parts of the spectrum characterized by the two dimensions: (a) Speedometer, (b) Thermometer, (c) Working Head, (d) Question mark-Head, (e)Smoking Head, (f) Personalized Entity, (g) Multiplying Circles, (h) Colored Circle, and (i) Scaling Circle.

Study Design and Measures All visualizations were presented to each participant following a within-subject design. The order of visualizations was randomized according to Latin square and for each visualization the following five dependent variables have been assessed on a Likert-item scale ranging from 1 (=I do not agree at all) to 4 (=I fully agree):

Intuitive Comprehensibility The statement "*The visualization is comprehensible.*" referred to the effort a user has to make to understand what a visualization is showing. This aspect has been derived from Nielsen's [187] *10 Usability Heuristics for User Interface Design* which phrase recommendations to make a system intuitively comprehensible.

Subjective Distraction Asking about the distraction, the statement was phrased as follows: "*The visualization is not distracting.*" assessing if the user was prevented from concentrating on something. According to Kirkham et al. [137] this requirement was particularly valuable for deployments in the workspace.

Unobtrusiveness While the latter aspect considered the visualization's potential to annoy users, obtrusiveness refers to the ambient quality of it. Hence, it was stated "*The visualization is unobtrusive*". As can be seen in prior work [64, 113, 290], many visualizations aim to be unobtrusive for users.

Aesthetic Pleasantness Since also visually appealing factors play a role in the evaluation of visualizations- even if it is subconsciously judged, the statement "*The visualization is aesthetically pleasant.*" has been included in the assessment. Hereby again one of Nielsen's [187] *10 Usability Heuristics for User Interface Design*, namely the *aestics and minimalist design* have been included in the assessment.

Sufficiency of Information Content The last aspect which was rated refers to the amount of information being presented in the visualization. It has been identified as an important criteria in related research [212]. To get a sense for whether the data has been presented without too much detail but with sufficient information to be able to be understood, the final statement was "*The information content of the visualization is sufficient*".

Furthermore, the participants were asked to provide qualitative feedback on each visualization. Therefore, ten questions were asked partly reassessing what has been inquired through the Likert-item scales. Accordingly, the first two questions asked what was understood from the visualization and in how far the visualization has been understood intuitively. Then, the participant should explain his or her impression of the visualization and what is being associated with it. Referring to it's potential to distract the users, the participant was asked in how far the visualization has not been distracting first, before asking in how far the visualization has been unobtrusive. Afterwards a similarly phrased question considered the aesthetic pleasantness. Lastly, two questions addressed the amount of information content. While the first question: "*Do you miss any information in this visualization?*" referred to the sufficiency, the latter question targeted to assess "*Which information do you perceive as redundant?*". Finally, the participants were asked if they would like to add some additional comments. The qualitative assessment part was complemented by the individual rating of the three best visualizations given the order 1, which is the best and 3 being the least best. If they preferred, the participants could add explanations why they had chosen this sequence for these triple of visualizations.

Inducing and Measuring Cognitive Load Since the ambient visualization should provide feedback on the user's cognitive load, the so-called N-back task was used to elicit cognitive load in the users. Among the different realizations that have been applied in prior research and shown to induce cognitive load effectively [117, 196], for this study a sequence of numbers ranging from one to nine had been played to the participant only on audio basis using headphones. Without knowing when this sequence ends, the task was to remember the nth element of the sequence and speak out loud what the it was. Hereby, five N-back tasks ranging from 1-back to 5-back have been performed in difficulty increasing sequence, to avoid potential feelings of overload among the participants resulting in dropouts. For measuring physiological data, namely heart rate variability (HRV) has been recorded throughout the performance of the N-back task. This physiological measure has shown to be a reliable indicator of cognitive load [161, 164, 174, 247]; while high HRV values signify a low mental demand, low values indicate high cognitive load. To ensure that participants felt cognitively challenged, their experiences regarding mental, physical, and temporal demand, as well as performance, effort, and frustration should be rated using the NASA Task Load Index (NASA-TLX) [103].

Data Transfer The physiology recording wearable Empatica E4 [13] has been attached to the participants' wrist. The device stored the physiological data locally and transferred them via Bluetooth Low Energy to the Android "E4 realtime" mobile application [14] running on a Motorola Moto G3 smartphone. After each sub-task of the N-back task had been completed, the session data was uploaded from the mobile application to the cloud-based "E4 Connect Web Dashboard" repository [15] being requested from a Sony Vaio laptop. Via the user interface a CSV file with the raw data could be downloaded.

Data Processing For the processing of the heart rate values being recorded with a sample frequency of 64 Hz, the R-R interval values were used to calculate the rMSSD value and threshold values to be used for visualizing the cognitive load. The calculation of the rMSSD value follows the formula

[13] `https://www.empatica.com/research/e4/`

[14] `https://play.google.com/store/apps/details?id=com.empatica.e4realtime`

[15] `https://support.empatica.com/hc/en-us/sections/200002718-E4-connect-Web-Dashboard`

6.3.1 being frequently used in related work [161, 164, 165]. Hereby, the rMSSD is defined as the square root of the averaged sum of all squared neighboring values (t_i). Thus, the rMSSD value refers to the time in milliseconds between two heart beats [247].

$$\Sigma_{\mathrm{sd}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[t_{\mathrm{i}+1} - t_{\mathrm{i}}]^2} \tag{6.1}$$

Participants and Procedure 18 individuals ($M = 22.7$, $SD = 2.72$ years) among them three females participated in the study. They were recruited via personal contact or mailing list. The study took place in a quiet room and took between 90 and 120 minutes per participant depending on the individual length of the qualitative part. Participants received 5 Euro per each 30 minutes, meaning at least 15 Euro for their participation.
Before the user study started all participants were explained the purpose and the sequence of the study addressing particularly the collection of physiological data. After the participants had provided their consent, the Empatica E4 wristband was attached to their non-dominant arm. Then, they were asked to put on the provided headphones and to relax for two minutes, while physiological data had been recorded as a baseline measurement. Subsequently, the cognitively demanding N-back task was explained in detail; each participant started with the 1-back task followed by the incrementally increasing more difficult task up to the 5-back task. For reassuring that cognitive load had been induced, the NASA-TLX was administered after each performed N-back task using an online version[16] for the assessment. When the last N-back task had been accomplished, all physiological data was exported from the Empatica E4 to CSV files and processed as described in Section 6.3.2 resulting in a cognitive load value given in percent which needed to present the ambient visualizations. Accordingly, the participants could take a short break before the nine visualizations were shown in randomized order. Each visualization was presented showing the full spectrum of cognitive load that had been perceived throughout the preceding task ranging from the individually determined minimal value to the maximum value. For each visualization the five dependent variables were assessed quantitatively using the Likert-item scales, and qualitatively asking the questions set (cf., Section 6.3.1). All interviews were audio recorded given the participants' consent. The entire study procedure is illustrated in Figure 6.4.

[16] `http://jensgrubert.bplaced.net/nasa-tlx-short/TLX-English-short.html`

6.3.2 Implementation

In the following, the system components will be presented and the it will be explained how the nine different visualizations have been realized.

System Architecture There are three main parts: First, the data transmission from the Empatica E4 to the Server, then the calculations performed on a server, and finally the data transfer and consequently the visualization of the physiology-based cognitive load on an Android tablet. After the raw data transmission being described in Section 6.3.1 was completed, the CSV files were processed on a local server using the Python web framework Flask [17]. Based on the calculated rMSSD values (cf., Section 6.3.1) for the baseline session and the particular N-back task (1-back to 5-back tasks), two threshold rMSSD values can be deduced signifying the minimal and the maximum cognitive load perceived throughout the experiment. Using the following formula a value of cognitive load in percentage is being calculated by subtracting the minimal rMSSD value from the obtained input value and the preliminary derived maximum rMSSD value and dividing both, so that the result needs to be multiplied with 100 to receive the percentage:

$$P_{\text{cog}} = 100 \left[1 - \frac{\Gamma_{\text{sd}} - \Gamma_{\text{sd,min}}}{\Gamma_{\text{sd,max}} - \Gamma_{\text{sd,min}}} \right] \tag{6.2}$$

For the final part, namely representing the cognitive load values given in percentage in an ambient visualization, the values are provided as JavaScript Object Notation (JSON) objects based on the Hypertext Transfer Protocol (HTTP). The resulting JSON objects were sent to a 13.3 inch Hannspree HSG 1351 tablet operating on Android version 5.1. They were displayed using an Android application implemented in Java and Kotlin presented via the system component Android System WebView [18].

Visualization Implementation To be able to display the visualizations in a mobile application using Android System WebView, the JavaScript libraries D3.js [19], Highcharts JS[20], and p5.js[21] were used. The environment was created in

[17] `http://flask.pocoo.org/`

[18] `https://play.google.com/store/apps/details?id=com.google.android.webview&hl=de`

[19] `https://d3js.org/`

[20] `https://www.highcharts.com/`

[21] `https://p5js.org/`

HTML5, a version of the HyperText Markup Language (HTML) and Cascading Style Sheets (CSS). In the following nine implemented visualizations will be described in detail starting with those visualizations that have the highest degree of abstraction up to the least degree of abstraction as explained in Section 6.3.

Speedometer Like Kirkham et al. [137] did with their barometer, an object being known from everyday life was implemented with the speedometer. Indicating a range between -100 and 100, more information content was added. This range of numbers was divided into intervals of 20, whereas the zero point signified the equilibrium between under- and overload. The negative numerical range up to -60 referred to medium, respectively the -80 to -100 range to heavy underload. Likewise, the positive numbers within this range indicated overload. For this visualization again the gradient from green to red was used (cf., Figure 6.5 (a)) and the speedometer needle placed in the center was adjusted according to the cognitive load.

Thermometer Following the approach of providing a number-based representation of cognitive load, the thermometer covered the same numerical range from -100 to 100 again determining the zero point as the equilibrium between under- and overload. While again hereby the similarity to a known object was paramount, the difference to the previously explained visualization was, that there was no item indicating the cognitive load but the colored bar inked in the green-to-red gradient signified the cognitive load itself (cf., Figure 6.5 (b)). Thus, the numerical range was visible but the color in addition to the negative or positive range indicated if under- or overload had been detected.

Working Head For this visualization the motive of a laterally shown abstractly sketched head was used inspired by the approach to present metaphorical visualizations [290]. Hereby, the amount and size of five rotating cog wheels signified the recorded cognitive load. The numerical range between 0 and 20% was represented when the smallest part was rotating; the next level indicating cognitive load from 20% to 40% was reached when the smallest and the next bigger cog wheel were working; taking the third cog wheel in addition, the range 40% to 60% was covered; the indication of 60% to 80% of cognitive load was achieved by the rotation of the fourth cog wheel; for the representation of the maximum cognitive load, respectively 80% to 100% all cog wheels were working. In this visualization the speed of all wheels was kept constantly and they did not differ in their speed but in their size signified more cognitive load. This visualization was kept in black and grey (cf., Figure 6.5 (c)).

Question mark-Head Requiring less abstraction skills, this visualization consisted of the lateral view of an abstractly sketched head whose top of the skull opened in correspondence with the amount of cognitive load. From the skull's inside multiple question marks appeared with increasing opening. This visualization was also kept in black and grey (cf., Figure 6.5 (d)).

Smoking Head The final visualization taking up the motive of a head, was inspired by the image of a smoking head. Hereby, the frontal view of a head including the face was shown and the amount of smoke surrounding it functioned as an indication of cognitive load (cf., Figure 6.5 (e)). This visualization provided little information on the exact amount of cognitive load detected and relied on the intuitive understanding that more smoke was linked to a higher amount of cognitive load.

Personalized Entity Representing the extreme of Yu et al.'s [290] metaphorical visualization, hereby a user could pick any object he or she associated either with overload or with relaxation to be his or her "personalized entity". Accordingly, the object's size was adjusted corresponding to whether the amount of cognitive load was high or low (cf., Figure 6.5 (f)). For example, an object reminding the user on cognitive strain would increase it's size when cognitive load would rise.

Multiplying Circles As another visualization, the amount of circles varying in their size and color have been used for representing the amount of cognitive load; this approach resembles Fan et al. [64] visualization for indicating the number of steps. For signifying minimal cognitive load the ratio between the blank black background and the number of circles was high, meaning only few circles. Correspondingly, a full screen with hardly any blank black background visible meant a high cognitive load (cf., Figure 6.5 (g)).

Colored Circle The colored circle applied a green-to-red gradient for representing the cognitive load value given in percentage. Hereby, green referred to the minimal cognitive load, while red signified the maximum (cf., Figure 6.5 (h)). This traffic light color scheme is commonly suggested [169] for providing clear signals being understood intuitively.

Scaling Circle The idea to use a scaling circle was inspired by [64], who represented the number of steps per day with the varying amount of rings. Similarly, the scaling circle used for the presented user study changed the diameter according to the cognitive load value given in percentage. Since the circle's color should not signify anything, a neutral lighter gray was picked (cf., Figure 6.5 (i)).

6.3.3 Results

From the correct answers and the subjective self-reported data of the NASA-TLX it can be inferred, that participants were cognitively demanded by the N-back tasks according to their increasing difficulty. In Table 6.3, it is shown that the number of correct answers diminished throughout the incrementally ascending N-back tasks. Likewise, the NASA-TLX scores indicate that participants perceived their cognitive load higher with rising difficulty. Interestingly, the standard deviation for both measures is highest in the most difficult level, namely the 5-back task suggesting that the variance among participants' performance and respectively cognitive load rating differed more tan for the other tasks.

N-back Task	**Cognitive Load Measures** (M,SD)	
	Correct Answers	NASA-TLX Score
1-back Task	6.83 (0.37)	33.37 (13.27)
2-back Task	5.5 (1.74)	46.48 (17.44)
3-back Task	4.72 (1.52)	62.36 (17.12)
4-back Task	3.28 (1.85)	67.54 (17.26)
5-back Task	2.9 (2.08)	68.01 (21.17)

Table 6.3: Means (M) and standard deviations (SD) for the cognitive load measures correct answers, and the NASA-TLX score per each N-back task.

The Visualization's Suitability With the aim to explore the suitability and preference as well as qualities of the ambient visualizations, intuitive comprehensibility (IC), subjective distraction (SD), unobtrusiveness (Un), aesthetic pleasantness (AP), and sufficiency of information content (SoIC) had been rated for each visualization (cf., Table 6.4). Additional to the qualitative data, the qualitative feedback on the particular visualizations had been gathered in semi-structured interviews.

Quantitative Data Looking at the mean results over all variables (see Table 6.4), the working head (c) received the best scores ($M = 3.22$). It was the only visualization which received a 3 (="I agree") on four of the five scales. However, the working head (c) underperformed with respect to

Variable	Visualization (M,SD)								
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
IC	3.28	3.17	3.11	3.28	3.72	3.11	2.56	3.06	3.33
	(0.75)	(0.79)	(1.08)	(0.83)	(0.46)	(0.96)	(0.86)	(0.87)	(0.77)
SD	2.94	2.78	3.44	2.72	2.61	2.72	2.33	3.22	3.39
	(0.94)	(0.73)	(0.7)	(1.07)	(0.7)	(1.02)	(0.97)	(0.73)	(0.61)
Un	2.89	3.0	3.5	3.06	2.89	2.89	2.61	3.06	3.44
	(0.9)	(0.87)	(0.71)	(0.87)	(0.83)	(0.83)	(0.78)	(0.8)	(0.62)
AP	3.28	2.89	3.39	3.0	2.33	2.0	2.89	2.89	2.83
	(0.67)	(0.9)	(0.7)	(0.84)	(0.77)	(0.84)	(0.76)	(0.68)	(0.79)
SoIC	3.28	3.22	2.67	2.94	2.89	2.61	2.72	2.89	2.89
	(0.57)	(0.81)	(0.97)	(0.64)	(0.76)	(0.92)	(0.67)	(0.76)	(0.83)
Mean	3.13	3.01	3.22	3.00	2.89	2.67	2.62	3.02	3.18
(SD)	(0.20)	(0.19)	(0.34)	(0.20)	(0.52)	(0.42)	(0.21)	(0.14)	(0.29)

Table 6.4: Means (M) and standard deviations (SD) for the subjective rating on intuitive comprehensibility (*IC*), subjective distraction (*SD*), unobtrusiveness (*Un*), aesthetic pleasantness (*AP*), and sufficiency of information content (*SoIC*) according to the following nine visualizations: (a) Speedometer, (b) Thermometer, (c) Working Head, (d) Question mark-Head, (e) Smoking Head, (f) Personalized Entity, (g) Multiplying Circles, (h) Colored Circle, and (i) Scaling Circle.

the sufficiency of information content. The majority of the visualizations, namely the speedometer (a), thermometer (b), question mark-head (d), smoking head (e), multiplying circles (g), colored circle (h), and scaling circle (i) had been rate better than the working head (c). Although their information content had been perceived more sufficient, they were given poor scores for at least two other variables. While the scaling circle (i) and the colored circle (h) had their weaknesses in the aesthetic pleasantness and in the sufficiency of information content, the question mark-head (d) also underperformed in the latter aspect but was also perceived as distracting ($M = 2.72$). Only the speedometer (a) and thermometer (b) were rated to provide sufficient information content. Hereby the speedometer (a) was slightly preferred among all visualizations also having the lowest standard deviation ($M = 3.28$, $SD = 0.57$). For the speedometer (a) the disadvantages had been it's potential to distract connecting to the missing unobtrusiveness. The thermometer (b) was also rated to have a comparibly high potential to distract, but had the second weakness in the aesthetic pleasantness. A little less preferred were the smoking head (e) ($M = 2.89$) performed best regarding it's intuitive comprehensibility also having the least standard deviation ($M = 3.72$, $SD = 0.46$). Likewise, the personalized entity (f) was rated to be less suitable ($M = 2.67$), although it was perceived to be intuitively comprehensible. Both visualizations performed comparably poor on the other criteria having not even 3's on the other scales. The lowest scores over all variables ($M = 2.62$) was given to the multiplying circles (g). For this visualization participants never rated to "agree" with the statements reflecting the variables.

Ordinal Rating As can be obtained from Figure 6.6 , the scaling circle (i) and the speedometer (a) had been rated almost equally good; both being mentioned by five participants each to be the most preferred visualization. Still liked but on the first rank positioned by only two users, was the working head (c). Those three further appeared two times, respectively four times (a) on the second rank and four times, respectively three times (a) on the third rank. The other visualizations b, e, and g were in the midfield being preferred by only a small number of users. In contrast the colored circle (h), the question mark-head (d), and the personalized entity (f) gained a minority of votes.

Qualitative Data The qualitative feedback on the suitability of the visualizations showed to be ambiguous among the participants. Those attributes some users appreciated the most, were perceived negatively by others. For examples, the aesthetic pleasantness of the scaling circle which was emphasized by 11 users to be "*beautifully unostentatious*" (P16) and "*simple*" (P12) was considered to be "*[...] a bit boring*" (P1). At large, the users emphasized that they disliked too many colors or objects, e.g. in the multiplying circles visualization, from an aesthetic perspective, but also because too complex visualizations required much effort for the interpretation, such as the working head. With respect to the information quantity the interviewees criticized those visualizations containing already numerical ranges and colors often heavier than those having a limited amount of information. Although they realized that these were missing detailed information, it seemed to be considered as a bigger advantage displaying less feedback concisely.

In conclusion, the smoking head (e) and the personalized entity (f) have been excluded due to their limited quality in the comprehensibility as confirmed by the ordinal ratings. Further, the multiplying circles (g) was rated to have the least suitability in general. The visualizations a,b,d,h, and i have been

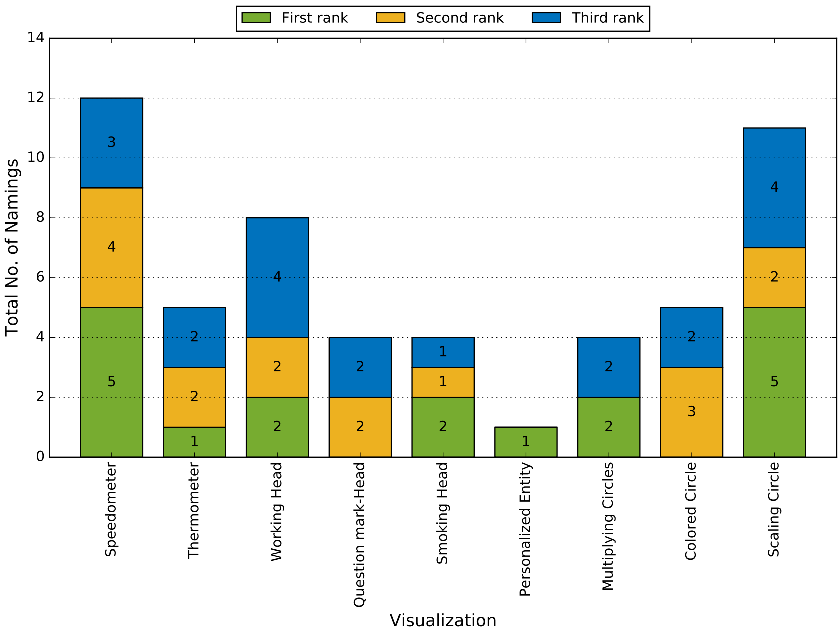


Figure 6.6: Results of the subjective ranking accumulated over all participants divided into the first, second, and third rank according to the nine presented visualizations.

perceived similarly revealing weaknesses for two out of the five assessed variables. Considering the ordinal ranking of subjective preferences, the scaling circle (i) and the speedometer (a) would be the most preferred ones for a majority of 11 and respectively 12 participants. Following the ratings for the five assessed variables, the working head (c) had good scores on even four out of the five scales and hence, one weakness less than the previously mentioned visualizations.

Facing these minor differences in the subjective preferences based on the quantitative ratings and ranks given, the qualitative feedback had also been taken into account for evaluating the suitability of the presented visualizations. Despite the fact that the working head (c) was argued to be the least distracting object among the presented alternatives, which was probably due to the "*choice of colors*" as P6 and P8 remarked and the subtle changes within the visualization (P1), the participants were divided on its comprehensibility. While some interviewees intuitively understood the cog wheels to be symbols for the proportional cognitive load, respectively work (P1,P10)), others admitted to have difficulties to interpret the exact meanings of each cog wheel (P4,P14,P15). The speedometer (a) was appreciated because it provided more information on the exact amount of detected cognitive load than the other head-based or circle visualizations. However, the interviewees missed some explanations on the negative numerical range, units, and colors (P1,P2,P7,P14). Positively highlighted by 16 users was the aesthetic pleasantness. Another advantage for this visualization is its comprehensibility. As P16 explains, the symbol is known from everyday life. Further it gives a detailed values of the cognitive load (P18) and combines numbers and colors (P11). However, P2 was "*confused*" in the beginning, what contributes to the impression of six interviewees that the speedometer is distracting because "*all colors are there any time*" (P12) and the permanent needle movements (P8,P11). The main strength of the scaling circle (i) was its intuitive comprehensibility, which was highlighted to be easy to retrieve as information due to the "*size ratio*" (P6,P11). Moreover, almost all users, namely 17 agreed that the scaling circle does not distract by being "*simple*" (P2) and not offering "*too many stimuli*" (P14). Being aesthetically pleasant for 11 interviewees due to being "*beautifully unostentatious*" (P16), the only weakness was the information content which could have been more sufficient by adding a value for the actual cognitive load (P5). Weighing the described advantages for each of the three favorites, the scaling circle (i) emerged as the most suitable visualization being easily comprehensible, unobtrusive, not distracting, and aesthetically pleasant. Thus, it was taken for the deployment in a real-world setting.

6.3.4 Prototypical Deployment in a Semi-Public Setting

With the aim to provide ambient visual feedback of a potential stressor, here cognitive load, the most suitable visualization, namely the scaling circle was deployed in-situ in a real-world scenario in the wild. For the deployment a desktop work setting had been chosen envisioning that individuals working in multiple scenarios could benefit from having their cognitive load visualized.

Study Design and Measures For deploying the visualization the semi-public setting of a shared office has been chosen. For this, participants were asked to use the ambient visualization of their cognitive load for five consecutive working days. Hereby, the HRV was constantly measured with the Empatica E4 and processed in real time, so an in-situ feedback of the cognitive state was displayed on a 9.7 inch Samsung Galaxy Tab S2 tablet with an Android 7 operating system (cf., Figure 6.4). In an in-depth interview after these five usage days, open-ended questions were asked to gather feedback on potential effects of the deployment on in general, as well as specificly on (a) the user's productivity, (b) the user's perception of reactions of potential bystanders, and (c) the user's perception of his or her privacy preservation in this context.

Questions Set First, the general impression on the visualization and whether it was liked or not including the reasons were inquired. Then, it was asked whether potential effects, e.g. on the everyday work life had been experienced. In conjunction to this, the particular setting of having the display visible for colleagues in the office had been addressed. Hereby, it was asked for the user's feelings regarding this visibility and what the office colleagues' reactions in this context had been. Lastly, potential improvements and further remarks could be mentioned, if the participant wanted.

Participants and Procedure Two participants ($M = 47.0$, $SD = 4.2$years), one female and one male, both working as administrative staff at a university had been recruited via personal acquisition. Before the study started, they were explained the purpose and the sequence, as well as the recording of physiological data with the wristband Empatica E4. After they had signed the consent form, they were asked to perform the N-back task (see Section 6.3.1 for a detailed description) eliciting cognitive load for obtaining a baseline measurement which was used to calculate the individual minimum and maximum values, which were needed for the in-situ feedback visualization. After this initial task had been performed, the users were given a self-written manual, which explained how the data recording could be started and stopped including a legend on the LED colors that have different meanings referring to the Empatica E4's status. After the usage period, the open-ended questions (cf., Section 6.3.4) were asked in an in-depth individual interview.

6.3.5 Implementation

To deploy the system, physiological data had been recorded using an Empatica E4. The data transmission and processing followed the same steps as realized in the *Preliminary Evaluation of Ambient Visualizations' Attributes* (cf., Section 6.3.1). Therefore, the system architecture and the formulas used for the processing can be obtained from the description in the Sections 6.3.1 and 6.3.2. Based on the findings from the *Preliminary Evaluation of Ambient Visualizations' Attributes*, the scaling circle was implemented for the deployment of ambient feedback on cognitive load.

6.3.6 Qualitative Feedback on the Deployment of an Ambient Feedback for Cognitive Load

In general, the participants found the study interesting and were happy to test this approach in a real-world setting. Regarding the visualization used, P1 perceived it as "*pleasant*" and "*not distracting at all*". Particularly, the lack of red and green which are related to judging criteria had been appreciated, since it had not been perceived as "*offending*". For P2 providing more information would have been desirable. He suggested an overview of one's data including the past hours because "*for [him] it would be interesting to see what was going on in the past quarter-hour*". Speaking of potential improvements, both users questioned the utility of the system, since they had the impression that their subjectively perceived physiological state did not always correspond to the displayed feedback which was based on the in-situ recorded HRV values. P1 reported that she had tried to deliberately control the visualization what she could and that this diminished the trust in the system. Based on this observation, she admitted to call into doubt whether her physiological values correlated with her cognitive load. P2 added that although the relation between cognitive load and physiological had been "*postulated*" in science, he was unsure whether the recorded activity reflect the reality. As a reason for this lack of belief, P2 mentioned the influence of "*external factors leading to a varying heart activity*". Nevertheless, P1 and P2 emphasized that such a system would be useful to "*sensitize*" (P1) and thereby become fully aware of one's own body signals. While P1 stated that she could imagine using it for a few days or maybe weeks to train the self-reflection and utilize the system as a mean to be reminded to "*listen to [one]self*", P2 suggested further usage contexts, such as therapy settings. However, both interviewees considered the feedback as a temporal mean for supporting self-reflection highlighting that it would be preferable to not need such a system because it can be "*critical to rely on systems*" (P1) enabling us to be aware of body signals. For P2 it is important that "*one learns to pay attention to one's own senses without measurement tools*", since "*the human has sufficient sensors to sense it*" (P2) and it could be questioned if such a system might not be "*counterproductive*" (P1) in the end. Though, P2 admitted that for people lacking this ability, such a feedback system could be used to train them.

With respect to sharing the cognitive load level, P1 reported that colleagues were "*interested in the system*" in the beginning. She highlighted that the system also leaded to misinterpretations, since colleagues were understanding the visual representation of the participant's cognitive load as a sort of "*thought-reading*". Accordingly, some situations had been commented on like "*Oh, this stresses you* *a lot*" (from an office colleague). As a result, P1 realized that, if the relation to her colleagues would not have been as good and relaxed as it is, such reactions would have made her feel embarrassed. Despite the casual atmosphere, P1 admitted that she would have felt "*uncomfortable*" when her colleagues drew conclusions from the cognitive load feedback and commented on it, particularly when bystanders from other offices noticed the system. Hence, the interviewee disliked the idea of a permanent deployment in her office, particularly among people she does not know very well. Although P2 did not perceive the problem as serious for himself, when he had visitors in his office, he mentioned that he could imagine that such a system would violate one's privacy when being deployed in open-plan offices. Saying that "*we have the tendency to abuse such data*", P2 added that these kind of information "*should be private any time*". Both users agreed that the presentation of individual data of one's cognitive load, particularly not in private settings is a highly sensitive topic that needs to be treated carefully.

6.3.7 Study Conclusion

As could be inferred from the preliminary evaluation, the suitability of ambient visualizations for representing cognitive load is perceived ambiguously. Thus, three visualizations, namely the scaling circle (i), the speedometer (a), as well as the working head (c) received almost equally good results in the quantitative ratings and preference ranking. However, favoring these three divergent visualizations, each fitting into a different quadrant of the design space (cf., Figure 6.5), indicates the complexity of finding suitable visualizations. Thus, the qualitative feedback gathered throughout the experiment had been taken into account for identifying the most suitable visualizations for the purpose of signifying cognitive load in a semi-public desk work setting. To conclude, the scaling circle was found to have the least disadvantages given the deployment scenario. For example, participants appreciated the unobtrusiveness and the aesthetic pleasantness, since it was kept simple.

Consequently, in the deployment the scaling circle was used to provide visual feedback on the physiologically sensed cognitive load using HRV values. Gathering qualitative feedback from two participants deploying the system for five consecutive working days in their (shared) offices, revealed that the system is more appreciated in specific situations. Hereby, participants referred to the natural quality of the human body to send and interpret signals that indicate cognitive overload, highlighting that only if this ability needs to be trained such a feedback system would be useful. Regarding the particular semi-public setting, both users agreed that making such kind of information publicly visible the privacy could be potentially threatened when colleagues would start to draw conclusions from the shown data. To sum up, the findings from the preliminary evaluation and the prototypical deployment indicate that the design of feedback visualizations representing sensitive private data, such as the cognitive load level must be handled with care. Not only the preferences for how suitable visualizations should look like to be comprehensible, unobtrusive, not distracting, aesthetically pleasant, and sufficiently informative are underlying subjective impressions, but also the perceived benefit of a system that provides ambient feedback on one's cognitive state. In any case, already the preoccupation with the sensing and representation of cognitive load supported the users' consciousness and awareness regarding the importance of listening to one's own body signals as an indication of stress.

6.4 Visualizing Stressors Requiring User Action

After having explored, what the effects were when users took a passive role in the visualization process of stressors, the present study investigated the consequences of requiring users' actions when making the stressor publicly visible. Hereby, not only the users' and bystanders' attention is drawn to the stressor, but in contrast to the previously described approach, the user needs to actively adjust how prevalent he or she perceives the stressor. This manipulation opportunity also allows the wearer him- or herself to manipulate one's own perception, when he or she is cheating at the self-reported feeling of being busy, respectively stressed. A public setting had been chosen to examine potential implications for privacy preserving systems that use physiological data. For the user study, a wearable prototype being able to display the perceived level of business was manufactured. This prototypical display could be worn as a necklace being visible to anyone in the spatial sphericity of its user.

6.4.1 Manufacturing a Wearable Display Prototype

For the wearable prototype plywood had been cut into the shape of a circle with a diameter size of six cm, having three slight notches which were painted in green, yellow, and red. These colors were chosen for their signaling functionality resembling a traffic light. Red referred to being totally busy, yellow signaled that

(a) (b) (c)

Figure 6.7: The manufacturing of the prototype consisted of different steps. The background plywood discs needed to be painted (a) before the second layer was fixed with a screw (b); a white knob was finally attached to the screw to facilitate rotating the second disc (c).

the wearer is somewhat busy, and green indicated not being busy. In the middle of the circle a tiny screw holds the second layer consisting of another wooden disc. To facilitate rotating the screw with the second disc, a white knob was glued onto the screw. The manufacturing process is depicted in Figure 6.7.

To allow users to modify and customize their prototype, the outer edges have been perforated framing the word "*busy*" or "*kire*" in Finnish language to clarify what the display is showing. Enabling the attachment of a band to carry the prototype around one's neck, two wholes were laser cut in right above the lettering.

6.4.2 Diary Study

For conducting the diary study paper-based booklets have been designed and distributed among the participants being collected after the study ended.

Study Design and Measures The study was carried out in the field, so that no restrictions were made for the participants regarding when or for how long to wear the *BuSiNec*. The 12 questions in the diary were complemented by the invitation to "*write or draw [one's] thoughts, ideas, notes, etc.*" and repeated

Figure 6.8: User wearing the *BuSiNec* indicating one's subjective feeling of being busy with the colors red meaning to be busy, yellow referring to an intermediate state, and green signaling to have time.

for each of the potential five usage days. As measures to evaluate the impact of wearing a self-adjustable stressor visualizing display, the usage behavior, the effects noticed with regard to one-self and to others, as well as the perceived comfort and usefulness had been assessed.

Questions Set The first three questions targeted the individual usage behavior. They referred to the length of wearing time, the situations when the prototype had been used and what had been initiated the adjustment of the display. The next two questions inquired about the feelings and experiences while wearing the *BuSiNec* before the reactions in the social environment and the amount of people, who showed interest should be reported. Besides naming what the prototype should be able to do or display, it was further asked if any effects on the stress level had been noticed. Then, on a Likert-item scale ranging from 1(=not useful at all, very uncomfortable) to 7 (=very useful, very comfortable) the usefulness and respectively the comfort should be rated. Lastly, the distribution of the level of business (green, yellow, red) should be given as value in percentage.

Participants and Procedure From the initially recruited 16 potential users, 13 participants returned their diaries. Two out of these were excluded from the data analysis since they handed in incomplete data. Acquiring the users via personal contact, each one was informed about the purpose of the study and their rights to end or continue it right away. After they had agreed to participate, the prototypical display was handed out together with the paper-based diary. The only instructions given to them were that they should wear the *BuSiNec* (cf., Figure 6.8) whenever they felt comfortable and that they should report on their experiences in the diary. The length of the study varied because users were left their own choice for how long they wanted to use the prototype.

6.4.3 Results

In the following, the results from the diary study will be presented including quantitative and qualitative data. The participants were left a choice whether to write down their thoughts in English or native language; those from Finnish language have been translated into English language.

Usage Behavior In the context of capturing genuine usage behavior, the participants were free to decide how much an in which situations they wanted to wear the prototype. From the diary entries it could be inferred that they used the *BuSiNec* 4.00 days at average ($SD = 1.35$). While eight participants wore it for four or five days, it was used by one user each for three, two, and one day comprising 44 usage days in total over all wearers. All users wore the wearable self-adjustable display throughout their working days and their work activities, such as in their offices, while teaching students, during lunch and coffee breaks, and even during meetings and workshops. Based on their estimations the *BuSiNec* had been used for 6.55 hours at average per day ($SD = 3.63$). While one wearer admitted to have forgotten to take it on one day in the morning, four others reported to continue wearing it after work. Revealing more details, two participants mentioned that their kids "*understood the color coding immediately*" (P12). P3 said that for her 6 year old daughter it was easier to notice when they were in a hurry due to the red color. Correspondingly P12's kids reminded her to change the color in the evening, since she was "*no longer in a hurry*" from her kids' perception.

Inquiring what exactly had initiated changes in the color and respectively indicating the level of business, more than one third of the users (P2,P4,P10,P13) named that they changed the business level display to red when they had specific tasks or "*urgent duties*" (P2,P10) to do. Those activities for which they adjusted the prototype according to their needs, had been for example phone calls, prototyping, reading or processing papers, or simply being in a hurry as P3, P5, and P12 admitted. P3, P9, and P11 used the wearable display to let others know whether they had time for them or not corresponding to having the "*feeling to be more/less busy*". Five participants agreed on showing green when they went to coffee breaks or lunch (P4,P7,P9,10,P13). For 27% of the usage days over all wearers when the *BuSiNec* had been worn, participants reported that they did not change the color because their subjective feelings had not changed either or some just had forgotten to adjust it in the particular situations.

Utility of the Wearable Display Prototype Regarding the perceived utility of the *BuSiNec* the participants had divergent opinions, which even differed

Participant	Displayed Business Level (M,SD)			Usage Days
	Green	Yellow	Red	
P2	0%	76% (35.78)	24% (35.78)	5
P3	52% (8.37)	40% (0.0)	8% (8.37)	5
P4	15% (5.00)	55% (22.91)	30% (26.46)	3
P5	54% (16.73)	29% (21.33)	17% (9.75)	5
P6	100%	0%	0%	1
P7	30% (0.0)	35% (7.07)	35% (7.07)	2
P9	22% (22.80)	18% (14.83)	60% (31.62)	5
P10	14.25% (10.90)	43.75% (37.72)	17% (22.42)	4
P11	47.50% (34.03)	17.50% (12.58)	10% (11.55)	4
P12	40% (54.77)	50% (50.00)	10% (22.36)	5
P13	50% (39.53)	42% (32.71)	8% (11.51)	5
Total	35.05% (32.32)	39.89% (31.39)	20.52% (25.20)	

Table 6.5: Means (M) and standard deviations (SD) indicating a self-reported estimation of how much the wearer adjusted the specific level of business including the total usage days.

among the individual usage. Thus, the standard deviation was almost 2 points ($SD = 1.97$) on the Likert-item scale. At average the users perceived wearing it as somewhat useful ($M = 3.66$). In contrast, the subjective comfort while wearing was much higher at average ($M = 4.99$, $SD = 1.45$) indicating that although the *BuSiNec* had not been regarded to be particularly useful, wearing did not affected their comfort negatively. Looking at the mean values depicted in Table 6.5, it becomes obvious, that the green and yellow level had been used almost equally often throughout the day, while red was used 15-20% less. Remarkably, the standard deviation was extremely high for all levels displayed, which can be also seen in the individual means for each participant.

Perceived Effects and Public Reactions In general, users liked the *BuSiNec* naming positive outcomes like: "*I paid attention to how I use my time*" (P5). Accordingly, they observed that they were "*more aware*" (P3) and "*more conscious*" (P12) of their business levels after using it only for a short period of time, but "*noticed better when [they] got more relaxed or stressed or were in a hurry*" (P3) after wearing it a couple of days. Following on this, others said that they started "*thinking about the meaning of being busy*" (P5), and that using the diary in addition to the prototype "*you pay attention to the feeling of business and how you feel*" (P12). P10 referred to actual benefit

saying that "*BuSiNec helps me also recognize the not-busy moments*". Further the positive experience of allowing oneself "*to be in the 'green state'*" (P10) had been addressed. In this context P13 made an interesting observation reporting that she noticed while playing with the physical wearable display, that she was actually more stressed than the color represented at that time. For her the haptic component was most important she found, when using it during meetings as a "*stresstoy*" (P13). However, one participant argued that "*thinking how much of your stress you want to communicate to others*" (P12) increased her stress level. Interestingly a three days before, the same wearer had observed that the prototype and the diary increased the consciousness for one's feelings.
Despite the positive feedback, users also faced unpleasant situations while wearing the *BuSiNec*. P9 reported that a curious bystander grabbed the prototype for having a closer look, which made her feel uncomfortable given that the wearable display hangs closely to the chest, a particular sensitive body region for females. Another difficult situation was brought up by P12 explaining that she did not want to signal to students or colleagues that she is highly busy because they should not "*feel like they could not ask [her] questions*".

The prototype as a gadget itself evoked great interest, particularly among other colleagues who also tool part in the same research. For instance, P9 had been asked by another participating colleague what her "*busy level*" was, what coincides with P12's experiences and what corresponds to P3's impression to feel "*connected to those who also wore the necklace*". Similarly, the wearable prototype "*raised curiousness and wondering*" among bystanders, as P7 mentioned. Often questions like "*what is this*" (P5,P6,P7,P9) were posed or some bystanders referred to the *BuSiNec* as a "jewellery" (P5,P9,P12,P13). Once the wearable display even stimulated a conversation among colleagues about being busy as P5 explained.

From the results it can be inferred that wearing such a wearable stressor visualization display however requires some time to get used it (P3, P5,P11,P12,P13). While this familiarity on the one hand is desirable, since users reported to not pay much attention to it and do not bother wearing it, the prototype's effect can also wear off. As P7 reports "*after the first excitement [...] it seems people don't have any reactions to the BuSiNec [...]*". Therefore a persisting improvement of the working atmosphere is questionable and highly depending on the usage context probably. Even the participants themselves, as P10 admitted that after the first excitement and curiosity, the enthusiasm decreased.

6.4.4 Study Conclusion

The presented diary study aimed to evaluate the utility and the potential effects of the wearable self-adjustable display representing the users' perceived level of business. From gathered feedback it became obvious that participants used the prototype during their working days and some even beyond the professional context. Hereby, they observed that particularly for smaller kids the color coding was intuitively comprehensible and that the *BuSiNec* facilitated understanding stressful situations from the children's perspective. Speaking of the effects on social interaction, nine out of eleven participants reported to have been approached by colleagues showing interest in the prototype. While some were curious to know what it was and what the study was about, in other cases inspiring discussions about the topic of being busy have been stimulated. Correspondingly, a common feeling among the participants had been established since they felt a stronger link towards each other. In this context, the diary entries revealed that being aware of one's own level of business raised their self-reflection and made them more conscious about their feelings.

Despite the positive comments of the display's effect, the results also suggested that long-terms effects are questionable since the initial excitement for the prototype wears off. Consequently, future research will have to focus on investigating how the positive impact can persist over a longer period. Accordingly, when making use of the wearers' engagement for the design of interactive systems, the advantages and disadvantages of visualizing one's mental state have to be taken into account. Hereby, the benefit of the colleagues' increasing mutual consideration for each others moods and feelings has to be weighted up against the potential peer group pressure to wear the prototype in order to not feel excluded. For this, preserving privacy when exploiting the benefits of public visualizations remains a significant issue that needs to be handled carefully. As the present study shows, individuals react differently to visualizing their feelings. While seven users noticed that their mind went more consciously into the "green", and respectively relaxed state, another two participants stated that they did not observe any effect; another two admitted to be indecisive regarding the prototype's efficiency. This finding has been supported by the quantitative assessment of the perceived utility and comfort. To conclude, while the subjective comfort of wearing was given for all users, the utility of the wearable self-adjustable display visualizing one's perceived level of business lies in the eye of the beholder.

6.5 Discussion

In the following, the findings from the three presented studies manipulating stressors in distinct ways will be discussed addressing the effects and implications.

Eliminating Stressors The approach of stressor manipulation consisted of the elimination of the source of stress. For this, a private scenario where no bystander were involved had been chosen. Hereby, smartphone notifications were delayed in three different modes to avoid undesired interruptions for the users and to take away the response pressure from them, as two sources of stress.

As part of the analysis of the interview data, some dropped-out participants reported, that for them not being reachable or contactable for their social circle have had severe effects on their social interaction. P9b admitted that it felt "annoying" and "cased a feeling of uncontrollability". This subjective perception corresponds to the recent development that most of the everyday-life communication is being done over text-messaging. As previous work also revealed, phone calls –which would have been a possibility to reach our participants at any time throughout the study– are not the primary mean of communication anymore [52, 91, 102]. Moreover, this reflects what it is being known as the fear of missing out [190] and what could be also observed with respect to smartphone usage [61]. Accordingly, the fear to not being reachable, which has been described as "no-mobile-phone phobia" or "nomophobia" in literature [241] was further observed among the participants. This phenomenon has been researched in different disciplines, such as psychology [27], medicine [135], and neurology [136]. For example, Wang and Suh investigated how organizational workers feel if they were unable to use their smartphones [269]. They found that while some had been more productive and relaxed, others reported the opposite. The findings from Pielot and Rello [201] are in line with this indicating that while participants admitted to be able to better concentrate due to less disruptions, they also reported to be more anxious and disconnected to their social circle, what is being supported by the presented study (cf., Section 6.2) too.

All in all, the user experience regarding the delaying of notifications as stress reducing technique had been positive for the majority of participants. Almost all interviewees felt less distracted during the absence of notifications which applied for each of the three modes. P3b explained that before the study he did not think of how pleasant a day could be without checking the phone all the time. Two participants reported that they had noticed that they checked their phone less often what goes hand in hand with the self-reflection which has been triggered by the notification delivery manipulation explicitly mentioned by six interviewees.

In particular, being able to set the interval length of the delaying time according to their needs was perceived as a huge advantage compared to the fixed interval length. For this, the *UDI Mode*was rated highest. Although this mode got very positive reviews, the users wished for a mixture of the *UDI Mode*and the *SD Mode*. Considering the degree of user-control in managing notifications, they preferred being in control of the interval length meaning that it should be flexible rather than static. Further, it was crucial that important messages would be able to send at any time, which is why they would have appreciate to combine these two modes.

Having assessed the initial stress level of the participants before and after the study, it could be seen that there had been huge differences among the stress perceptions. Looking at the standard deviation of 8.12 and at what participants stated, the stress reports had been strongly dependent on the temporary state and not much influenced by their smartphone usage. Therefore, it might not have been surprising that only barely noticeable effects on the users' stress level could be seen in the quantitative data. Correspondingly, a large survey asking users about the reasons for the stress when receiving notifications and what their effort was to reduce the stress conducted by Yoon et al. [289], revealed that users could be clustered into four main groups including a set of 25 participants who experienced stress although they had adjusted their notifications. This could be another explanation for why the subjectively assessed stress level does not indicate significant effects over all user groups. With respect to the other two rated variables, happiness and activeness, apparently the happiness level had not been influenced by the notification delaying which is remarkable, since participants seemed not to be unhappier when receiving their notifications delayed. As the standard deviation is always far below 1.0, even the unsatisfied users fearing to miss something have not been affected that severely that they rated the negative extreme on the happiness scale. The results for the assessment of activeness deviate even less. Since again the standard deviation has been between 0.59 and 0.76, the users' activeness was apparently not influenced by the notification delivery manipulation.

Visualizing Stressors The other stressor manipulation approach which had been evaluated in the present chapter, is how the visualization of a stressor can affect the user. Within this question, two distinct characteristics of targeting users have been applied. While the first study put users in a passive role, namely just receiving feedback on their source of stress, the latter study required active user action for visualizing the source of stress.

The deployment of the ambient feedback application in a semi-public, namely office work scenario revealed that hereby the individual sensitivity towards one's own body and mind plays a decisive role. Accordingly, the participants suggested to use the ambient feedback representation for learning to become aware or increase the awareness of one's body signals indicating stress. In connection with the utility of the concept, the usage of such biofeedback systems in therapeutic areas was discussed. With regard to novel usage contexts, also the last study employing a wearable self-adjustable display led to interesting findings. Since some participants broadened the usage context of the stressor visualization and wore the self-adjustable when spending time with their family, they reported that the color mapping to stressed and relaxed states war particularly easy to understand for children. But also apart from this observation, the prototypical implementation of a wearable self-adjustable display had been perceived positive by the vast majority of the users. Interestingly, the results revealed that the lack of utility did not interfere with the perceived comfort of wearing the prototype. Hence, even though the amount of how much one benefited from it differed greatly indicated by the high standard deviation, wearing the prototype was still continued. Consequently, testing out the individual effects while wearing *BuSiNec* can be performed by anyone who is interested, since it does not evoke an unpleasant feeling. Speaking of the highly depending individual benefit, the noticed effects are also underlying subjective feelings, which will be addressed in the following paragraph.

Speaking about the effects of the presented manipulation techniques, the users' willingness to actually utilize such a systems needs to be considered. As has been found in the results of the second study (cf., Section 6.3), users could imagine to deploy such a system in their office for learning to listen to their body signals. Consequently, P1 said that she would use the feedback for a few days or weeks depending on when she wants to end the training. Similarly, the wearable display prototype was considered to be used "*from time to time*" (P10). Nevertheless, the high awareness can be regarded as good and bad at the same time depending on (a) how much pressure the wearer feels to communicate the actual stress state and (b) how much the wearer feels empathy for those bystanders respecting the signaled color, e.g. students do not wanting to ask something if it is red shown.

Effects of Stressor Manipulation Approaches As a main finding from the study presented in Section 6.2, it became obvious that smartphone users felt an urge to reply to notifications immediately, also because they were afraid to miss something. This is supported by the results from Church and De Oliveira [44], who identified in their research that indeed there exist expectations regarding the response raised by visual indicators of a message's delivery, i.e., the WhatsApp ticks. One of their participants explicitly phrased her frustration saying that "*[...] If you're offline then I don't expect but if you're online, it sort of means* *that it's in front of you and you are doing other stuff and you are ignoring me*". This phenomenon can be related to a *response pressure* described by Renaud et al. [220], who found that recipients felt pressure to respond when they had to deal with e-mails. Taking away pressure source, which can be a potential stressor, the approach of manipulating a stressor in terms of elimination can be regarded to result in the successful reducing resulting stress.

Another valuable observation when speaking of the overall effect of stressor manipulation had been that delaying notifications but giving the opportunity to receive messages had reduced the subjectively felt stress in smartphone users. While the *FI Mode*mode, the *UDI Mode*mode and the control condition were rated almost equally, the *SD Mode*seemed to evoke the least stress in the users. This was also supported by the qualitative findings, reflecting that users felt more comfortable when they knew that a text-message could be received in an emergency. When analyzing the interviews, it became obvious that participants were much more occupied reflecting about their smartphone usage, and particularly their notification management, when not receiving those all the time. P12a said: "*I don't want to be too dependent on my phone but I know that I am, so it was kind of nice to think about this*". Thus, the positive impact on how users started to consciously consider their smartphone usage was one of the main findings from the evaluation of eliminating stressors. Following up on the investigation of the effects when testing the stressor visualization approach, the results from the deployment study presented in Section 6.3 revealed that the pure visualization does not necessarily mean that users reflect about what is being represented through the visual feedback. As the participants reported, a considerable capacity of thoughts went into scrutinizing the representativeness and utility of the visual feedback on one's cognitive load. Instead of considering the link between work tasks and the amount of cognitive load, users were rather occupied with considering how much such feedback is needed depending on user's ability to listen to one's own body signals telling how much cognitive load is perceived as stress. From a meta view, the visualization can be therefore regarded as a catalyst for becoming aware of the fact to support one's awareness regarding one's own physiological signals. The second visualization approach putting users in an active role instead of presenting passively provided feedback, also focused on the effects of stressor visualization but in public visibility embracing the other extreme on the privacy continuum. Although, four participants doubted that wearing the *BuSiNec* had a significant effect on their stress level, almost two thirds perceived the effects of the *BuSiNec* on their stress level as "*positive*" (P5,P11), namely "*lowering*" (P5,P10). P9 confirmed that she felt, when she turned it to green, she was "*more relaxed and in red [she is] more focused and feeling more* *busy*". P10 added that the conscious change of the color fitting the change in one's mind regarding feeling busy also "*eases your mind, when you know what's going on*". One participant even stated that "*BuSiNec helps me also recognize the not-busy moments*". Correspondingly, for P10 the self-adjustable display was a tool that helps allowing oneself to switch consciously between mental states referring to the "*green state*" with respect to the color coding. Altogether, this indicates that self-reflection had been positively affected by the visualization requiring to make one's own adjustments.

Moreover, the last study presented in Section 6.4 showed that the exposure of one's mental state affects the working atmosphere. Hereby, the wearable display facilitated it for colleagues to ask in a jokingly and more playful way, whether the wearer is available, since according to the reported experiences it was easier for others to react to the visualization of the business level (P3). This observation could be traced back to what is known as the so-called "*Honeypot Effect*". This phenomenon prevalent in public displays and interactive installations [286] describes that individuals are more likely to interact and approach systems, when they are attracted by the other's presence or action [291]. Accordingly, the findings could lead to the conclusion that it is easier for people to communicate, if they are provided with an anchor point to talk about, here the *BuSiNec* prototoype. The results suggest that wearing the prototype could be beneficial for overall communication structure, since it provides the opportunity to friendly or jokingly ask for someone's availability instead of crashing into unpleasant situations. An example for this was reported by P12, who talked with her colleagues about the "*levels and subjects of business*".

Moreover, the fact that a group of employees took part in the study created a sense of community. Remarks like: "*oh you are participating too*" (reaction reported by P5) signify that connectedness had not only been felt but also expressed, which facilitates again communication from another perspective. Despite the potential positive impact regarding the mutual consideration, creating a peer group of users can also be seen ambiguously. While on the one hand the people belonging to this group could increase sensivity and awareness among colleagues, on the other hand having peer groups using such devices could lead to pressure among those, who do not want to reveal their mental states. Given the huge differences in individuals' willingness to share feelings, preserving the user's autonomy in how stress-ware interactive systems can be used becomes even more considerable, what will be discussed in the subsequent paragraph.

Implications for Interactive Systems when Manipulating Stressors
Looking at the findings from all three approaches, the degree of privateness

does not affect the ability to self-reflect considerably. Nevertheless, it makes a difference whether stress mitigating approaches have been deployed on a smartphone, on an ambient display, or using a manual prototype. The technology being used determines how visible in terms of present you want the stress reduction method to be. For this, the usage context has to be considered when an approach is being realized, bearing in mind that smartphone applications can demand the user's attention much more intensively than ambient feedback deployments. In this context, also the privacy preservation plays an important role. While employing smartphones is surely representing the most privacy respecting option, the qualitative feedback from the deployment study (cf., Section 6.3) revealed that colleagues' comments on the visual feedback of one's mental states can sometimes evoke unpleasant feelings. On the other hand, from the interviewees' statements in the last study presented in Section 6.4, it can be inferred that the design of manual wearable display raised curiosity and interest among bystanders and stimulated valuable conversation about, for instance the meaning of feeling busy. Consequently, for the design decision on which platform a stress-mitigating technique is being deployed, the advantages and disadvantages regrading privacy preservation and in particular, its effects, have to be taken into account.

Besides the consideration of the deployment technology and the privacy, it is necessary that not only the user him- or herself but also potential bystanders, such as family members or colleagues can easily comprehend the manipulation technique. While in the first study presented in Section 6.2 a mixed approach has been used, where the chat partner was only informed about the delay of his or her notification in one of the three tested modes, the second study (cf., Section 6.3) waived to provide explanations on the manipulation, particularly bystanders were left without clarification. In the final study (cf., Section 6.4) an intuitively understandable visual mapping based upon the commonly known familiarity with color meanings has been applied. As could be inferred from the participants statements, the color mapping of signifying one's level of business using the traffic light color scheme red, yellow, and green has been effective without needing further explanations. This had become particularly obvious when the users' kids reacted to the displayed colors and reminded their mum to stop indicating that she was still busy, when she was actually not. In contrast, the unknowingness of the bystanders and respectively chat partners in the two other performed studies led to misunderstandings, such as that the biofeedback application was misinterpreted as thought-reading or that not responding to text-messages had been taken offensively. Consequently, the means and elements a stressor manipulation technique employs need to be easily comprehensible from an design perspective also paying attention to not exclusively the user him- or herself.

Another important finding from the self-reports is, that people like using the stressor manipulation technique voluntarily and not because they are told to or because it is part of a working culture. In the qualitative feedback retrieved from each of the presented studies it becomes obvious that individuals appreciate to have as much freedom and autonomy in their usage behavior. Thinking of the smartphone application, users wanted to stay in control of whose text-messages should be blocked. Hereby, freedom would be granted by letting users customize the *NotModes* application to a maximum. Eventually, allowing users to be reachable as they prefer would even include those participants who reported signs of the fear of missing out. In the context of stressor visualization, users would have preferred to have more transparency regarding the working mechanism of the visual representation of their cognitive load. And referring to the wearable display, users appreciated the freedom to take the *BuSiNec* off whenever they wanted to, for example in an official meeting or just because they did not feel like wearing it. These examples support previous work by Barkhuus and Dey [10] indicating that users do not want systems that take over the full control and thereby neglect the user's freedom of choice, but rather favor approaches that combine computer-based sense-making derived from the collected information allowing them to customize and change settings according to their individual needs. The results further emphasize that the user's autonomy in using interactive systems is a simple requirement to be easily respected when designing stress-aware technology.

Limitations From the three performed studies described in the present chapter, it could be inferred that manipulating stressors supports the self-reflection regarding the coping with stressors. Although the definition of a stressor is broad and refers to those stimuli that evoke stress in individuals, it is arguable if the three constructs made subject to the studies are suitable stressors. Since not only the negative consequences of receiving notifications have been shown in prior work [179, 200], but also the mental strain through cognitive load, as well as the link between stress and perceived feeling of business [23, 175, 252], an appropriate representation of potential stressors can be assumed. Another critical issue like in many HCI studies, is the lack of participants. Despite the initially high number of recruited users, the rate of successfully accomplished participants per study was not as high as originally aimed for. One of the main reasons is surely the sensitivity of the topic; reporting or even visualizing one's mental state bears of potential of violating one's privacy. Additionally, all three presented studies took place in the field to avoid the artificiality of laboratory studies and to make usage of technology which is already embedded in our environment [51]. Consequently, this implies that users granted access to their real data, no matter if it had been notifications or the representation of their mental state. Facing this particularity, the participant numbers are still providing sufficient data to investigate the aimed research questions.

6.6 Chapter Summary and Conclusion

Answering **RQ3a** the present Chapter 6 demonstrated exemplarily that the manipulation of stressors can take different forms ranging from the elimination of stressors via the visual passive representation of the stressor and up to the dynamic self-adjustability of the stressor complemented by the requirement to actively manipulate the stressor by the user him- or herself. Moreover, such manipulations can comprise the role of context factors, like for example the variation of privacy preservation or the deployment of technical accessories. For manipulating stressors, different means of technology can be included. In the presented studies three different stages of technical aids have been used. While the first study (cf., Section 6.2) relied on the usage of smartphone enabling to investigate the effects of eliminating stressors on a personal level, for the second study (cf., Section 6.3) an ambient display in form of a tablet has been deployed to ensure that not only the participant, but also bystanders could see the visual feedback. For the final study (cf., Section 6.4) the usage of technology has been neglected and a physical display prototype has been built which had to be adjusted manually. Thus, the manipulation of stressors can further compromise the degree of technical accessories being used. This consequently affects the user's interaction opportunities and must be considered when choosing the technical platform for the design of an interactive system. As another mean to shape the occurrence of stressors, the privacy continuum can be exploited as has been done in the presented research. Accordingly, three different scenarios involving the elimination of stressors in one's personal private setting (cf., Section 6.2), the visual passive representation of the stressor in a semi-public setting (cf., Section 6.3), and finally, the visual representation of the stressor putting the user in an active role in a fully public environment (see Section 6.4) have been evaluated. Another context factor addressing particularly the interaction design perspective (**RQ3a**), is the degree of user control and respectively user involvement being applied when manipulating stressors. The three presented works include different stages of user involvement. While in the first user study a mixture of different user control elements have been provided with the three notification delivery modes, in the second user study the participant was given a passive role without any autonomous control over the ambient feedback which was provided. In the final user study, users were required to take an active part, since they needed to adjust their wearable display to be able to manipulate the stressor. From the findings, it can be inferred that varying the user's degree of involvement has an impact on the user experience referring to the next research question **RQ3b**. Having no or only limited control as tested in the deployment scenario (see Section 6.3) and also partly in the first study with the *FI Mode*(see Section 6.2) was disliked and received the most critical feedback. Conversely, the users appreciated to stay in control while having maximum freedom to customize the application. Those delivery modes which granted the participants more autonomy had been explicitly preferred. Hence, this observation can be regarded as one effect of how users perceive the manipulation of stressors. Further, this is interesting from an interaction perspective, since the results suggest that involving the user actively, even partly increases self-reflection about the potential stressors. In contrast, if the user is required to get involved too heavily in the manipulation of the stressor, it can lead to a feeling of overload and consequently even to abandoning the interactive system could be argued for the final study (see Section 6.4). Thus, maintaining a balance between user involvement in the manipulation of stressors is crucial for the success of the interactive system incorporating such techniques to reduce stress, since otherwise it could be reversed in an feeling of overload or disinterest in the system.

These findings are contributing to answering **RQ3b** targeting the effects of stressor manipulation on the user. As indicated by the results from the first study described in Section 6.2, a measurable impact on the subjectively perceived stress level cannot be detected. Although there are some hints in the ESM data showing that while the *sender dependent mode* was activated, the stress level decreased slightly, the standard deviation and the limited number of participants do not allow to draw any generalizable conclusions on the stress-reducing effects of such a mode. In contrast, the second study presented in Section 6.3 provides answers to **RQ3b** revealing that the visualization of a stressor putting the user in the passive role of receiving feedback without needing to take action can be beneficial for becoming aware of one's own body signals substituting the feedback. Although the utility of the prototypical deployment of the system, which visually represented cognitive load on a ambient display, was not necessarily perceived to be useful by the participants, they agreed that for people who have difficulties noticing their physiological state by themselves, such a system could be helpful. Similarly, the findings from the last study requiring the user to take action for visualizing one's level of business lead to the conclusion that the active confrontation with something that stresses the user, supports the ability to become aware and conscious about one's mental state. This important finding further leads to **RQ3c** showing that the visualization of a stressor can facilitate self-reflection in the user. Accordingly, even if the participants would not like to rely on the automatic feedback permanently, they were encouraged to deal with the topic, here the visual representation of potential stressors and how it affects them, just by using the system. Moreover, the latter presented evaluation of the manual wearable display revealed that encouraging the self-reflection can further lead to interesting conversations with others. By wearing the physical prototype publicly and together with co-participants, also the mutual consideration of each others' personal sensitivities had been enabled. Thus, creating an appropriate usage scenario, the visualization of stressors in groups can support the interaction behavior going beyond one's self-reflection.

Since the second study (cf., Section 6.3) has been carried out in a semi-public setting, namely the workplace environment, it allowed to study what implications the non-private presentation of sensitive data on the user has (**RQ3d**). As reported in the interviews, the ambient display evoked interest in the colleagues what was accompanied by misinterpretations that the system could do thought-reading. Although the working atmosphere was relaxed according to the participant, the public visibility of the cognitive state lead to some comments which made her feel uncomfortable, particularly when bystanders from other offices took note of the visual representation. This observation responds to **RQ3d** when suggesting that visualizing one's mental state allows misinterpretations, which addresses the challenge how privacy can be preserved when stress-aware interfaces are designed. From the in-the-wild deployment it can be inferred that users are afraid of their abuse of sensitive data on the one hand, but also of bystanders jumping to conclusions leading to misjudgements of their cognitive state on the other. Correspondingly, it is debatable whether ambient displays visualizing feedback based on physiological data are a suitable technological mean given the potential to violate the user's privacy. Instead, the feedback could be either shown on personal devices not being exposed to bystanders, such as feedback icons in the taskbar on a computer screen. Further, the opportunity to easily switch on and off a device which is exposed to bystanders and could tell privacy sensitive data could be a solution for preserving as much the user's privacy as desired. In general, it is an important finding that all participants agreed that sharing the source of stress, such as the feeling of being busy, is preferred instead of explicitly admitting to be stressed by revealing one's mental state. This indicates that stress is still a highly sensitive topic for individuals, which can be considered as a sign of weakness that one does not want to disclose.

Chapter 7

Conclusion and Outlook

In the following, the conclusions drawn from the performed research activities will be presented. With the aim to explore the effectiveness and the effects of potentially stress mitigating techniques, the foundation provided by prior work on how stress responses can be measured physiologically (**RQ1a**) and the demonstration of the physiological correlates of subjective stress (**RQ1b**) have laid the foundation for the identification of prevalent shortcomings regarding existing sensing technology (**RQ1c**). Further, the *Design Space for Wearable Physiological Measurement Tools* have been presented as an outcome of the identification and summarization process referring to the evaluation of physiological measurement hardware (**RQ1d**). Evaluating the suitability of thermal feedback for notifying users about stress by applying different feedback (**RQ2a**) resulted in the conclusion that tactile feedback does not seem to function as an effective notifier, answering the question how such a stimulus needs to be designed (**RQ2b**). Finally, three research probes consisting of comprehensive studies that exemplarily illustrate by which means stressors can be manipulated (**RQ3a**) have led to the subsequent research contributions and insights tackling the effects of such manipulation techniques (**RQ3b**). Special emphasis was put on the investigation of the users' self-reflection (**RQ3c**) and privacy perception (**RQ3d**) and these aspects had been affected by the elimination, respectively visualization of stressors. The present chapter concludes with an outlook regarding the prospective challenges future work will be targeting to research and the concluding remarks of the presented research.

7.1 Summary of Research Contribution

Through the investigation of different stress reducing approaches and resulting interventions varying in multiple aspects, this work provides three main contributions. The detailed inquiry of two stakeholder groups utilizing physiology-sensing hardware has firstly, led to the *Design Space for Wearable Physiological Measurement Tools* reflecting consumers' and researchers' criteria for choosing such equipment; and secondly, five *Design Recommendations for Wearable Physiology Sensing Devices* targeting current challenges in their practical application have been deduced. With respect to the evaluation of different approaches to reduce stress, the third contribution consists of three minor contributions all responding to the question, what to consider when designing interactive systems to mitigate stress. While the exploration of tactile feedback for notifying users about stress has led to the conclusion that users question the stress-relieving function and rather tend to focus on the detected stress. Hence, in our research we did not further focus on the approach to notify users about stress. However, this result is valuable for future work to be tested in an in-the-wild scenario where it should be taken into account that thermal feedback provided at unobtrusive body locations is preferred. To move on, I explored an alternative intervention method and manipulated the stressors differently. Hereby, it was shown that manipulating stressors affect positively the ability of self-reflect about one's sources of stress. Moreover, dependent on the context of visual presentation of one's stressor and their visibility to bystanders, it could be even seen that the mutual consideration for each other and the social interaction benefits from the visualization of a stress source. The final minor contribution providing insights for the practical implementation of stress-aware systems, is the investigation of privacy preservation. As the results indicate, revealing one's mental state is a highly sensitive topic for individuals, which is why they prefer sharing the source of stress, such as the feeling of being busy instead of explicitly admitting to be stressed. In this context, the preservation of privacy by, for example allowing users to utilize applications as they like or granting as much freedom to customize systems as possible is crucial for the success of such stress-aware technology.

7.1.1 Identifying and Summarizing Criteria for Successful Physiological Measurements

Since the reliable recognition of stress builds the foundation for designing effective interactive systems to mitigate stress, physiological measurement tools are essentially important. Recently, more and more researchers have used such tools to develop and validate affective computing systems [63, 83, 198]. As a result of the increasing demand of such hardware, the variety of consumer and medical standard devices also rose within the past years. Thus, the investigation of considerable criteria when deciding for physiological measurement tools from both, consumers and researchers, perspectives contributes to the state-of-art. The *Design Space for Wearable Physiological Measurement Tools* presented in Chapter 4 is the first of its kind embracing six dimensions and 17 sub-dimensions addressing crucial issues for dealing with such technology and thereby revealing constraints of sensing devices (**RQ1c**). It's value does not exclusively lay in the identification inferred from an extensive corpus of qualitative data, but further in the summarization of the criteria building a clear overview for any stakeholder group responding to **RQ1d**. Consequently, it provides a valuable starting point for further discussions, particularly among professionals using physiology-aware hardware for measuring stress and affective states.

7.1.2 Design Recommendations for Stress-Aware Interactive Systems

For the design of interactive systems developers, designers and user-experience experts alike need to view the human-computer interaction from the end-user perspective. For this, user studies, particularly collecting qualitative data from genuine practitioners' experiences as presented in this thesis, provide an essential source of information and bear valuable insights when wanting to tackle consumer needs and requirements. Accordingly, the *Design Space for Wearable Physiological Measurement Tools* emphasizes the shortcomings in current sensing hardware addressing **RQ1c**, but more importantly provides an overview on critical issues. As part of the extensive research on physiology-aware systems (cf., Chapter 4) including the elaboration of the *Design Recommendations for Wearable Physiology Sensing Devices* (cf., Section 4.4), the following recommendations for designing stress-aware interactive systems have been derived.

Enhance Self-Monitoring Users benefit from self-monitoring applications providing feedback upon, e.g. their stress level. For this concepts, such as the quantified self approach should be supported in future systems.

Respect the Ownership of Private Data Hand in hand with the previous recommendation goes the requirement to provide maximum transparency for the users regarding their data, which applies particularly for sensitive data, such as stress level indicators. This includes not only what and how the data is collected, but also what is being done with it and where it is stored. Although not each user might be interested in all of this information, it is important to respect the ownership of one's data and act accordingly.

Support System Usability With the advancements of interactive systems sensing stress, often their complexity increases too. Thus, the user effort to operate and maintain such systems should be kept low and intuitive.

Ensure the Reliability of System's Outcomes Following the preceding aspect, stress-aware interactive technology needs to be resistant against external confounders affecting their results. Since we know that multiple factors, such as the caffein intake or the physical fitness can influence physiological indicators of stress, there is a special need to develop systems that are able to provide reliable results.

7.1.3 Evaluating Interventions to Reduce Stress

The present work contributes a multidimensional investigation of different approaches aiming to reduce stress. While Chapter 5 concentrates on the exploration of effective stimuli to provide feedback on stress states, in Chapter 6 three concepts of how stressors can be manipulated are validated in the field collecting insightful user feedback. The reasoning for focusing on those constructs which elicit stress in users lays in the conclusion that notifications about stress can lead to opposite effects. In practice, signaling the user that the stress level has risen was seen as critical by participants (see Section 5.3.2), since it could increase one's stress level instead of reducing it. Consequently, the approach to adjust the stressor had been explored extensively in this work. For this, the first study eliminated stressors as a potential opportunity to reduce subjectively felt stress elicited by interruptions from notifications, response pressure or the fear of missing out. Subsequently, the effects of visualizing stressors had been examined taking two different characteristics. Before the self-report on one's perceived level of business has been presented publicly requiring the users take action, the visual representation of one's cognitive load in form of ambient feedback had been evaluated in a semi-public setting.

Referring to **RQ3a**, there exist several possibilities of which this work varies three means to manipulate stressors, namely the utilization of technological means, the deployment scenario, and the variation of user control. In concrete, despite the variation in the applied deployment, namely smartphone, ambient display, and manual display, further the privacy continuum has been exploited ranging from private usage (smartphone) through semi-public (ambient display in a shared office) up to public (necklace worn in an university). For the modification of user control the users' ability to change the stressor manipulation had been varied. For example, the smartphone application had two different modes that delayed notifications for a fixed time interval and one for which the delay could be adjusted.

Suitability of Tactile Feedback In the course of the exploration of feedback for stress states several types of work have been performed involving tactile feedback. Due to its advantages, namely the unobtrusiveness, which also preserves privacy and is less revealing, as well as the unoccupation of this sense by other technology so far **RQ2a**, this modality was used to design two experiments evaluating the effectiveness of different tactile stimuli. Finding an answer to research questions **RQ2b**, initially the effect of presenting vibrotactile feedback in comparison to pressure-based stimulation had been evaluated in a user study. Since the latter was indicated to evoke only a little less stress in the users when being applied, the third common type of tactile feedback, thermal stimulation, was researched in the subsequent study. Hereby, it was found that an effective feedback stimulus is mild cold and has a high rate of change (**RQ2b**). Regarding the body location, where users would like to receive the feedback, it showed to depend on whether users valued the privacy preservation and therefore chose hidden spots, such as the back. Unlike others, who appreciated the familiarity and unobtrusiveness when incorporating thermal feedback in existing smart watches or other wrist-worn wearables. However, the participants' opinions were divided regarding the overall effectiveness of and utility of the approach to notify users about their stress level, which is why I explored the manipulation of the source of stress as another concept.

Effects of Manipulating Stressors One of the essential outcomes of the three performed evaluation studies was that users partly had the feeling that the manipulation of the stressor affected their stress level (cf., Section 6.2) and their perceived feeling of being busy (cf., Section 6.4) positively. Although there could not be shown significant differences for the stress assessments, the fact that users subjectively felt less stressed or calmed down is a notable effect (**RQ3b**). As another interesting finding tackling research question **RQ3c** users reported that they were encouraged to reflect about the manipulated stressors. In practice, the smartphone usage and notification management was viewed more critically throughout the study participation. Further, the feedback presented on the ambient display supported the participants in listening to their own body signals indicating mental overload. Similarly, the adjustment of one's level of business assumed that users became aware of how much they actually felt stressed or rushed. Thus, not only the self-reflection, but moreover the increased awareness of sources of stress made the stressor manipulation a successful stress-mitigating technique. Interestingly, the reflective potential of this approach was shown to exceed and even affected the mutual consideration among colleagues as the lastly performed study could show. In this context, it was revealed that visualizing the stressor stimulated insightful conversation and could even facilitate the understanding of the participants' kids regarding their parents' stress state. For this, the effects of manipulating stressors can be regarded as overall positive ranging from subjectively perceived less stress through increased awareness of stressors up to the facilitation to reflect about oneself and even others.

Respecting Privacy in Stress-aware Systems As has been shown to be particularly relevant for studies involving the public (or semi-public) representation of one's mental state, the preservation of privacy plays a key role in the design of stress-aware interactive systems being addressed by **RQ3d**. Moreover, presenting sensitive private information unobtrusively is important, since it had been revealed that individuals have concerns disclosing their stress state in front of others as an outcome of the research probes focalizing the visualization of stressors. However, thinking of the presentation of tactile feedback and in particular thermal stimuli (cf., Chapter 5), the approach to discreetly notify about one's stress has not been shown to be promising. Whereas exploiting the main advantage of thermal feedback by combining it with the manipulation of stressors could be desirable. Reviewing the findings from Chapter 6, it can be obtained that granting users as much freedom to utilize the technology according to their needs and requirements seems to be successful for causing a feeling of privacy due to the maintained control. Correspondingly, the transparency of what data is being recorded, what is being used for, and where it is being stored must be provided when incorporating physiological data. As the study participants emphasized, the fear of seeing one's data abused is a prevalent thread when dealing with technology. Thus, to build trust and support the users' confidence, interactive systems must provide comprehensive explanations for the aforementioned questions and further let users maintain control, when aiming for successful and effective systems.

7.2 Limitations

Since dealing with physiological data is sensitive due to the various factors which influence the physiological state, the reliability of collected data requires carefulness when interpreting the results. Besides being prone to caffeein intake or physical overall fitness, the signal quality can be easily worsened through movements during the data collection. Although it was attempted to isolate and control such factors throughout the performed studies, there remains a small amount of uncontrollability.

Likewise, assessing self-reported data, for example when prompting users with ESMs underlies subjective feelings and experience. While is partly the strength of such data, it surely yields disadvantages for the of the replicability of the results. While having considered both, quantitative and qualitative data together for drawing conclusions, the presented findings give an impression on what has been found for a particular application tested in a specific scenario. Hence, to make sure the conclusions are generalizable much more extensive research involving more participants and tests in different scenarios using distinct applications or systems would be needed. Nevertheless, this work represents a valuable starting point to deepen further research in the addressed domains.

7.3 Future Work

In the course of the investigation of different stress mitigating approaches based on physiological stress detection it has been revealed that some research domains have so far, remained untouched. Since their exploration goes beyond the scope of this work, they will be illustrated briefly in the following.

Extending Physiological Sensing Capabilities As one of the prerequisites for the successful design of interactive systems aiming to incorporate stress mitigating techniques the detection of stress needs to work reliably. To show which hardware works best for collecting physiological signals, the study presented in Chapter 3 was performed. Although the results showed significant correlations, there were some shortcomings revealed, which were researched extensively in Chapter 4. Facing the difficulties when collecting physiological data under specific circumstances, such as physical activity, there is still a lot of research needed to improve the reliable sensing of stress. As has been highlighted in the *Design Recommendations for Stress-Aware Interactive Systems* (see Section 7.1.2) future attempts from engineers and developers should focus on the enhancement of wearables collecting stress data. Hereby, there should be more emphasis put on the users' needs, as for example increasing the transparency of how physiological data is being recorded and why it might deviate from similar, e.g. step counter information. Instead of promoting high-technology to potential buyers, future work should work out opportunities to facilitate the user's understanding of the complex mechanisms. In this context subsequent research also needs to invest in in-depth user inquiries to gain valuable feedback, as the presented work already builds a starting point.

Testing Stress Mitigating Systems in the Wild Looking into the future of the design of interactive systems aiming to incorporate stress mitigating techniques, there is an enormous potential to be researched. While the present work has focused on the exploration of stress reducing approaches shaped by the revealed limitations and opportunities of physiological sensing, there is still a lack of knowledge in how far stress mitigation can be regarded as an effective mean. Since stress is a highly sensitive topic due to its subjective nature, it is extremely difficult to design experiments validating if the applied approach successfully reduces stress in the user. As could be also seen in the results of the presented studies individuals experience stress differently and sometimes their subjective perception does not correspond to the physiologically indicated stress. For this, interactive systems need to be evaluated extensively using various measures, such as self-assessments, observations and physiological sensing, and further the context must not be artificial. Hence, one of the most pressing targets for subsequent research should be the comprehensive testing of potentially stress mitigating systems in real-world scenarios.

Developing Novel Application Use Cases As had been revealed in the final study, kids and also smaller children had understood the the color mapping of *BuSiNec* immediately. From the participants' statements it becomes obvious that they were not quick in the comprehension of what each color meant, but that they further had less effort to retrieve the meaning of the signaled color. Thus, visualizing stressors, or respectively the stress state could be particularly useful for parents telling their kids in an easily understandable way if they are stressed or relaxed. Hence, future work should focus on the investigation of utilizing intuitive stressor representations in playful context. Hereby the mutual consideration between parents and kids regarding their stress level could be facilitated. As P3 from the *BuSiNec* sample mentioned, she would have loved to test out such a concept at home for visualizing if you are in hurry for bringing the kids to the kindergarten or school.

Since there are also other situations when communication can be difficult, the research gap embracing challenging target groups besides children should be aimed at. For example, the utilization of stress visualization could be beneficial for therapeutic contexts. As a result of the deployment study the visual representation of physiologically indicated stress could be beneficial for those users who have difficulties listening to their own body signals. In particular, patients having personality disorders or suffering from a lack of self-reflection represent potentially promising use cases to be explored in the future.

Thinking of such scenarios, it could be further interesting to examine whether the actual representation of one's mental state is important or if self-awareness could be achieved using fake systems functioning as placebos. Based on the findings from the presented work, subsequent research could test what exactly triggers the self-reflection when dealing with stressors.

7.4 Concluding Remarks

Facing the enormous technological developments society has faced and that will be imminent, the role of stress has grown in importance. While in past times humans had time to adapt to environmental changes over hundreds of years, the prevalent challenge today entails the fast evolvements of the usage of digital technologies, which are perceived more and more as stressful. This continuous feeling of information overload due to the omnipresent usage of ubiquitous technology leads to a number of severe consequences, such as attention or memory deficits. Therefore the investigation of coping strategies and effective interventions to reduce stress in users represents an important strand of work. The present thesis has focused on the exploration of stress-mitigating techniques involving the tactile notification of users about their stress level and further the manipulation of stressors. With our technology built in the research probes, the user's self-reflection for his or her stress level was increased. This provided a starting point for future research incorporating such intervention techniques in interactive systems. For the encounter of such challenges and transferring the gained knowledge into practice, the *Design Recommendations for Stress-Aware Interactive Systems* have been presented.

Bibliography

[1] P. Adams, M. Rabbi, T. Rahman, M. Matthews, A. Voida, G. Gay, T. Choudhury, and S. Voida. Towards personal stress informatics: Comparing minimally invasive techniques for measuring daily stress in the wild. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '14, pages 72–79, ICST, Brussels, Belgium, 2014. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[2] J. Alvina, S. Zhao, S. T. Perrault, M. Azh, T. Roumen, and M. Fjeld. Omnivib: Towards cross-body spatiotemporal vibrotactile notifications for mobile phones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2487–2496, New York, NY, USA, 2015. ACM.

[3] S. Andersson and A. Finset. Heart rate and skin conductance reactivity to brief psychological stress in brain-injured patients. *Journal of Psychosomatic Research*, 44(6):645 – 656, 1998.

[4] J. Auda, D. Weber, A. Voit, and S. Schneegaß. Understanding user preferences towards rule-based notification deferral. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages 1–6, New York, NY, USA, 2018. ACM.

[5] Y. Ayzenberg and R. W. Picard. Feel: A system for frequent event and electrodermal activity labeling. *IEEE Journal of Biomedical and Health Informatics*, 18(1):266–277, Jan 2014.

[6] B. P. Bailey, J. A. Konstan, and J. V. Carlis. The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In *Interact*, volume 1, pages 593–601, 2001.

[7] L. M. Baker and W. M. Taylor. The relationship under stress between changes in skin temperature, electrical skin resistance, and pulse rate. *Journal of experimental psychology*, 48(5):361, 1954.

[8] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In A. Ferscha and F. Mattern, editors, *Pervasive Computing*, pages 1–17, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[9] L. Barkhuus. Context information vs. sensor information: A model for categorizing context in context-aware mobile computing. *SIMULATION SERIES*, 35(1):127–133, 2003.

[10] L. Barkhuus and A. Dey. Is context-aware computing taking control away from the user? three levels of interactivity examined. In A. K. Dey, A. Schmidt, and J. F. McCarthy, editors, *UbiComp 2003: Ubiquitous Computing*, pages 149–156, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[11] A. Barreto and J. Zhai. Physiological instrumentation for real-time monitoring of affective state of computer users. *WSEAS Transactions on Circuits and Systems*, 3(3):496–501, 2003.

[12] A. Barreto, J. Zhai, and M. Adjouadi. Non-intrusive physiological monitoring for automated stress detection in human-computer interaction. *Human–Computer Interaction*, 4796:29–38, 2007.

[13] A. Baum, J. E. Singer, and C. S. Baum. Stress and the environment. *Journal of Social Issues*, 37(1):4–35, 1981.

[14] D. Bawden and L. Robinson. The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, 35(2):180–191, 2009.

[15] M. Becker, A. Kolbeck, C. Matt, and T. Hess. Understanding the continuous use of fitness trackers: A thematic analysis. In *PACIS*, pages 40–53, 2017.

[16] G. Berg-Beckhoff, G. Nielsen, and E. Ladekjær Larsen. Use of information communication technology and stress, burnout, and mental health in older, middle-aged, and younger workers–results from a systematic review. *International Journal of Occupational and Environmental Health*, 23(2):160–171, 2017.

[17] F. Bianchi, S. J. Redmond, M. R. Narayanan, S. Cerutti, and N. H. Lovell. Barometric pressure and triaxial accelerometry-based falls event detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 18(6):619–627, Dec 2010.

[18] I. Biederman. Mental set and mental arithmetic. *Memory & Cognition*, 1(3):383–386, 1973.

[19] A. Blandford, D. Furniss, and S. Makri. Qualitative hci research: Going behind the scenes. *Synthesis Lectures on Human-Centered Informatics*, 9(1):1–115, 2016.

[20] A. Blythe, S. E. Crouter, and S. R. LaMunion. Validity of consumer-based physical activity monitors for estimating energy expenditure in youth. Master's thesis, University of Tennessee, Knoxville, USA, 2017.

[21] K. Bodine and F. Gemperle. Effects of functionality on perceived comfort of wearables. *Proceedings of the 7th IEEE International Symposium on Wearable Computers*, pages 57–61, 2003.

[22] G. Böhme. *Invasive technification: Critical essays in the philosophy of technology*. Bloomsbury Publishing, 2012.

[23] L. E. Bourne Jr and R. A. Yaroush. Stress and cognition: A cognitive psychological perspective. `https://ntrs.nasa.gov/search.jsp?R=20040034070`, September 2003. Retrieved August 2, 2019.

[24] C. V. C. Bouten, K. T. M. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Biomedical Engineering*, 44(3):136–147, March 1997.

[25] R. E. Boyatzis. Transforming qualitative information: thematic analysis and code development, 1998.

[26] M. M. Bradley and P. J. Lang. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal Of Behavior Therapy And Experimental Psychiatry*, 25(1):49–59, Mar 1994.

[27] N. L. Bragazzi and G. Del Puente. A proposal for including nomophobia in the new dsm-v. *Psychology Research and Behavior Management*, 7:155–160, 2014.

[28] V. Braun and V. Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.

[29] S. Brewster and L. M. Brown. Tactons: structured tactile messages for non-visual information display. In *Proceedings of the fifth conference on Australasian user interface - Volume 28*, pages 15–23. Australian Computer Society, Inc., 2004.

[30] M. Bruns Alonso, M. Varkevisser, and D. V. Keyson. Expressive stress relievers. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, NordiCHI '12, pages 761–764, New York, NY, USA, 2012. ACM.

[31] J. Burke. *Health analytics: gaining the insights to transform health care*. John Wiley & Sons, 2013.

[32] C. Caldeira and Y. Chen. *Seniors and Self-tracking Technology*, pages 67–79. Springer International Publishing, Cham, 2019.

[33] A. I. Canhoto and S. Arp. Exploring the factors that support adoption and sustained use of health and fitness wearables. *Journal of Marketing Management*, 33(1-2):32–60, 2017.

[34] W. B. Cannon. *The Wisdom of the Body*. Norton & Co., Oxford, England, 1939.

[35] J. Casselman, N. Onopa, and L. Khansa. Wearable healthcare: Lessons from the past and a peek into the future. *Telematics and Informatics*, 34(7):1011 – 1023, 2017.

[36] M. A. Cavanaugh, W. R. Boswell, M. V. Roehling, and J. W. Boudreau. An empirical examination of self-reported work stress among us managers. *Journal of applied psychology*, 85(1):65, 2000.

[37] CCS Insight. Ccs insight forecast predicts apple watch and hearables to fuel growth in wearables. `https://www.ccsinsight.com/press/company-news/3375-ccs-insight-forecast-predicts-apple-watch-and-hearables-to-fuel-growth-in-wearables/`, February 2018. Retrieved June 11, 2019.

[38] G. Chen and D. Kotz. A survey of context-aware mobile computing research. *Dartmouth Computer Science Technical Report TR2000-381*, 2000.

[39] E. Chiauzzi, C. Rodarte, and P. DasMahapatra. Patient-centered activity monitoring in the self-management of chronic health conditions. *BMC Medicine*, 13(1):77, Apr 2015.

[40] E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 1143–1152, New York, NY, USA, 2014. ACM.

[41] G. P. Chrousos and P. W. Gold. The Concepts of Stress and Stress System Disorders: Overview of Physical and Behavioral Homeostasis. *JAMA*, 267(9):1244–1252, March 1992.

[42] G. P. Chrousos, D. L. Loriaux, and P. W. Gold. Mechanisms of physical and emotional stress. In *Advances in Experimental Medicine and Biology*, volume 245. Springer US, 1988.

[43] S. H.-W. Chuah, P. A. Rauschnabel, N. Krey, B. Nguyen, T. Ramayah, and S. Lade. Wearable technologies: The role of usefulness and visibility in smartwatch adoption. *Computers in Human Behavior*, 65:276 – 284, 2016.

[44] K. Church and R. de Oliveira. What's up with whatsapp?: comparing mobile instant messaging behaviors with traditional sms. In *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '13, pages 352–361. ACM, 2013.

[45] P. A. Clark. Healthcare marketing in the 21st century. *Health Marketing Quarterly*, 23(3):1–7, 2007.

[46] S. Cohen, T. Kamarck, and R. Mermelstein. A global measure of perceived stress. *Journal of Health and Social Behavior*, 24(4):385–396, 1983.

[47] S. Consolvo and M. Walker. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing*, 2(2):24–31, April 2003.

[48] M. Consult. How likey are you to purchase wearables such as fitbit or smartwatches?, 2018. Retrieved June 11, 2019.

[49] A. D. J. Cooke and P. P. Zubcsek. The connected consumer: Connected devices and the evolution of customer intelligence. *Journal of the Association for Consumer Research*, 2(2):164–178, 2017.

[50] L. Coorevits and T. Coenen. The rise and fall of wearable fitness trackers. In *Academy of Management*, 2016.

[51] A. Crabtree, A. Chamberlain, R. E. Grinter, M. Jones, T. Rodden, and Y. Rogers. Introduction to the special issue of "the turn to the wild". *ACM Trans. Comput.-Hum. Interact.*, 20(3), July 2013.

[52] F. Culwin and X. Faulkner. When fingers do the talking: a study of text messaging. *Interacting with Computers*, 17(2):167–185, 12 2004.

[53] R. J. Davidson, K. R. Scherer, and H. H. Goldsmith. *Handbook of Affective Sciences*. Oxford University Press, 2003.

[54] M. E. Dawson, A. M. Schell, and D. L. Filion. The electrodermal system. *Handbook of Psychophysiology*, 2:200–223, 2007.

[55] Z. Dharmawan. Analysis of computer games player stress level using eeg data. Master's thesis, Delft University of Technology, Netherlands, 2007.

[56] D. M. Diamond, A. M. Campbell, C. R. Park, J. Halonen, and P. R. Zoladz. The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the yerkes-dodson law. *Neural Plasticity*, 2007, 2007.

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

[58] G. M. Dominick, K. N. Winfree, R. T. Pohlig, and M. A. Papas. Physical activity assessment between consumer- and research-grade accelerometers: A comparative study in free-living conditions. *JMIR mHealth and uHealth*, 4(3), 2016.

[59] J. T. Durlak. A typology for interactive med. *Annals of the International Communication Association*, 10(1):743–757, 1987.

[60] F. El-Amrawy and M. I. Nounou. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? *Healthcare Informatics Research*, 21(4):315, 2015.

[61] J. D. Elhai, J. C. Levine, R. D. Dvorak, and B. J. Hall. Fear of missing out, need for touch, anxiety and depression are related to problematic smartphone use. *Computers in Human Behavior*, 63:509 – 516, 2016.

[62] K. R. Evenson, M. M. Goto, and R. D. Furberg. Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1):159, Dec 2015.

[63] A. Exler, A. Schankin, C. Klebsattel, and M. Beigl. A wearable system for mood assessment considering smartphone features and data from mobile ecgs. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, pages 1153–1161, New York, NY, USA, 2016. ACM.

[64] C. Fan, J. Forlizzi, and A. K. Dey. A spark of activity: Exploring informative art as visualization for physical activity. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 81–84, New York, NY, USA, 2012. ACM.

[65] G. Farnell and J. Barkley. The effect of a wearable physical activity monitor (fitbit one) on physical activity behaviour in women: A pilot study. *Journal of Human Sport and Exercise*, 12(4):1230–1237, 2017.

[66] S. Fedor, P. Chau, N. Bruno, R. W. Picard, J. Camprodon, and T. Hale. Can we predict depression from the asymmetry of electrodermal activity. *Journal of Medical Internet Research*, 18(12):1–1, Dec 2016.

[67] S. Feng, G. Wilson, A. Ng, and S. Brewster. Investigating pressure-based interactions with mobile phones while walking and encumbered. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '15, pages 854–861, New York, NY, USA, 2015. ACM.

[68] T. Ferguson, A. V. Rowlands, T. Olds, and C. Maher. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *International Journal of Behavioral Nutrition and Physical Activity*, 12(1):42, Mar 2015.

[69] P. Ferreira, P. Sanches, K. Höök, and T. Jaensson. License to chill!: How to empower users to cope with stress. In *Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges*, NordiCHI '08, pages 123–132, New York, NY, USA, 2008. ACM.

[70] K. Finstad. The usability metric for user experience. *Interacting with Computers*, 22(5):323–327, 2010.

[71] A. J. Fisher and M. G. Newman. Heart rate and autonomic response to stress after experimental induction of worry versus relaxation in healthy, high-worry, and generalized anxiety disorder individuals. *Biological Psychology*, 93(1):65–74, 2013.

[72] R. R. Fletcher, M. Z. Poh, and H. Eydgahi. Wearable sensors: Opportunities and challenges for low-cost health care. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 1763–1766, Aug 2010.

[73] T. A. Floegel, A. Florez-Pregonero, E. B. Hekler, and M. P. Buman. Validation of consumer-based hip and wrist activity monitors in older adults with varied ambulatory abilities. *The Journals of Gerontology: Series A*, 72(2):229–236, 2017.

[74] B. J. Fogg. Persuasive technologies: Introduction. *Commun. ACM*, 42(5):26–29, May 1999.

[75] I. O. for Standardization. *ISO 9241-210:2010 – Ergonomics of human-system interaction: Part 210: Human-centred design for interactive systems*. International Organization for Standardization (ISO), 2010.

[76] J. Frey and J. R. Cauchard. Remote biofeedback sharing, opportunities and challenges. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, UbiComp '18, pages 730–733, New York, NY, USA, 2018. ACM.

[77] T. Fritz, E. M. Huang, G. C. Murphy, and T. Zimmermann. Persuasive technology in the real world: A study of long-term use of activity sensing devices for fitness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 487–496, New York, NY, USA, 2014. ACM.

[78] Gallup. Gallup 2019 Global Emotions Report. `https://www.gallup.com/analytics/248909/gallup-2019-global-emotions-report-pdf.aspx`, 2019. Retrieved August 6, 2019.

[79] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti. Empatica e3 – a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *2014 EAI 4th International Conference on Wireless Mobile Communication and Healthcare*, MobiHealth 2014, pages 39–42. IEEE, 2014.

[80] F. C. M. Geisler, N. Vennewald, T. Kubiak, and H. Weber. The impact of heart rate variability on subjective well-being is mediated by emotion regulation. *Personality and Individual Differences*, 49(7):723–728, Nov 2010.

[81] P. J. Gianaros, I. C. Onyewuenyi, L. K. Sheu, I. C. Christie, and H. D. Critchley. Brain systems for baroreflex suppression during stress in humans. *Human Brain Mapping*, 33(7):1700–1716, 2012.

[82] S. Gillinov, M. Etiwy, R. Wang, G. Blackburn, D. Phelan, A. M. Gillinov, P. Houghtaling, H. Javadikasgari, and M. Y. Desai. Variable accuracy of wearable heart rate monitors during aerobic exercise. *Medicine & Science in Sports & Exercise*, 49(8):1697–1703, Aug 2017.

[83] P. A. Gloor. Consistent excitement correlates with happiness - predicting mood through body sensing with smartwatches. *ICKN White Paper*, January 2017. `http://www.ickn.org/documents/Mood_Tracking_Smartwatch.pdf`.

[84] L. Goldman. Using customer intelligence to support a customer-focused strategy. *Information Management*, 14(3):62, 2004.

[85] D. Gooch. *An Investigation into Communicating Social Presence With Thermal Devices*. PhD thesis, MSc Dissertation, 2009.

[86] A. Gorlick. Media multitaskers pay mental price, stanford study shows. *New Scientist*, 5:11, 2009.

[87] N. Gorm. Personal health tracking technologies in practice. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '17 Companion, pages 69–72, New York, NY, USA, 2017. ACM.

[88] N. Gorm and I. Shklovski. Sharing steps in the workplace: Changing privacy concerns over time. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4315–4319, New York, NY, USA, 2016. ACM.

[89] J. G. Greene, R. Smith, M. Gardiner, and G. C. Timbury. Measuring behavioural disturbance of elderly demented patients in the community and its effects on relatives: A factor analytic study. *Age and Ageing*, 11(2):121–126, 1982.

[90] S. Greene, H. Thapliyal, and A. Caban-Holt. A survey of affective computing for stress detection: Evaluating technologies in stress detection for better health. *IEEE Consumer Electronics Magazine*, 5(4):44–56, Oct 2016.

[91] R. Grinter and M. Eldridge. Wan2tlk?: Everyday text messaging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 441–448, New York, NY, USA, 2003. ACM.

[92] M. Haller, C. Richter, P. Brandl, S. Gross, G. Schossleitner, A. Schrempf, H. Nii, M. Sugimoto, and M. Inami. Finding the right way for interrupting people improving their sitting posture. *Human-Computer Interaction–INTERACT 2011*, pages 1–17, 2011.

[93] M. Halvey, G. Wilson, S. Brewster, and S. Hughes. "baby it's cold outside": The influence of ambient temperature and humidity on thermal feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 715–724, New York, NY, USA, 2012. ACM.

[94] M. Halvey, G. Wilson, Y. Vazquez-Alvarez, S. A. Brewster, and S. A. Hughes. The effect of clothing on thermal feedback perception. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, ICMI '11, pages 217–220, New York, NY, USA, 2011. ACM.

[95] C. Hamilton-Craig, A. Fifoot, M. Hansen, M. Pincus, J. Chan, D. L. Walters, and K. R. Branch. Diagnostic performance and cost of ct angiography versus stress ecg – a randomized prospective study of suspected acute coronary syndrome chest pain in the emergency department (ct-compare). *International Journal of Cardiology*, 177(3):867 – 873, 2014.

[96] L. Han, Q. Zhang, X. Chen, Q. Zhan, T. Yang, and Z. Zhao. Detecting work-related stress with a wearable device. *Computers in Industry*, 90:42 – 49, 2017.

[97] P. A. Hancock and P. A. Desmond. *Stress, workload, and fatigue*. CRC Press, Boca Raton, 2000.

[98] K. Hänsel, R. Poguntke, H. Haddadi, A. Alomainy, and A. Schmidt. What to put on the user: Sensing technologies for studies and physiology aware systems. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 145:1–145:14, New York, NY, USA, 2018. ACM.

[99] K. Hänsel, N. Wilde, H. Haddadi, and A. Alomainy. Challenges with current wearable technology in monitoring health data and providing positive behavioural support. In *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare*, MobiHealth 2015, pages 158–161, ICST, Brussels, Belgium, Belgium, 2015. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[100] Y. Hao, J. Budd, M. M. Jackson, M. Sati, and S. Soni. A visual feedback design based on a brain-computer interface to assist users regulate their emotional state. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 2491–2496, New York, NY, USA, 2014. ACM.

[101] Y. Hao and R. Foster. Wireless body sensor networks for health-monitoring applications. *Physiological Measurement*, 29(11):R27, 2008.

[102] M. A. Harrison and A. L. Gilmore. U txt when? college students' social contexts of text messaging. *The Social Science Journal*, 49(4):513 – 518, 2012. Special Issue: National and state politics: A current assessment.

[103] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139 – 183. North-Holland, 1988.

[104] S. S. Hassellund, A. Flaa, L. Sandvik, S. E. Kjeldsen, and M. Rostrup. Long-term stability of cardiovascular and catecholamine responses to stress tests an 18-year follow-up study. *Hypertension*, 55(1):131–136, 2010.

[105] M. Hassib, D. Buschek, P. W. Wozniak, and F. Alt. Heartchat: Heart rate augmented mobile chat to support empathy and awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 2239–2251, New York, NY, USA, 2017. ACM.

[106] M. Hassib, M. Khamis, S. Schneegaß, A. S. Shirazi, and F. Alt. Investigating user needs for bio-sensing and affective wearables. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 1415–1422, New York, NY, USA, 2016. ACM.

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

[108] A. Hedge, S. Morimoto, and D. Mccrobie. Effects of keyboard tray geometry on upper body posture and comfort. *Ergonomics*, 42(10):1333–1349, 1999.

[109] J. M. Hektner, J. A. Schmidt, and M. Csikszentmihalyi. *Experience sampling method: Measuring the quality of everyday life*. Sage, 2007.

[110] W. S. Helton. Validation of a short stress state questionnaire. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 48, pages 1238–1242. SAGE Publications, 2004.

[111] W. S. Helton and K. Näswall. Short stress state questionnaire. *European Journal of Psychological Assessment*, 31, 2015.

[112] P. Hemp. Death by information overload. *Harvard business review*, 87(9):82–9, 121, September 2009.

[113] P. Henriquez, B. J. Matuszewski, Y. Andreu-Cabedo, L. Bastiani, S. Colantonio, G. Coppini, M. D'Acunto, R. Favilla, D. Germanese, D. Giorgi, P. Marraccini, M. Martinelli, M. Morales, M. A. Pascali, M. Righi, O. Salvetti, M. Larsson, T. Strömberg, L. Randeberg, A. Bjorgan, G. Giannakakis, M. Pediaditis, F. Chiarugi, E. Christinaki, K. Marias, and M. Tsiknakis. Mirror mirror on the wall... an unobtrusive intelligent multisensory mirror for well-being status self-assessment and visualization. *IEEE Transactions on Multimedia*, 19(7):1467–1481, July 2017.

[114] J. P. Herman and W. E. Cullinan. Neurocircuitry of stress: central control of the hypothalamo-pituitary-adrenocortical axis. *Trends in Neurosciences*, 20(2):78 – 84, 1997.

[115] J. Hernandez, D. McDuff, C. Infante, P. Maes, K. Quigley, and R. Picard. Wearable esm: Differences in the experience sampling method across wearable devices. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '16, pages 195–205, New York, NY, USA, 2016. ACM.

[116] J. Hernandez, R. R. Morris, and R. W. Picard. Call center stress recognition with person-specific models. In *International Conference on Affective Computing and Intelligent Interaction*, pages 125–134. Springer, 2011.

[117] L. M. Hirshfield, R. Gulotta, S. Hirshfield, S. Hincks, M. Russell, R. Ward, T. Williams, and R. Jacob. This is your brain on interfaces: Enhancing usability testing with functional near-infrared spectroscopy. In *Proceedings*

of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, pages 373–382, New York, NY, USA, 2011. ACM.

[118] M. Honal and T. Schlutz. Identifying user state electroencephalographic data. In *Proceedings of the International Conference on Multimodal Input*, ICMI, Trento, Italy, 2005. IEEE Computer Society Conference.

[119] J.-H. Hong, J. Ramos, and A. K. Dey. Understanding physiological responses to stressors during physical activity. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 270–279, New York, NY, USA, 2012. ACM.

[120] K. Hovsepian, M. al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar. cstress: Towards a gold standard for continuous stress assessment in the mobile environment. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 493–504, New York, NY, USA, 2015. ACM.

[121] E. Huhtamo. From cybernation to interaction: a contribution to an archaeology of interactivity. *The digital dialectic: New essays on new media*, (s 96):110, 1999.

[122] S. Hwang, J. Seo, H. Jebelli, and S. Lee. Feasibility analysis of heart rate monitoring of construction workers using a photoplethysmography (ppg) sensor embedded in a wristband-type activity tracker. *Automation in Construction*, 71:372 – 381, 2016.

[123] G. Ilie and W. F. Thompson. A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception: An Interdisciplinary Journal*, 23(4):319–330, Apr 2006.

[124] H. Jagos, J. Oberzaucher, M. Reichel, W. L. Zagler, and W. Hlauschek. A multimodal approach for insole motion measurement and analysis. *Procedia Engineering*, 2(2):3103 – 3108, 2010. The Engineering of Sport 8 - Engineering Emotion.

[125] S. M. Jex. *Stress and job performance: Theory, research, and implications for managerial practice.* Sage Publications Ltd, 1998.

[126] L. A. Jones and M. Berris. The psychophysics of temperature perception and thermal-interface design. In *Proceedings 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems. HAPTICS 2002*, pages 137–142, March 2002.

[127] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *IEEE Transactions on Information Technology in Biomedicine*, 10(1):156–167, Jan 2006.

[128] P. Karthikeyan, M. Murugappan, and S. Yaacob. A review on stress inducement stimuli for assessing human stress using physiological signals. In *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*, pages 420–425, March 2011.

[129] H. Kataoka, H. Kano, H. Yoshida, A. Saijo, M. Yasuda, and M. Osumi. Development of a skin temperature measuring system for non-contact stress evaluation. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 2, pages 940–943. IEEE, 1998.

[130] N. Kelly and S. Gilbert. The wear scale: Developing a measure of the social acceptability of a wearable device. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 2864–2871, New York, NY, USA, 2016. ACM.

[131] R. Kettner, P. Bader, T. Kosch, S. Schneegaß, and A. Schmidt. Towards pressure-based feedback for non-stressful tactile notifications. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, pages 89:1–89:8, New York, NY, USA, 2017. ACM.

[132] R. Khusainov, D. Azzi, I. E. Achumba, and S. D. Bersch. Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations. *Sensors*, 13(10):12852–12902, 2013.

[133] D.-j. Kim, Y. Lee, S. Rho, and Y.-k. Lim. Design opportunities in three stages of relationship development between users and self-tracking devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 699–703, New York, NY, USA, 2016. ACM.

[134] P. Kindness, C. Mellish, and J. Masthoff. Identifying and measuring stressors present in pre-hospital care. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, PervasiveHealth '13, pages 270–273, ICST, Brussels, Belgium,

Belgium, 2013. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[135] A. L. S. King, A. C. Valença, Alexandre M. Silva, F. Sancassiani, S. Machado, and A. E. Nardi. "nomophobia": impact of cell phone use interfering with symptoms and emotions of individuals with panic disorder compared with a control group. *Clin Pract Epidemiol Ment Health*, 10:28–35, 2014.

[136] A. L. S. King, A. M. Valença, and A. E. Nardi. Nomophobia: The mobile phone in panic disorder with agoraphobia: Reducing phobias or worsening of dependence? *Cognitive and Behavioral Neurology*, 23:52–54, 2010.

[137] R. Kirkham, S. Mellor, D. Green, J.-S. Lin, K. Ladha, C. Ladha, D. Jackson, P. Olivier, P. Wright, and T. Ploetz. The break-time barometer: An exploratory system forworkplace break-time social awareness. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '13, pages 73–82, New York, NY, USA, 2013. ACM.

[138] C. Kirschbaum and D. H. Hellhammer. Salivary cortisol in psychoneuroendocrine research: Recent developments and applications. *Psychoneuroendocrinology*, 19(4):313 – 333, 1994.

[139] F. Kiss, R. Poguntke, A. Schmidt, and P. W. Woźniak. S5: Selective sensing of single sound sources. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '18, pages 39–46, New York, NY, USA, 2018. ACM.

[140] J. Klamet, D. J. C. Matthies, and M. Minge. Wearelaxable: A wearable system to enhance stress resistance using various kinds of feedback stimuli. In *Proceedings of the 3rd International Workshop on Sensor-based Activity Recognition and Interaction*, iWOAR '16, pages 2:1–2:6, New York, NY, USA, 2016. ACM.

[141] P. A. J. Kolarsick, M. A. Kolarsick, and C. Goodwin. Anatomy and physiology of the skin. *Journal of the Dermatology Nurses' Association*, 3(4):203–2013, 2011.

[142] A. Kononova, L. Li, K. Kamp, M. Bowen, R. Rikard, S. Cotten, and W. Peng. The use of wearable activity trackers among older adults: Focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. *MIR Mhealth Uhealth*, 7(4), 2019.

[143] T. J. M. Kooiman, M. L. Dontje, S. R. Sprenger, W. P. Krijnen, C. P. van der Schans, and M. de Groot. Reliability and validity of ten consumer activity trackers. *BMC Sports Science, Medicine and Rehabilitation*, 7(1):24, Oct 2015.

[144] T. Kosch, R. Kettner, M. Funk, and A. Schmidt. Comparing tactile, auditory, and visual assembly error-feedback for workers with cognitive impairments. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '16, pages 53–60, New York, NY, USA, 2016. ACM.

[145] H. Koskimäki, H. Mönttinen, P. Siirtola, H.-L. Huttunen, R. Halonen, and J. Röning. Early detection of migraine attacks based on wearable sensors: Experiences of data collection using empatica e4. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, pages 506–511, New York, NY, USA, 2017. ACM.

[146] R. Kraut, M. Patterson, V. Lundmark, S. Kiesler, T. Mukophadhyay, and W. Scherlis. Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist*, 53(9):1017–1031, 1998.

[147] S. Kye, J. Moon, J. Lee, I. Choi, D. Cheon, and K. Lee. Multimodal data collection framework for mental stress monitoring. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, pages 822–829, New York, NY, USA, 2017. ACM.

[148] E. Labbé, N. Schmidt, J. Babin, and M. Pharr. Coping with stress: The effectiveness of different types of music. *Applied Psychophysiology and Biofeedback*, 32(3):163–168, Dec 2007.

[149] T. Lappalainen, L. Virtanen, and J. Häkkilä. Experiences with wellness ring and bracelet form factor. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*, MUM '16, pages 351–353, New York, NY, USA, 2016. ACM.

[150] A. Lazar, C. Koehler, J. Tanenbaum, and D. H. Nguyen. Why we use and abandon smart devices. In *Proceedings of the 2015 ACM International*

Joint Conference on Pervasive and Ubiquitous Computing, UbiComp '15, pages 635–646, New York, NY, USA, 2015. ACM.

[151] R. S. Lazarus. *Emotion and adaptation*. Oxford University Press on Demand, New York, NY, USA, 1991.

[152] R. S. Lazarus and S. Folkman. *Stress, appraisal, and coping*. Springer publishing company, 1984.

[153] D. Ledger and D. McCaffrey. Inside wearables: How the science of human behavior change offers the secret to long-term engagement. https://endeavour.partners/white-papers/, January 2014. Retrieved June 6, 2019.

[154] A. R. Lee, S.-M. Son, and K. K. Kim. Information and communication technology overload and social networking service fatigue: A stress perspective. *Computers in Human Behavior*, 55:51 – 61, 2016.

[155] K. Lee and H. Hong. Mindnavigator: Exploring the stress and self-interventions for mental wellness. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 572:1–572:14, New York, NY, USA, 2018. ACM.

[156] M.-H. Lee, S. Cha, and T.-J. Nam. Patina engraver: Visualizing activity logs as patina in fashionable trackers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 1173–1182, New York, NY, USA, 2015. ACM.

[157] W. Lee and Y.-k. Lim. Thermo-message: Exploring the potential of heat as a modality of peripheral expression. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 4231–4236, New York, NY, USA, 2010. ACM.

[158] J. A. Lepine, N. P. Podsakoff, and M. A. Lepine. A meta-analytic test of the challenge stressor-hindrance stressor framework: An explanation for inconsistent relationships among stressors and performance. *Academy of Management Journal*, 48(5):764–775, 2005.

[159] W. Liao, W. Zhang, Z. Zhu, and Q. Ji. A real-time human stress monitoring system using dynamic bayesian network. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pages 70–70. IEEE, 2005.

[160] W. Linden. What do arithmetic stress tests measure? protocol variations and cardiovascular responses. *Psychophysiology*, 28(1):91–102, 1991.

[161] C. D. B. Luft, E. Takase, and D. Darby. Heart rate variability and cognitive function: Effects of physical effort. *Biological Psychology*, 82(2):186 – 191, 2009.

[162] U. Lundberg, R. Kadefors, B. Melin, G. Palmerud, P. Hassmén, M. Engström, and I. Dohns. Psychophysiological stress and emg activity of the trapezius muscle. *International Journal of Behavioral Medicine*, 1(4):354–370, 1994.

[163] T. Lundeberg, R. Nordemar, and D. Ottoson. Pain alleviation by vibratory stimulation. *Pain*, 20(1):25–44, 1984.

[164] A. Luque-Casado, J. C. Perales, D. Cárdenas, and D. Sanabria. Heart rate variability and cognitive processing: The autonomic response to task demands. *Biological Psychology*, 113:83 – 90, 2016.

[165] A. Luque-Casado, M. Zabala, E. Morales, M. Mateo-March, and D. Sanabria. Cognitive performance and heart rate variability: The influence of fitness level. *PloS one*, 8(2):1–9, 02 2013.

[166] C. Ma, A. Osherenko, H. Prendinger, and M. Ishizuka. A chat system based on emotion estimation from text and embodied conversational messengers. In *Proceedings of the 2005 International Conference on Active Media Technology, 2005*, AMT 2005, pages 546–548, May 2005.

[167] C. Magnusson, H. Caltenco, S. Finocchietti, G. Cappagli, G. Wilson, and M. Gori. What do you like? early design explorations of sound and haptic preferences. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '15, pages 766–773, New York, NY, USA, 2015. ACM.

[168] A. Marakhimov and J. Joo. Consumer adaptation and infusion of wearable devices for healthcare. *Computers in Human Behavior*, 76:135–148, July 2017.

[169] A. Marcus and W. Wang. *Design, User Experience, and Usability: Understanding Users and Contexts: 6th International Conference, DUXU 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings*, volume 10290. Springer, 2017.

[170] E. Márquez Segura, J. Fey, E. Dagan, S. N. Jhaveri, J. Pettitt, M. Flores, and K. Isbister. Designing future social wearables with live action role play (larp) designers. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 462:1–462:14, New York, NY, USA, 2018. ACM.

[171] B. Marte, J. Finkelstein, and L. Anson. Skin biology. *Nature*, 445(7130):833–833, 2007.

[172] G. Matthews, L. Joyner, K. Gilliland, S. Campbell, S. Falconer, and J. Huggins. Validation of a Comprehensive Stress State Questionnaire: Towards a State 'Big Three'. *Personality Psychology in Europe*, 7:335–350, 1999.

[173] M. Matthews, J. Snyder, L. Reynolds, J. T. Chien, A. Shih, J. W. Lee, and G. Gay. Real-time representation versus response elicitation in biosensor data. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 605–608, New York, NY, USA, 2015. ACM.

[174] S. Mayya, V. Jilla, V. N. Tiwari, M. M. Nayak, and R. Narayanan. Continuous monitoring of stress on smartphone using heart rate variability. In *2015 IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 1–5, Nov 2015.

[175] D. J. McDuff, J. Hernandez, S. Gontarek, and R. W. Picard. Cogcam: Contact-free measurement of cognitive stress during computer tasks with a digital camera. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4000–4004, New York, NY, USA. ACM.

[176] K. A. McGonagle and R. C. Kessler. Chronic stress, acute stress, and depressive symptoms. *American Journal of Community Psychology*, 18(5):681–706, Oct 1990.

[177] S. J. McMillan and J.-S. Hwang. Measures of perceived interactivity: An exploration of the role of direction of communication, user control, and time in shaping perceptions of interactivity. *Journal of Advertising*, 31(3):29–42, 2002.

[178] A. Mehrotra, F. Tsapeli, R. Hendley, and M. Musolesi. Mytraces: Investigating correlation and causation between users’ emotional states and mobile phone interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(3):83:1–83:21, Sept. 2017.

[179] S. Misra, L. Cheng, J. Genevie, and M. Yuan. The iphone effect: The quality of in-person social interactions in the presence of mobile devices. *Environment and Behavior*, 48(2):275–298, 2016.

[180] A. V. Moere. Towards designing persuasive ambient visualization. In *Issues in the Design & Evaluation of Ambient Information Systems Workshop*, pages 48–52. Citeseer, 2007.

[181] C. M. Morin, S. Rodrigue, and H. Ivers. Role of stress, arousal, and coping skills in primary insomnia. *Psychosomatic Medicine*, 65:259 – 267, 2003.

[182] M. Najström and B. Jansson. Skin conductance responses as predictor of emotional responses to stressful life events. *Behaviour Research and Therapy*, 45(10):2456 – 2463, 2007.

[183] K. Nakagaki, S. Follmer, and H. Ishii. Lineform: Actuated curve interfaces for display, interaction, and constraint. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST '15, pages 333–339, New York, NY, USA, 2015. ACM.

[184] Y. H. Nam, Y. Bai, J. A. Lee, Y. Kim, J.-M. Lee, N. F. Meier, and G. J. Welk. Validity of consumer-based physical activity monitors in semi-structured activities. *Medicine & Science in Sports & Exercise*, 47(5S):260–261, May 2015.

[185] S. Nasir and Y. Yurder. Consumers' and physicians' perceptions about high tech wearable health products. *Procedia-Social and Behavioral Sciences*, 195:1261 – 1267, 2015.

[186] M. B. Nelson, L. A. Kaminsky, D. C. Dickin, and A. H. K. Montoye. Validity of consumer-based physical activity monitors for specific activity types. *Medicine & Science in Sports & Exercise*, 48(8):1619–1628, 2016.

[187] J. Nielsen. Heuristics for user interface design. `https://www.nngroup.com/articles/ten-usability-heuristics/`, April 1994. Retrieved August 20, 2019.

[188] J. Niess and P. W. Woźniak. Supporting meaningful personal fitness: The tracker goal evolution model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 171:1–171:12, New York, NY, USA, 2018. ACM.

[189] D. Norman. *The Design of Everyday Things*. Basic Books, 2013. revised and expanded edition.

[190] U. Oberst, E. Wegmann, B. Stodt, M. Brand, and A. Chamarro. Negative consequences from heavy social networking in adolescents: The mediating role of fear of missing out. *Journal of Adolescence*, 55:51 – 60, 2017.

[191] E. Ophir, C. Nass, and A. D. Wagner. Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences*, 106(37):15583–15587, 2009.

[192] D. Ottoson, A. Ekblom, and P. Hansson. Vibratory stimulation for the relief of pain of dental origin. *Pain*, 10(1):37–45, 1981.

[193] H. S. Packer, G. Buzogany, D. A. Smith, L. Dragan, M. Van Kleek, and N. R. Shadbolt. The editable self: A workbench for personal activity data. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 2185–2190, New York, NY, USA, 2014. ACM.

[194] P. Paredes and M. Chan. Calmmenow: Exploratory research and design of stress mitigating mobile interventions. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 1699–1704, New York, NY, USA, 2011. ACM.

[195] S. M. Patterson, D. S. Krantz, L. C. Montgomery, P. A. Deuster, S. M. Hedges, and L. E. Nebel. Automated physical activity monitoring: Validation and comparison with physiological and self-report measures. *Psychophysiology*, 30(3):296–305, 5 1993.

[196] E. M. M. Peck, B. F. Yuksel, A. Ottley, R. J. Jacob, and R. Chang. Using fnirs brain sensing to evaluate information visualization interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 473–482, New York, NY, USA, 2013. ACM.

[197] R. W. Picard. *Affective Computing*. MIT Press, Cambridge, Massachsusetts, USA, 1997.

[198] R. W. Picard. Automating the recognition of stress and emotion: From lab to real-world impact. *IEEE MultiMedia*, 23(3):3–7, July 2016.

[199] M. Pichlmair, C. Brandt, M. Henrich, A. Biederer, I. Aslan, B. Bittner, and E. André. Pen-pen: A wellbeing design to help commuters rest and relax. In *Proceedings of the Workshop on Human-Habitat for Health (H3): Human-Habitat Multimodal Interaction for Promoting Health and Well-Being in the Internet of Things Era*, H3 '18, pages 6:1–6:9, New York, NY, USA, 2018. ACM.

[200] M. Pielot, K. Church, and R. de Oliveira. An in-situ study of mobile phone notifications. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '14, pages 233–242, New York, NY, USA, 2014. ACM.

[201] M. Pielot and L. Rello. Productive, anxious, lonely: 24 hours without push notifications. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, pages 11:1–11:11, New York, NY, USA, 2017. ACM.

[202] J. Pietilä, S. Mehrang, J. Tolonen, E. Helander, H. Jimison, M. Pavel, and I. Korhonen. Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities. In *EMBEC & NBC 2017*, pages 145–148. Springer, Jun 2017.

[203] G. Plasqui, A. G. Bonomi, and K. R. Westerterp. Daily physical activity assessment with accelerometers: new insights and validation studies. *Obesity Reviews*, 14(6):451–462, 2013.

[204] R. Poguntke, M. Hassib, T. Mantz, A. Schmidt, and S. Schneegaß. Smile to me: Investigating emotions and their representation in text-based messaging in the wild. In *Proceedings of the 2019 Conference Mensch and Computer*. ACM, 2019.

[205] R. Poguntke, J. Ilk, A. Schmidt, and Y. Abdelrahman. Designing thermal feedback for notifying users about stress. In *13th EAI International Conference on Pervasive Computing Technologies for Healthcare - Demos and Posters*, PervasiveHealth 2019. EAI, 6 2019.

[206] R. Poguntke, F. Kiss, A. Kaplan, A. Schmidt, and S. Schneegaß. Rainsense: Exploring the concept of a sense for weather awareness. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '18, pages 9–15, New York, NY, USA, 2018. ACM.

[207] R. Poguntke, C. Tasci, O. Korhonen, F. Alt, and S. Schneegaß. Avotar: Exploring personalized avatars for mobile interaction with public displays. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '18, pages 1–8, New York, NY, USA, 2018. ACM.

[208] R. Poguntke, M. Wirth, and S. Gradl. Same same but different: Exploring the effects of the stroop color word test in virtual reality. In *Proceedings of the 17th IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT)*. ACM, 2019.

[209] M. Z. Poh, N. C. Swenson, and R. W. Picard. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering*, 57(5):1243–1252, May 2010.

[210] H. Pohl, D. Becke, E. Wagner, M. Schrapel, and M. Rohs. Wrist compression feedback by pneumatic actuation. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 315–318, New York, NY, USA, 2015. ACM.

[211] H. Pohl, P. Brandes, H. Ngo Quang, and M. Rohs. Squeezeback: Pneumatic compression for notifications. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 5318–5330, New York, NY, USA, 2017. ACM.

[212] Z. Pousman and J. Stasko. A taxonomy of ambient information systems: Four patterns of design. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '06, pages 67–74, New York, NY, USA, 2006. ACM.

[213] C. K. Prahalad and V. Ramaswamy. The new frontier of experience innovation. *MIT Sloan Management Review*, 44(4):12–18, 2003.

[214] J. T. Ptacek, R. E. Smith, and K. L. Dodge. Gender differences in coping with stress: When stressor and appraisals do not differ. *Personality and Social Psychology Bulletin*, 20(4):421–430, 1994.

[215] A. A. Quesada, R. M. Tristao, R. Pratesi, and O. T. Wolf. Hyper-responsiveness to acute stress, emotional problems and poorer memory in former preterm children. *Stress*, 17(5):389–399, 2014.

[216] A. Raij, A. Ghosh, S. Kumar, and M. Srivastava. Privacy risks emerging from the adoption of innocuous wearable sensors in the mobile environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 11–20, New York, NY, USA, 2011. ACM.

[217] J. Rantakari, V. Inget, A. Colley, and J. Häkkilä. Charting design preferences on wellness wearables. In *Proceedings of the 7th Augmented Human International Conference 2016*, AH '16, New York, NY, USA, 2016. ACM.

[218] R. E. Reid, J. A. Insogna, T. E. Carver, A. M. Comptour, N. A. Bewski, C. Sciortino, and R. E. Andersen. Validity and reliability of fitbit activity monitors compared to actigraph GT3X+ with female adults in a free-living environment. *Journal of Science and Medicine in Sport*, 20(6):578–582, Feb 2018.

[219] H. Ren, M. Q. H. Meng, and X. Chen. Physiological information acquisition through wireless biomedical sensor networks. In *2005 IEEE International Conference on Information Acquisition*, pages 6 pp.–, June 2005.

[220] K. Renaud, J. Ramsay, and M. Hair. 'you've got e-mail!'... shall i deal with it now? electronic mail from the recipient's perspective. *International Journal of Human-Computer Interaction*, 21(3):313–332, 2006.

[221] C. Reynolds and R. Picard. Affective sensors, privacy, and ethical contracts. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pages 1103–1106, New York, NY, USA, 2004. ACM.

[222] M. Riediger, C. Wrzus, K. Klipker, V. Muller, F. Schmiedek, and G. G. Wagner. Outside of the laboratory: Associations of working-memory performance with psychological and physiological arousal vary with age. *Psychology and Aging*, 29(1):103–114, 2014.

[223] M. E. Rosenberger, M. P. Buman, W. L. Haskell, M. V. McConnell, and L. L. Carstensen. 24 hours of sleep, sedentary behavior, and physical activity with nine wearable devices. *Medicine and Science in Sports and Exercise*, 48(3):457–465, Mar 2016.

[224] D. W. Rowe, J. Sibert, and D. Irwin. Heart rate variability: Indicator of user state as an aid to human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, pages 480–487, New York, NY, USA, 1998. ACM Press/Addison-Wesley Publishing Co.

[225] M. A. Rupp, J. R. Michaelis, D. S. McConnell, and J. A. Smither. The impact of technological trust and self-determined motivation on intentions to use wearable fitness technology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1):1434–1438, 2016.

[226] M. A. Rupp, J. R. Michaelis, D. S. McConnell, and J. A. Smither. The role of individual differences on perceptions of wearable fitness device trust, usability, and motivational impact. *Applied Ergonomics*, 70:77 – 87, 2018.

[227] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

[228] A. Sahami Shirazi, N. Henze, T. Dingler, M. Pielot, D. Weber, and A. Schmidt. Large-scale assessment of mobile notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 3055–3064, New York, NY, USA, 2014. ACM.

[229] P. Sanches, K. Höök, E. Vaara, C. Weymann, M. Bylund, P. Ferreira, N. Peira, and M. Sjölinder. Mind the body!: Designing a mobile stress management application encouraging personal reflection. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, DIS '10, pages 47–56, New York, NY, USA, 2010. ACM.

[230] J. Sauro. Measuring usability with the System Usability Scale (SUS). `http://www.measuringusability.com/sus.php`, 2011. Retrieved August 6, 2019.

[231] A. Schäfer and J. Vagedes. How accurate is pulse rate variability as an estimate of heart rate variability? *International Journal of Cardiology*, 166(1):15–29, Jun 2013.

[232] M. F. Scheier, J. K. Weintraub, and C. S. Carver. Coping with stress: divergent strategies of optimists and pessimists. *Journal of Personality and Social Psychology*, 51(6):1257, 1986.

[233] J. Scheirer, R. Fernandez, J. Klein, and R. W. Picard. Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers*, 14(2):93–118, 2002.

[234] U. Schimmack and A. Grob. Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality*, 14(4):325–345, 2000.

[235] A. Schmidt, A. Takaluoma, and J. Mäntyjärvi. Context-aware telephony over wap. *Personal Ubiquitous Comput.*, 4(4):225–229, Jan. 2000.

[236] S. Schneegaß, R. Poguntke, and T. Machulla. Understanding the impact of information representation on willingness to share information. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 523:1–523:6, New York, NY, USA, 2019. ACM.

[237] S. Schneegaß and R. Rzayev. Embodied notifications: Implicit notifications through electrical muscle stimulation. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '16, pages 954–959, New York, NY, USA, 2016. ACM.

[238] N. Schneiderman, G. Ironson, and S. D. Siegel. Stress and health: psychological, behavioral, and biological determinants. *Annu. Rev. Clin. Psychol.*, 1:607–628, 2005.

[239] D. Schuurman, D. Mahr, and L. De Marez. User characteristics for customer involvement in innovation processes: Deconstructing the lead user-concept. (XXII):1–9, 2012.

[240] V. Schwind, N. Deierlein, R. Poguntke, and N. Henze. Understanding the social acceptability of mobile devices using the stereotype content model. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 361:1–361:12, New York, NY, USA, 2019. ACM.

[241] SecurEnvoy. 66% of the population suffer from nomophobia the fear of being without their phone. `https://www.securenvoy.com/en-gb/blog/66-population-suffer-nomophobia-fear-being-without-their-phone`, February 2012. Retrieved April 5, 2019.

[242] H. Selye. *The Stress of Life*. McGraw-Hill Book Company, 1956.

[243] H. Selye. Forty years of stress research: principal remaining problems and misconceptions. *Canadian Medical Association Journal*, 115(1):53, 1976.

[244] H. Selye. *Stress without Distress*, pages 137–146. Springer US, Boston, MA, 1976.

[245] P. Seraganian, A. Szabo, and T. G. Brown. The effect of vocalization on the heart rate response to mental arithmetic. *Physiology & Behavior*, 62(2):221–224, 1997.

[246] 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.

[247] F. Shaffer and J. P. Ginsberg. An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5:258, 2017.

[248] F. Shaffer, R. McCraty, and C. L. Zerr. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in Psychology*, 5(26):139, Sep 2014.

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

[250] M. Sharmin, A. Raij, D. Epstien, I. Nahum-Shani, J. G. Beck, S. Vhaduri, K. Preston, and S. Kumar. Visualization of time-series sensor data to inform the design of just-in-time adaptive stress interventions. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, pages 505–516, New York, NY, USA, 2015. ACM.

[251] A. Shcherbina, C. M. Mattsson, and D. Waggott. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of Personalized Medicine*, 7(2):3, 2017.

[252] R. Sloan, P. Shapiro, E. Bagiella, S. Boni, M. Paik, J. Bigger, R. Steinman, and J. Gorman. Effect of mental stress throughout the day on cardiac autonomic control. *Biological Psychology*, 37(2):89–99, 1994.

[253] S. Song, G. Noh, J. Yoo, I. Oakley, J. Cho, and A. Bianchi. Hot & tight: Exploring thermo and squeeze cues recognition on wrist wearables. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, ISWC '15, pages 39–42, New York, NY, USA, 2015. ACM.

[254] K. Spiel, F. Kayali, L. Horvath, M. Penkler, S. Harrer, M. Sicart, and J. Hammer. Fitter, happier, more productive?: The normative ontology of fitness trackers. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages alt08:1–alt08:10, New York, NY, USA, 2018. ACM.

[255] A. A. Stanley and K. J. Kuchenbecker. Evaluation of tactile feedback methods for wrist rotation guidance. *IEEE Transactions on Haptics*, 5(3):240–251, March 2012.

[256] F. A. Storm, B. W. Heller, and C. Mazzà. Step detection and activity recognition accuracy of seven physical activity monitors. *PloS one*, 10(3):1–13, 03 2015.

[257] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. L. Griss. Activity-aware mental stress detection using physiological sensors. *MobiCASE*, 76(2):282–301, 2010.

[258] C. S. S. Tan, J. Schöning, K. Luyten, and K. Coninx. Investigating the effects of using biofeedback as visual stress indicator during video-mediated collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 71–80, New York, NY, USA, 2014. ACM.

[259] 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. *Circulation*, 93(5):1043–1065, Mar 1996.

[260] R. E. Thayer. *The Biopsychology of Mood and Arousal*. Oxford University Press, 1990.

[261] J. Tholander and S. Nylander. Snot, sweat, pain, mud, and snow: Performance and experience in the use of sports watches. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2913–2922, New York, NY, USA, 2015. ACM.

[262] J. Tomaka, J. Blascovich, and L. Swart. Effects of vocalization on cardiovascular and electrodermal responses during mental arithmetic. *International Journal of Psychophysiology*, 18(1):23–33, 1994.

[263] K. A. Townsend, J. W. Haslett, T. K. K. Tsang, M. N. El-Gamal, and K. Iniewski. Recent advances and future trends in low power wireless systems for medical applications. In *Fifth International Workshop on System-on-Chip for Real-Time Applications (IWSOC'05)*, pages 476–481, July 2005.

[264] C. E. Tudor-Locke and A. M. Myers. Methodological considerations for researchers and practitioners using pedometers to measure physical (ambulatory) activity. *Research Quarterly for Exercise and Sport*, 72(1):1–12, 2001. PMID: 11253314.

[265] C. H. Vinkers, R. Penning, J. Hellhammer, J. C. Verster, J. H. Klaessens, B. Olivier, and C. J. Kalkman. The effect of stress on core and peripheral body temperature in humans. *Stress*, 16(5):520–530, 2013.

[266] A. Visuri, R. Poguntke, and E. Kuosmanen. Proposing design recommendations for an intelligent recommender system logging stress. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2018, pages 411–417, New York, NY, USA, 2018. ACM.

[267] E. Vlemincx, I. Van Diest, and O. Van den Bergh. A sigh following sustained attention and mental stress: Effects on respiratory variability. *Physiology & Behavior*, 107(1):1–6, Aug 2012.

[268] M. P. Wallen, S. R. Gomersall, S. E. Keating, U. Wisløff, and J. S. Coombes. Accuracy of heart rate watches: Implications for weight management. *PloS one*, 11(5):e0154420, May 2016.

[269] G. Wang and A. Suh. Disorder or driver?: The effects of nomophobia on work-related outcomes in organizations. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 50:1–50:12, New York, NY, USA, 2018. ACM.

[270] R. Wang, F. Quek, D. Tatar, K. Teh, and A. Cheok. Keep in touch: Channel, expectation and experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 139–148, New York, NY, USA, 2012. ACM.

[271] R. Weitkunat, C. R. Coggins, Z. Sponsiello-Wang, G. Kallischnigg, and R. Dempsey. Assessment of cigarette smoking in epidemiologic studies. *Beiträge zur Tabakforschung/Contributions to Tobacco Research*, 25(7):638–648, 2013.

[272] T. Westermann, S. Möller, and I. Wechsung. Assessing the relationship between technical affinity, stress and notifications on smartphones. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '15, pages 652–659, New York, NY, USA, 2015. ACM.

[273] D. P. Williams, C. Cash, C. Rankin, A. Bernardi, J. Koenig, and J. F. Thayer. Resting heart rate variability predicts self-reported difficulties in emotion regulation: a focus on different facets of emotion regulation. *Frontiers in Psychology*, 6:122, Mar 2015.

[274] G. Wilson, S. Brewster, M. Halvey, and S. Hughes. Tempera-tour, hot apps, cool widgets: Thermal feedback for mobile devices. In *Proceedings of the 14th International Conference on Human-computer Interaction with*

Mobile Devices and Services Companion, MobileHCI '12, pages 211–212, New York, NY, USA, 2012. ACM.

[275] G. Wilson, S. Brewster, M. Halvey, and S. Hughes. Thermal icons: Evaluating structured thermal feedback for mobile interaction. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '12, pages 309–312, New York, NY, USA, 2012. ACM.

[276] G. Wilson and S. A. Brewster. Multi-moji: Combining thermal, vibrotactile & visual stimuli to expand the affective range of feedback. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 1743–1755, New York, NY, USA, 2017. ACM.

[277] G. Wilson, S. A. Brewster, and M. Halvey. The effects of walking and control method on pressure-based interaction. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 2275–2280, New York, NY, USA, 2011. ACM.

[278] G. Wilson, S. A. Brewster, M. Halvey, A. Crossan, and C. Stewart. The effects of walking, feedback and control method on pressure-based interaction. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 147–156, New York, NY, USA, 2011. ACM.

[279] G. Wilson, G. Davidson, and S. A. Brewster. In the heat of the moment: Subjective interpretations of thermal feedback during interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2063–2072, New York, NY, USA, 2015. ACM.

[280] G. Wilson, D. Dobrev, and S. A. Brewster. Hot under the collar: Mapping thermal feedback to dimensional models of emotion. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 4838–4849, New York, NY, USA, 2016. ACM.

[281] G. Wilson, M. Halvey, S. A. Brewster, and S. A. Hughes. Some like it hot: Thermal feedback for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2555–2564, New York, NY, USA, 2011. ACM.

[282] G. Wilson, H. Maxwell, and M. Just. Everything's cool: Extending security warnings with thermal feedback. In *Proceedings of the 2017*

CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '17, pages 2232–2239, New York, NY, USA, 2017. ACM.

[283] M. Wilson, J. Doyle, and G. McTaggart. Designing stress management interventions for older adults to improve wellbeing. In *Proceedings of the 2015 British HCI Conference*, British HCI '15, pages 307–308, New York, NY, USA, 2015. ACM.

[284] T. Winograd and D. Woods. The challenge of human-centered design. `http://www.ifp.illinois.edu/nsfhcs/bog_reports/bog3.html`, April 1997. Retrieved July 24, 2019.

[285] R. Winsky-Sommerer, B. Boutrel, and L. de Lecea. Stress and arousal. *Molecular Neurobiology*, 32(3):285–294, Dec 2005.

[286] N. Wouters, J. Downs, M. Harrop, T. Cox, E. Oliveira, S. Webber, F. Vetere, and A. Vande Moere. Uncovering the honeypot effect: How audiences engage with public interactive systems. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, DIS '16, pages 5–16, New York, NY, USA, 2016. ACM.

[287] H. I. Xie. Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. *Journal of the American Society for Information Science*, 51(9):841–857, 2000.

[288] R. M. Yerkes and J. D. Dodson. The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5):459–482, 1908.

[289] S. Yoon, S.-s. Lee, J.-m. Lee, and K. Lee. Understanding notification stress of smartphone messenger app. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, pages 1735–1740, New York, NY, USA, 2014. ACM.

[290] B. Yu, M. Funk, J. Hu, and L. Feijs. Stresstree: A metaphorical visualization for biofeedback-assisted stress management. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, DIS '17, pages 333–337, New York, NY, USA, 2017. ACM.

[291] R. B. Zajonc. Social facilitation. *Science*, 149(3681):269–274, 1965.

[292] J. Zhai and A. Barreto. Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In *2006*

International Conference of the IEEE Engineering in Medicine and Biology Society, pages 1355–1358, Aug 2006.

[293] J. Zhai and A. Barreto. Stress recognition using non-invasive technology. In *Proceedings of the 19th International Florida Artificial Intelligence Research and Society Conference FLAIRS*, pages 395–400, 2006.

[294] Z. Zhang, Z. Pi, and B. Liu. Troika: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. *IEEE Transactions on Biomedical Engineering*, 62(2):522–531, Feb 2015.

[295] Y. Zheng, E. Su, and J. B. Morrell. Design and evaluation of pactors for managing attention capture. In *2013 World Haptics Conference (WHC)*, pages 497–502, April 2013.

[296] F. Zhou, Y. Ji, and R. J. Jiao. Affective and cognitive design for mass personalization: status and prospect. *Journal of Intelligent Manufacturing*, 24(5):1047–1069, Oct 2013.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12. Juli 2011, § 8, Abs. 2 Pkt. 5.)

Hiermit erkläre ich an Eides statt, dass die Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt wurde.

Stuttgart, den 26.05.2020

Romina Poguntke geb. Kettner

Romina Poguntke

Understanding Stress Responses Related to Digital Technologies

Feeling stressed is a common phenomenon which is known among all age groups, cultures, and societal classes. Although its perception and recognition may follow universal rules too, humans experience the reasons for stress and the success of stress mitigating techniques differently. Research in various domains, such as psychology, biology, or neuroscience have made great progress in understanding the evolutionary roots of stress and its stressors, namely stimuli that elicit stress reactions. However, the nature of stressors has changed significantly over time. While thousands of years ago humans experienced stress physiologically when facing a dangerous animal in the wilderness, nowadays stress is a harmful consequence of the permanent challenge for our brain to process the constant stream of information. As a result, the modern human feels stressed by the omnipresent access to information resulting in information overload, going hand in hand with the societal expectations to be reachable and so-called "online“ anytime. The consequential problems for users, as well as design challenges for designers and developers to avoid such disagreeable effects have been recently addressed in the field of Human-Computer Interaction (HCI).

In this thesis, I identify, describe and successively explore the character of different inventions for reducing the stress level of users and their impact on them. Hereby, I present three exemplary research probes differing in their degree of privateness and their degree of using digital technologies.