KU LEUVEN

FACULTEIT INGENIEURSWETENSCHAPPEN

Analysis of noninvasive biomarkers for oxytocin treatment effects in social stress regulation

Jef Masereel

Thesis submitted for the degree of
Master of Science in
Biomedical Engineering

Thesis supervisors:
Prof. dr. ir. S. Van Huffel
Prof. dr. ir. C. Varon Perez
Prof. dr. K. Alaerts

Assessors:
ir. D. Huysmans
ir. A. Rozo
prof. dr. B. Boets

Mentors:
ir. J. Morales
N. Daniels

Academic year 2020 – 2021

Preface

The goal in thesis was to provide the MOX research team with additional quantitative insights in their experimental data. My hope is that it serves them well, and that it can provide a few interesting leads for future research.

It has been a strange year, but I am very grateful for the people around me and the opportunity to have learned more about myself. I wish to everyone the stability and freedom to make it through this wild new world with a fresh perspective, and a positive outlook on life. Personally, I am very happy with the lessons learned and experiences shared with my friends, family and colleagues during these times. I want to thank all of them for their support during the development of this thesis, in no particular order.

I am very grateful for my parents and siblings, who provided a warm place to land and share our good and bad days. It has been great fun to be home again for a longer time and catch up with my not-so-little-anymore siblings. My friends Jeffrey, Stef, Karen, Nina, Hannah, Marnik, Annelien, Stacy ... are deeply appreciated, for sharing their stories and adventures when it was time to go outdoors and take a break, and their motivation and support when it was time to work again. I write this in loving memory of Majo, Monique, Paul, Kiefer and Johnny, whose lessons I will never forget. Although they could not be here today, I look forward to carry them with me in the years to come.

Finally, a massive thanks to John, Andrea, Dorien, Nicky, and professors van Huffel, Alaerts and Varon for their technical advice and feedback, as well as their patience and understanding during the first months of 2020. I could not have made it this far without you.

Jef Masereel

Contents

Abstract

Oxytocin (OT) has recently gained increased interest as a potential treatment for disorders related to social stress (dys)regulation. Early studies attributed prosocial and anxiolytic effects ("tend and defend") to exogenous OT, while later findings indicated more complex contextual dependencies affecting these response mechanisms. This thesis aims to contribute to recent work in this domain by performing a broad explorative analysis of the expression of social stress regulation in blood volume pulse (BVP), electrodermal activity (EDA) and respiration (RSP) under influence of social stress and single-dose OT treatment (24 IU).

Daniels et al. collected the BVP, EDA and RSP for a group of 56 healthy, neurotypical (NT) adult men at rest and during intermittent eye-contact (social stress), before and after OT administration. Additionally, the participants' attachment styles were assessed using the State Adult Attachment Measure (SAAM), in order to assess the role of attachment (dys)regulation in social stress regulation and how it is influenced by OT treatment. Three feature sets are extracted from this data: BVP-based measures of pulse rate variability, features of amplitude and response variability in EDA, and respiratory variability. To do so, each of the three signal modalities must first be filtered and processed, which requires certain trade-offs to be made. The extracted features are then analyzed with a combination of univariate tests, complemented by a positive false detection rate (pFDR) correction to compare findings across features and estimate the risk of false detections. Additionally, correlation tests provide further insight in the interactions of OT treatment with social context and participants' attachment style.

With this approach, the feature set was found to capture multiple expressions of the targeted response mechanisms. OT-specific treatment effects on recovery features in the EDA aligned with prior findings and a new treatment effect was found in the mean respiratory rate, but PRV-based power features were not found to be significant despite expectations from prior studies. The correlation tests found multiple strong interactions, but might not extrapolate well to other target populations due to the limited variability in SAAM scores for the participants in this dataset.

In conclusion, this thesis reports the effects of OT and social stress on healthy, adult NT men as captured by a broad set of features extracted from noninvasive biomarkers. As such, it can be used as a point of comparison when studying different target populations or physiological response mechanisms. More research should be done to confirm the potential of exogenous OT as a treatment for social disorders in neurodivergent patients.

List of Figures and Tables

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

ASD	Autism spectrum disorder
OT	Oxytocin
ANS	Autonomic nervous system
IU	International units (intranasal OT dosage)
IPPA	Inventory of Parent and Peer Attachment
SRS	Social Responsiveness Scale
RBS	Repetitive Behavior Scale
SAAM	State Adult Attachment Measure
NT	Neurotypical
BVP	Blood volume pressure
PPG	Pulse photoplethysmography
EDA	Electrodermal activity
SC	Skin Conductance (measure of EDA [μS])
GSR	Galvanic Skin Response (idem SC [μS])
RSP	Respiration signal (approximate measure from thoracic breathing belt)
HRV	Heart Rate Variability (ECG-derived)
PRV	Pulse Rate Variability (BVP/PPG-derived HRV)
PSD	Power spectral density
MAD	Mean absolute deviation
STD	Standard deviation
MN	Statistical means
PDA	Pulse decomposition analysis
SCR	Skin conductance response
SMNA	Sudomotor nerve activation
IRF	Impulse response function
PVM	Poral valve model
PO	Pore opening
RVT	Respiratory volume per unit time
CV	Coefficient of variation
AR	Autocorrelation at one breath lag (for respiratory biomarker)
pFDR	positive false detection rate [3]

Symbols

f_s	Sample frequency (Hz)
$x_{BVP}(n)$	Discrete time signal for BVP activity
n_A	Apex point of BVP signal
n_B	Base point of BVP signal
n_M	Midpoint of BVP signal
$IRF(t)$	Impulse Response Function for EDA decomposition
$IRF_k(t)$	k-th iteration of the $IRF(t)$ estimate
λ	Power value for Box-Cox transformation
α	Significance threshold for statistical tests
p	p-value of a statistical hypothesis test
q	q-value obtained by pFDR correction [3]
ftr	Dataset with normalized feature values (cfr. Table 4.2)
dftr	Dataset with pre- to post-treatment differences of *ftr*
bcsr	Dataset with baseline-corrected stress responses (from *ftr*)
dbcsr	Dataset with pre- to post-treatment differences of *bcsr*

Chapter 1

Introduction

This thesis explores the broad quantitative effects of oxytocin treatment on social stress regulation in NT adult men, as part of recent studies conducted by Daniels et al. [4, 5, 6, 7, 8]. This chapter introduces the key concepts of this analysis to support further decisions and interpretations. In the first section, the topic of oxytocin as a prosocial medical treatment is situated in its broader clinical context. Once the basic mechanisms at play are understood, the second section discusses which noninvasive biomarkers can provide insight in the expression of these mechanisms. The third section then specifies the details of the studied dataset. Finally, the chapter closes with the formulation of the problem statement for this thesis, and provides a summary of the technical approach used to address these questions in further chapters.

1.1 Clinical context

1.1.1 Key concepts of oxytocin as a treatment

Although oxytocin (OT) has become popularly known as the "love hormone", its exact neurophysiological functions and pathways are still under discussion. According to Bakermans et al., it is produced in the hypothalamus, from where it is projected into the bloodstream as well as onto various receptors within the brain. The OT that enters the bloodstream acts *"as a hormone"*, in interaction with the autonomic nervous system (ANS). For the receptors within the brain, OT acts *"as a neurotransmitter and neuromodulator"* of complex regulatory mechanisms. It is considered to be an important actor in social behaviors related to caregiving, social bonding, cooperation and the protection of loved ones. Note that this last element is important to understand the increased aggression and reduced trust towards unfamiliar peers in competitive situations reported in some studies [9].

Dosage and administration method

Early studies explored the use of exogenous OT as a treatment to address disorders related to these social regulatory mechanisms. Bakermans et al. reviewed several studies, concluding that low-dosage intranasal administration of OT yields the most

desireable outcomes [9]. Compared to intravenous administration, the uptake of OT (measured via peripheral OT levels such as in saliva) is much more consistent when using a nasal spray. Additionally, Bakermans found that participants are unable to distinguish placebo from oxytocin when using intranasal administration, *"enabling truly double-blind, placebo-controlled experiments"* [10]. In terms of dosage, the evidence suggests that low quantities (16-24 international units [IU]) are sufficient to achieve the desired response. The increase in peripheral OT levels reportedly remains stable for up to seven hours thanks to a natural feedforward mechanism. Increases in dosage (up to 7000 IU) returned less consistent findings. Generally, high dosage OT administration is expected to result in neutral or even adverse effects compared to low dosage studies. Although it must be studied further, most current research focuses on intranasal dosages of 16 or 24 IU.

Interactionist approach

So far, the "true" effects of exogenous OT are still under discussion. Prior studies have reported many findings that are inconsistent with each other at first sight. The interactionist approach proposed by Bartz et al. reframes these inconsistencies as important insights in the relation between various observed factors and the underlying mechanism. By documenting all known features of the studied individuals, as well as any contextual factors imposed in an experiment, the reported findings become more valuable when studying the observed effects across studies. [11]

Bartz, Graustella and Bakermans used this interactionist approach to systematically collect findings from earlier studies regarding OT treatment in the social domain, and formulate a more cohesive approach for future work. Based on the original interpretation of OT as a "tend and defend" regulator, the main interest lies in its potential for the treatment or alleviation of social (aspects of) disorders such as social anxiety, attachment dysregulation, autism spectrum disorder (ASD), depression, obsessive compulsive disorder (OCD) and bipolar disorder (BPD).

Known effects of OT treatment

The core effects of OT administration are two-fold: first, evidence suggests a facilitating role in social cognition, meaning the detection, recognition and understanding of social, emotional and affective cues that are relevant in social interactions. Secondly, OT was found to have anxiolytic and prosocial effects, which expresses itself as a constructive improvement in behaviors related to the formation and maintenance of social bonds [11, 12].

The expression of these effects has been attributed to various hypothetical models. Bartz et al. interpret their findings as the results of an interaction between anxiety reduction, affiliative motivation and perceptual selectivity or social salience [11]. More pragmatically, Bakermans et al. formulate a set of important factors affecting the underlying response mechanism [9]. They suggest that social context, personality traits and personal history must be taken into account when discussing the effects of OT. Note that factors such as neurodiversity and social stress disorders are typically

attributed to both personality factors as a combination of genetic predisposition and elements of upbringing.

Taking all these considerations into account, Bakermans et al. published important conclusions regarding the role of OT in social stress regulation. Across all reviewed studies, only autism-related studies reported a consistently positive outcome. In those studies, the observed effects included a reduction in repetitive behaviors, positive effects on social cognition and improved recognition of facial emotions. Additionally, significant evidence was found for increased (defensive) aggressive behavior in a competitive setting. Individuals under influence of OT showed increased trust and preference towards "better" players in a competitive game, which increased their chances of success. This helps to explain the increased alertness in some subjects when interacting with unfamiliar peers [9].

1.1.2 Assessment of person-specific response factors

As mentioned earlier, the effects of OT treatment are often found to be influenced by person-dependent factors such as attachment style and/or social responsiveness. Dysregulation of such mechanisms is considered to be caused by interactions of fixed personality traits and learned/developed changes such as early adversity in upbringing or traumatic events. Note, however, that complete verification of this assumption is outside the scope of this thesis. A few different assessment methods exist which can be used to quantify these inter-personal variations in the (dys)regulation of social stress. By including this information in the data collection of a study, important moderation effects within the sample can be detected and addressed for a more thorough conclusion.

For the assessment of attachment style, there are two commonly used methods. The Inventory of Parent and Peer Attachment (IPPA, 1987) provides a measure for anxious, avoidant and secure attachment under the assumption that these are fixed values across the life of a given subject [13]. More recently, the State Adult Attachment Measure (SAAM, 2009) was published. This newer method uses the same three scores as ordinal scales from one to seven, but removes the stationary assumption such that it can be used to assess changes between different conditions or time periods [14]. High scores on the anxiety and/or avoidance metrics and/or a low score on secure attachment are a reliable indicator of attachment disorders.

Two more methods are relevant for the analysis of regulation disorders related to ASD. The Social Responsiveness Scale (SRS, 2003) is a broad assessment of social symptoms that are commonly observed in people with autism [15]. The Repetitive Behavior Scale (RBS, 2007) quantifies patterns of restricted repetitive behavior by means of a questionnaire [16]. Both are considered to be reliable predictors of autism spectrum disorder as defined by DSM-IV criteria (Diagnostic and Statistical Manual of Mental Disorders, 1994) [17]. Note, however, that ASD diagnostics remain heavily debated in modern research (e.g. see [18]). A revised diagnostic (DSM-V) was published by the American Psychiatric Association in 2013 [19].

1.1.3 Effects of OT treatment on social stress regulation

Recent studies coordinated by professor Kaat Alaerts[1] are exploring the mechanisms responsible for social stress regulation, and how OT treatment might be used to alleviate related symptoms. To do so, the hypotheses formulated in previous studies require further testing and refinement. This thesis contributes to their research regarding the interactions between OT treatment, attachment variations and social stress induced by eye-contact. The following paragraphs summarize the state of the art in OT research, which will provide the relevant context for the problem statement defined in Section 1.4. The technical concepts mentioned below are discussed in more detail in later sections.

All studies cited here share a few common factors to build a more cohesive body of findings. OT was always administered as a nasal spray to maintain a consistent dosage of 24 IU (3 sniffs of 4 IU in each nostril). The SRS, RBS, IPPA and/or SAAM scores were collected before experiments to assess inter-personal variations that might interact with the OT treatment [4, 5, 6, 7, 8].

Reference group neurotypical adults

The first set of experiments targeted adult neurotypical (NT) men to assess the influence of OT in the absence of significant social disorders. Prinsen et al. measured the motor-resonance of participants using transcranial magnetic stimulation (TMS). They found that, although no uniform treatment effect was observed for the OT group, motor-resonance in subjects with high self-reported attachment avoidance was significantly enhanced by OT treatment under conditions of direct eye-contact [4].

Additional studies by Daniels et al. analyzed the effects of OT on measures of skin conductance and pulse rate variability in another group of adult NT men. The first study reports a significant reduction in the recovery time of skin conductance responses after OT administration. The treatment effects were further enhanced in subjects with low SRS scores [5]. The second study focuses on autonomic cardiac function, measured by estimates of low and high frequency components in the subjects' heart rate variability (LF-HRV and HF-HRV, respectively). In resting conditions, OT enhanced HF-HRV and reduced LF-HRV, which indicates a shift from sympathetic to parasympathetic tone. During eye-contact, these effects were suppressed due to the need for social alertness [6].

Adults with autism spectrum disorder

The second set of experiments studied adult men with ASD in a similar approach, to compare findings with those for NT men. Alaerts et al. measured peripheral OT levels before and after single-dose administration to explore effects on connectivity levels in brain regions of the central oxytocinergic system (by means of fMRI). Higher levels of endogenous OT (pre-treatment) were found to correlate with lower functional coupling between the targeted brain regions. Interestingly, measures of core autism

[1] Research Group for Neuromotor Rehabilitation, Department of Rehabilitation Sciences, KU Leuven, Belgium

traits (SRS, RBS) did not correlate with the measured OT levels, while attachment measures (IPPA, SAAM) did show a positive correlation with the secure attachment score [7].

Bernaerts et al. conducted a one year follow-up of the same group of men with ASD to gain insight in the long-term influence of OT when administered once daily (24 IU/day) for four weeks. No OT-specific effects were observed in SRS scores. However, subjects did report reduced feelings of avoidance (SAAM) and a reduction in repetitive behavior (RBS) after those four weeks of treatment. The improvements in attachment avoidance were still significant for the OT group at one year follow-up despite ending the treatment at four weeks [8].

These results are very promising, but they confirm again that the treatment effects of exogenous OT are very dependent on situational and personal context. To gain a better understanding of the regulation mechanisms as a whole, it would be beneficial to analyze various different biomarkers and their interactions with person-dependent factors such as the assessment scales discussed earlier.

1.2 Noninvasive biomarkers

1.2.1 Psychological stress in the ANS

The response mechanisms of interest are deeply rooted in the ANS, which subconsciously regulates bodily functions in response to external stimuli [20]. The research introduced in the previous section used social stimuli in the form of directed and/or averted eye-contact to induce psychological stress. The resulting state of increased stress levels under such conditions is further referred to as social stress.

M. Ziegler describes how the ANS regulates physiological activity under influence of psychological stress [20]. Stressful events or experiences are interpreted by the ANS, which can trigger a cascade of response mechanisms in function of the nature of the stimulus and various person-dependent factors. These responses are typically described as an activation of the sympathetic nervous system (SNS) in combination with inhibition of the parasympathetic nervous system (PNS). The SNS is related to functions of "fight or flight", such as increased alertness and muscle activation, while the PNS is related to functions of rest and recovery ("rest and digest").

Such a response to psychological stress is then expressed in various physiological processes. Well known indicators include increases in sweat gland activity, heart rate, blood pressure and cardiac output [20]. Since social stress is assumed to be closely related to general psychological stress, one would expect these biomarkers to behave similarly in the OT studies discussed in Section 1.1. Further literature review confirms their relevance for the assessment of mental or social stress. Additionally, measures of respiratory variability also show promise for these purposes [21, 22, 23, 24, 25, 26].

The dataset studied in this thesis (cfr. Section 1.3) contains recordings of blood volume pulse, electrodermal activity and respiration based on noninvasive acquisition methods. As such, the derived features can be reproduced in later studies without the need for medical intervention and with minimal risk to participants or patients.

The following section introduces each signal modality to support design decisions and interpretation in later chapters.

1.2.2 Blood volume pulse

Measurement of the blood volume pulse (BVP) by means of photoplethysmography (PPG) provides quantitative insight in cardiac activity. A PPG sensor is placed on the subject's finger, earlobe or toe such that it clamps the soft tissue between a light source and an optode. The light, typically in the infrared and/or ultraviolet range, passing through this tissue is attenuated in function of the blood perfusion and oxygenation levels. With the appropriate calibration, this attenuation is measured by the optode in the PPG device and translated to a time signal representing the blood-volume changes in the clamped tissue [27].

The variations of blood-volume in the observed tissue are directly linked to cardiac activity via the cardiovascular system: pulsatile contraction of the heart creates a pressure differential which pumps blood through the vascular system in order to transport oxygen and other molecules from and to the body. Bifurcations in the vascular pathway reflect incoming pressure waves, such that BVP measurements at a specific body site represent the superposition of a series of smaller reflections along the corresponding pathway onto the original cardiac profile. This means that the measured BVP signal contains information on the original cardiac pressure-flow profile, albeit transformed by a site-specific response function. Low frequency components in the BVP are related to SNS activity and thermoregulation, while high-frequency components provide insight in direct mechanical cardiac activity. Although no electrical activity is measured, PPG captures the heart rate close enough to ECG standards if implemented correctly [27, 28].

Figure 1.1 illustrates the three morphological variations that will become relevant in the next chapter. Typically, PPG waveforms follow a sawtooth profile: each heartbeat is represented by a left-leaning triangular pulse, with a small secondary peak on the right flank. The small deviation between primary and secondary peaks is the dicrotic notch. This is shown in (a). In some cases, however, this secondary peak appears to separate from the primary, creating an exceptionally pronounced dicrotic notch. An example of this can be seen in (b). As indicated before, these morphological variations are influenced by sensor placement, the subject's posture and their cardiovascular state [29, 28]. Additionally, PPG is still susceptible to motion artefacts and low signal quality if the clamped tissue is too dense. These external sources of noise can make it difficult to distinguish the BVP waveform in signals such as (c).

1.2.3 Electrodermal activity

Another relevant biomarker for psychosocial stress is electrodermal activity (EDA). As mentioned earlier, the SNS has been found to facilitate an increase in sweat gland activity in response to social stress, among other factors. By placing two or more skin electrodes on affected regions such as the hands, the sweat gland activity can

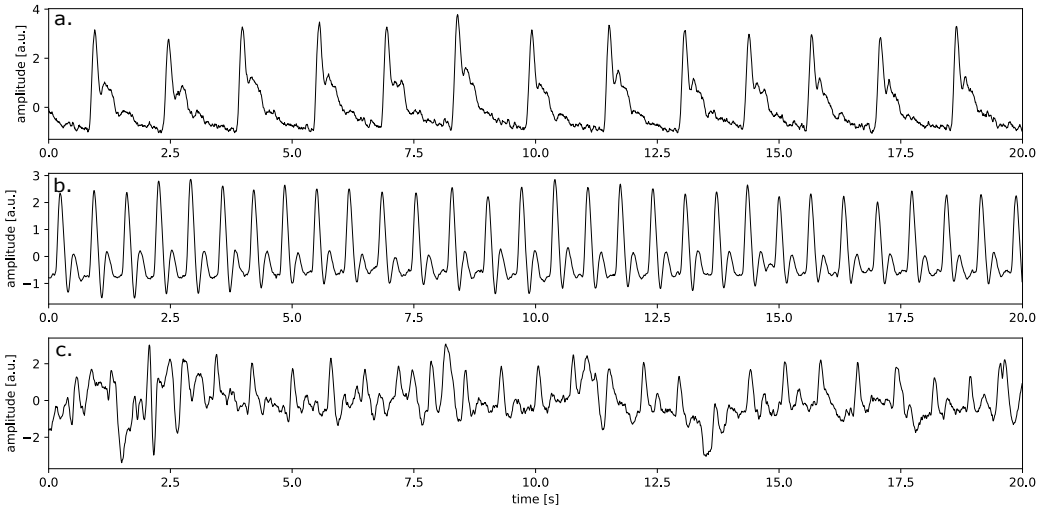

FIGURE 1.1: Morphological variations in the available BVP recordings: conventional PPG waveform (a), PPG with deep dicrotic notches, causing separation of the secondary pulse (b) and PPG with background noise (c).

be measured as fluctuations in electrical conductance between the electrodes (in microSiemens μS), hence the name electrodermal activity. [30].

The morphology of EDA signals is typically described as the superposition of high-frequency phasic activity on a low-frequency tonic level. The tonic level is a steady change in conductivity related to parasympathetic activity, while the phasic activity consists of a series of local peaks related to discrete sweat gland activations by the SNS [30]. These individual peaks are further referred to as skin conductance responses (SCR). Prior studies by Grundlehner and Setz et al. demonstrate how features of the tonic and phasic components can be used to assess various states of psychological arousal, which will become useful for later analysis [21, 23].

There are, however, some caveats when processing EDA signals. W. Boucsein describes in detail how the signal quality can be affected by various factors. A first point of concern is the contact between electrodes and skin. When using electrolyte gel to maintain contact, air bubbles and/or evaporation will cause a drift in the measured tonic level over time. Dry electrodes on the other hand are at risk of being affected by local moisture buildup from sweat excretion at the point of contact. Secondly, motion artefacts can be picked up from electrode movement, muscle activity near the electrode (background EMG). If possible, the subject should remain still during acquisition, which can be a challenge in some experiments. A more preventable issue is electromagnetic interference. Proper shielding of the electrode leads and filtering of known artefacts' frequencies, such as 50/60 Hz for power line noise, will limit EM background noise in most cases. Lastly, it is important to be aware of demographic variations: prior studies have found significant differences in response

threshold between people, as well as differences in recovery times after a response and effects of habituation and extinction when exposed to repeated stimuli. Although the details of these mechanisms are outside the scope of this thesis, it is important to be aware of their expression in signal morphology when reviewing a new set of recordings [31].

Figure 1.2 shows four segments, which are representative of the variations that will become relevant for this thesis. Segment (a) is a "normal" EDA signal with a relatively stable tonic level and well defined phasic responses. The second segment (b) shows a similar tonic baseline, but is completely void of any phasic activity. As introduced in the previous paragraph, this is likely a person with a very high activation threshold. Boucsein refers to them as nonresponders, since the normal stimulus levels do not manage to trigger a measurable phasic response [31]. Similarly, people with very low activation thresholds are referred to as hyperresponders, due to their tendency towards high response rates for the same stimulus levels. Segment (c) is an example of mild hyperresponse. The fast succession of SCRs can make it difficult to reliably assess the underlying nervous activity, but certain algorithms are able to address this. Lastly, segment (d) is an example of a corrupted recording. Although unconfirmed, the first half is likely excessively noisy due to poor skin contact or technical interference, while the jump in baseline level in the second half is likely a motion artefact. Such signals are no longer valid for use and should be removed from further analysis.

1.2.4 Respiration

Respiration (RSP) offers a third biomarker for the assessment of social stress regulation. Despite its relatively low popularity compared to the previous biomarkers, recent studies have found strong indicators of relevant regulation mechanisms in measures of respiratory variability and cardiorespiratory coupling [22, 32, 33]. Interestingly, while cardiac and electrodermal activity are fully autonomic functions, respiration is an autonomic function that can be overtaken by intentional motor control whenever the subject is conscious. Although this can make it more challenging to collect consistent data, respiratory features can provide useful insights if extracted correctly. Additionally, improvements in noninvasive acquisition methods have made respiration monitoring more accessible for applications in the realm of wearable and mobile health devices [34, 35].

For such applications, respiratory inductive plethysmography (RIP) has become a good alternative to traditional capnometry. Instead of requiring tracheal intubation to measure the pulmonary airflow directly, RIP measures the circumference of the thoracic and abdominal walls to estimate the volume of exchanged air. This is done noninvasively by strapping a belt with inductive coils or Hall effect sensors around the chest and/or abdomen, such that changes in the length of the belt can be converted to an electrical signal. Different versions of such breathing belts have been developed with varying degrees of accuracy and repeatability [36, 35, 37].

Despite the benefits of RIP, the measured signals are subject to certain limitations. Besides the previously mentioned variations in device quality, there are three technical

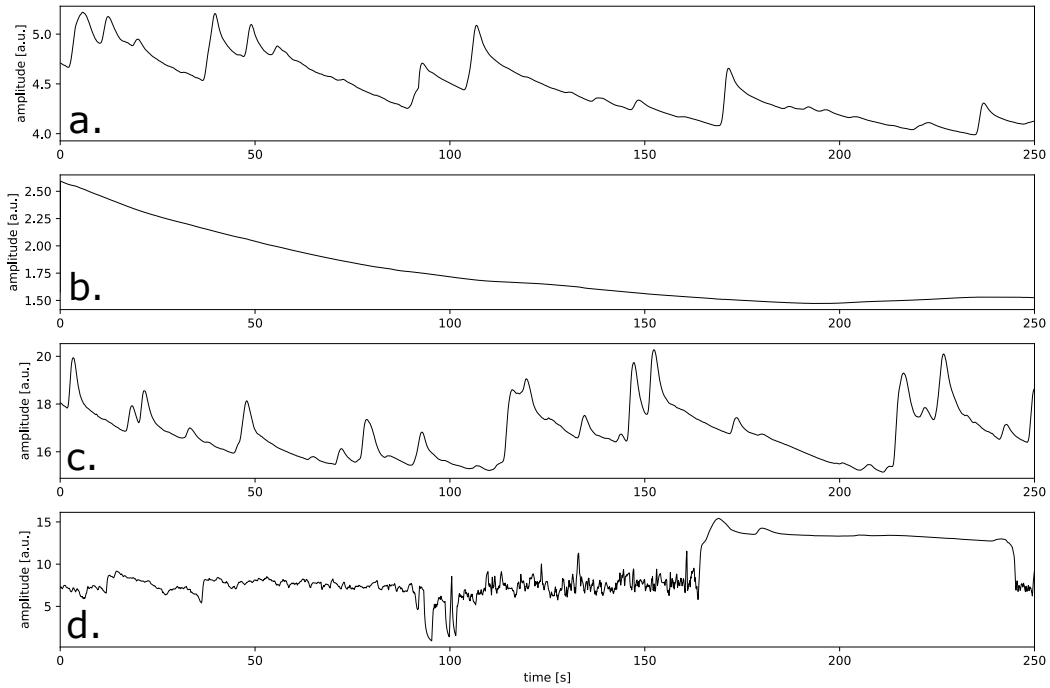

FIGURE 1.2: Morphological variations in the available EDA recordings: conventional EDA (a), EDA without phasic activity as seen in nonresponders (b), EDA with relatively high phasic activity (c) and a corrupted EDA segment (d).

challenges when processing RIP signals: first, a single belt can only measure relative changes in amplitude and should never be interpreted as absolute values of pulmonary volume or airflow. If absolute values are needed, it is recommended to use two belts (thoracic and abdominal) and calibrate them correctly [38]. Second, any breathing belt is susceptible to motion artefacts when the belt is displaced during acquisition. These artefacts can be seen in the signal as a drift in baseline amplitude. Although most cases can be resolved with a simple baseline correction method, the transient of the artefact can be more difficult to address correctly. Lastly, various natural breathing patterns occur that can make it difficult to automate the interpretation of individual breaths. These are typically present as sudden changes in respiratory volume or rate, "missed breaths" and/or deep sighs at irregular intervals. All of these patterns are natural and should be accounted for if possible [34, 39].

Figure 1.3 shows four practical examples demonstrating such variations. Note that the distinct peaks and valleys in the signals are used as indicators of exhale and inhale onsets respectively [35]. Segment (a) shows regular breathing at a stable respiratory rate, interrupted by deep sighs. According to Vlemincx et al., a high sigh rate is indicative of an increase in mental stress and should be treated as relevant information instead of unwanted artefacts [22]. Segment (b) illustrates a pattern

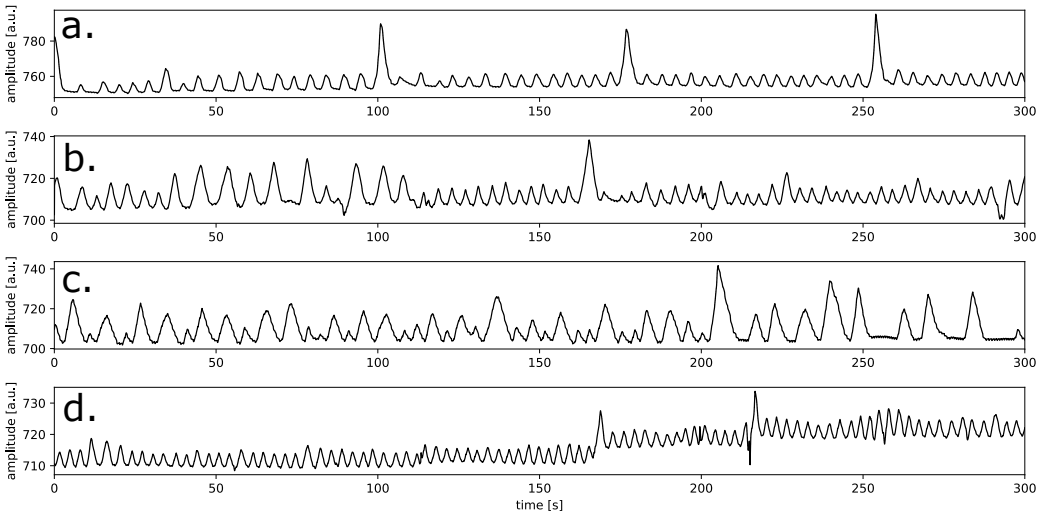

Figure 1.3: Morphological variations in the available RSP recordings: regular respiration, interrupted by single deep breaths/sighs (a), transition in breathing amplitude (b), erratic breathing with inconsistent timing and depth (c) and a signal with two motion artefacts (d).

of deep breathing transitioning into more superficial breathing, followed by a deep inhale before continuing with shallow breathing. Segment (c) shows more abrupt breaths of inconsistent depths, likely related to a state of elevated stress [39]. Finally, segment (d) is a good example of baseline drift due to a displacement of the breathing belt. Notice the local transient peaks where the baseline amplitude shifts, these artefacts can be difficult to correct.

1.3 Dataset specifications

The experimental data, collected by Daniels et al. for their work in [5] and [6], is analyzed further in this thesis. Note that although most elements are already reported in those studies, some additional information was later made available by Daniels to broaden the scope of this thesis. First, Section 1.3.1 discusses all important details of the experiment design. Section 1.3.2 then specifies how all information was collected. The resulting dataset contains the SAAM scores for each subject prior to the experiment, as well as recordings of blood volume pulse, electrodermal activity and thoracic respiration for each subject at each phase of the experiment over a duration of five minutes.

Phase	Condition	Time	Description
1	rest	pre	baseline activity in resting conditions
2	stress	pre	baseline activity under social stress
3	rest	post	post-treatment activity in resting conditions
4	stress	post	post-treatment activity under social stress

TABLE 1.1: Overview of recorded phases in the studied experiment

1.3.1 Experiment design

In context of the research discussed in Section 1.1, Daniels et al. recruited 56 healthy adult men to study the effects of exogenous OT on a NT reference group. All participants are right-handed and Dutch speaking, their mean age at the time of recording was approximately 22.2 ± 3.32 years. SAAM scores of each participant were collected prior to the experiment to provide insight in inter-personal variations of attachment style. This group was then divided into two groups by a random-selection program to create a double-blind, placebo-controlled study [5].

The experiment consisted of four phases in fixed order. Table 1.1 summarizes these phases with the same condensed descriptions as used for later test specifications (cfr. appendix B). In short, the phases switch between conditions of rest and induced social stress before and after placebo-controlled treatment. The key elements are illustrated in Figure 1.4.

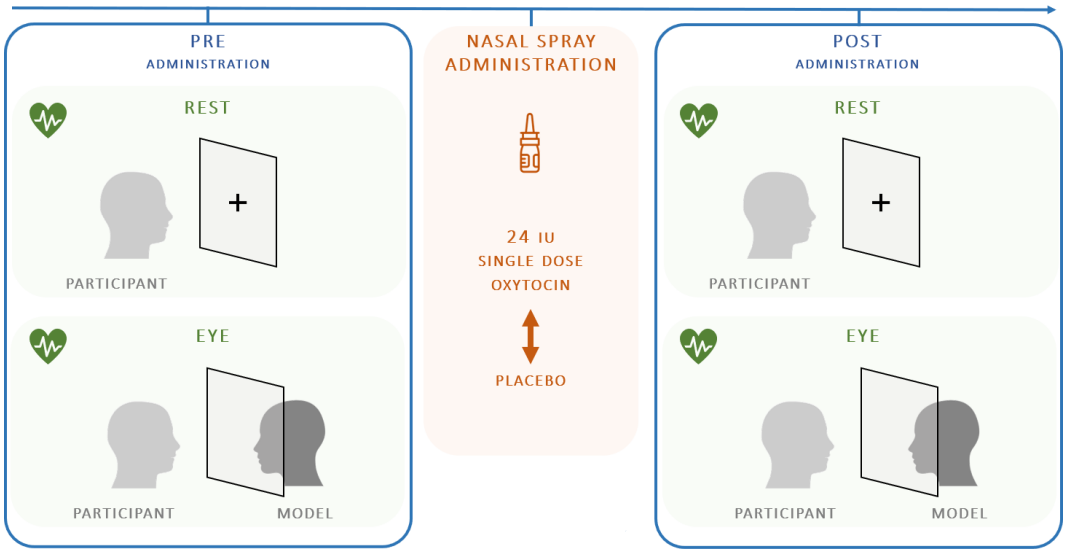

FIGURE 1.4: Experiment design: participants sit in front of a screen during four phases (cfr. Table 1.1) to create conditions of rest (screen closed) and social stress (eye-contact on screen openings at random intervals). Image courtesy of N. Daniels.

Signal	$f_s[Hz]$	Sensor type and body site	NeXuS-32 / BioTrace+ settings
BVP	128	fingertip PPG (ring finger)	default, raw export, notch filter OFF
EDA	128	fingertip SC/GSR (middle & index)	default, raw export
RSP	128	thoracic breathing belt (hall sensor)	default, raw export

Table 1.2: Acquisition specifications as used for MindMedia acquisition devices

For every phase, each individual participant was seated in front of a table with a closed screen. In the resting condition, the subject was asked to fixate their gaze on a cross marker on the screen for a duration of at least five minutes. In the "social stress condition", a model was seated on the opposite side of the screen so that only their face is visible when it opens. This screen was then set to open at random intervals of minimum 20 seconds for a five seconds at each screen opening over a duration of at least five minutes. The subject was asked to maintain eye-contact with the model on screen openings and focus their gaze on the cross for the remaining duration. As such, the social stimuli should be as consistent as possible across all recordings.

After recording the first two phases without treatment, 28 randomly selected subjects received OT as a single dose of 24 IU (3 sniffs of 4 IU in each nostril). The remaining 28 participants received a placebo spray (saline natriumchloride solution) to form a valid control group. Phases three and four were then executed identically to one and two respectively.

1.3.2 Signal acquisition

Three signals were recorded for each participant at each phase of the experiment: blood volume pressure (BVP), electrodermal activity (EDA) and thoracic respiration (RSP). As introduced in Section 1.2, these signals should provide relevant insights into the response mechanisms of interest. The relevant acquisition specifications are summarized in Table 1.2.

The physical placement of the required sensors is relatively straightforward. A fingertip PPG sensor was placed on the ring finger of the subject's non-dominant hand to measure the BVP $[a.u.]$. Galvanic skin response (SC/GSR) sensors on the index and middle fingers of the same hand measured EDA $[\mu S]$. Lastly, a breathing belt around the subject's chest measures the circumference with a Hall effect sensor to provide an estimate of the breathing volume $[a.u.]$.

The Nexus-32 multimodal acquisition system from MindMedia[2,3] was used to collect all signals from the applied sensors with minimal preprocessing. This raw information was then exported to `.txt` files with BioTrace+ software (version 2015a, MindMedia) for further analysis. In case of double/retake measurements, only the first file was used to maintain consistency. Further processing steps are documented in the next chapters.

[2] https://www.mindmedia.com/en/products/nexus-32/
[3] https://www.mindmedia.com/en/documentation/

1.4 Problem statement

1.4.1 Research goals

The end goal of this thesis is to contribute to a better understanding of social stress regulation mechanisms under influence of single-dose OT administration in a population of healthy, NT adult men. Additionally, these findings could provide useful reference metrics for similar studies with neurodiverse individuals.

Section 1.1 discussed recent studies by Prinsen and Daniels on the analysis of single-dose intranasal OT administration in adult NT men. Inter-subject variations in SAAM attachment scores (a.o.) were found to correlate with OT-specific treatment effects. Since these interactions provide important insights in the benefits and limitations of OT treatment, they will be assessed in this thesis as well.

The primary research goals are to confirm whether the reported anxiolytic and prosocial effects of OT can also be found in this dataset. And if so, to quantify and compare the observed effects across the available biomarkers. Once these effects are known, a secondary assessment is performed to test the influence of person-specific attachment scores. All relevant assumptions are verified in order to validate the obtained results.

1.4.2 Research questions

The above research goals are broken down into four actionable research questions: Which features can be used to measure the targeted response mechanisms? Is the studied data valid, and are the extracted features reliable? Do these features measure any OT-specific treatment effects? And lastly, are there any context-dependencies, such as attachment style, influencing these effects? The following paragraphs elaborate on these questions.

Which features capture the targeted effects? As indicated in the previous section, the dataset under analysis provides raw recordings for three noninvasive biomarkers. Although methods exist to assess such time series directly, it is more effective to extract a set of well-considered features from each recording.

Is the extracted data valid? Before applying any further tests it is important to verify whether the data under analysis can provide reliable information. A general assessment of relevant assumptions should be complemented by an inspection for outliers and missing data. Lastly, the size and distribution of the sample should be taken into account when selecting the appropriate analysis methods.

Any significant treatment effects? Statistical analysis of the extracted features will provide insight in the effects of OT treatment. To do so, the pre- to post-treatment changes are compared between groups for activity in rest and stress. Changes in activity from rest to stress are referred to as (social) stress response. They provide a baseline-corrected measure of stress activity, to account for inter-subject variations

in resting activity. The placebo group can then be used to differentiate treatment effects from any shared effects.

Influence of attachment style? Correlation tests with the subjects' SAAM attachment scores will provide insight in context-dependencies of social stress regulation and OT treatment. First, are there any interactions with pre-treatment observations? And second, are there any interactions with the observed treatment effects? The placebo group serves as a statistical reference to distinguish OT-specific interactions from other external factors (if any).

1.4.3 Approach

As illustrated in Figure 1.5, a systematic approach is used to address these questions. The signal processing and feature extraction and feature analysis are first documented in terms of methods and design decisions, in chapters 2 and 3. Their intermediate results are reported in Section 5.1. After introducing the appropriate statistical methods in Section 4.1, the feature analysis approach is then specified in Section 4.2. The results of this analysis are collected in Section 5.2.

The interpretation and discussion of all results can be found in chapter 6: first, Section 6.1 interprets the main findings and formulates answers to the research questions. The findings for this specific dataset are then compared to reference literature in Section 6.2. The strengths and limitations of the documented approach are briefly discussed in Section 6.3. Chapter 7 finally concludes the thesis with a summary of findings and suggestions for future work.

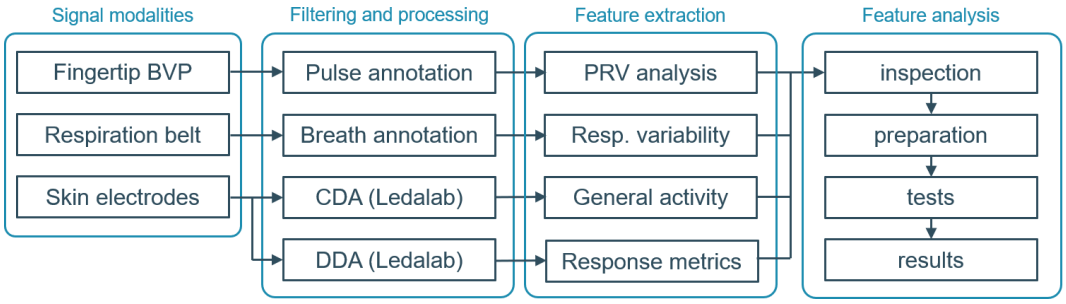

Figure 1.5: Systematic overview of the applied signal analysis

Chapter 2

Signal processing methods

This chapter discusses the processing steps applied to the dataset, for the BVP, EDA and RSP signals introduced in chapter 1. First, each signal modality is filtered to remove any unwanted spectral contents. After this, the algorithms to either annotate reference points on the signal or separate the signal into several functional components are applied. The resulting information can then be used for feature extraction. In order to standardize the extracted features, only the first 300 seconds of each signal are used.

2.1 Blood volume pressure

2.1.1 Preprocessing

The original BVP signals are sampled at 128 Hz, which is sufficient time resolution for pulse rate variability analysis (cfr. Section 3.1) according to the standards published by Malik et al. [40]. Visual inspection of the data shows no significant background noise in most signals, although a small subset is too corrupted for reliable use. These are removed from further analysis. Besides this, no filters[1] or resampling are applied

Although the available BVP data is generally of good quality, certain recordings contain significant morphological variations as introduced in Section 1.2.2. These are often explained by the influences of sensor placement, body posture and person-specific factors related to physiology, stress levels or even specific pathologies [27, 29, 28]. To obtain consistent results for all waveform variations in the studied dataset, it is important to select the appropriate annotation method.

2.1.2 Signal annotations

The annotation of reference points in the BVP signals is needed for accurate estimation of the features used in this thesis. First, various peak detection methods can be used

[1] Although a bandpass filter of 0.03 to 40 Hz is typically applied to PPG signals [41], this was left out in practice since it showed limited effect during visual inspection. In hindsight, it remains good practice to apply this filter. No harm is done to the signal contents of interest, while any baseline variations or HF noise are removed without the need to rely on visual inspection of all signals.

to localize the peaks of BVP pulses. This already provides sufficient information for basic feature extraction. Afterwards, other peak-related fiducial points can be considered to refine the quantification of BVP variability.

PPG pulse detection

Initially, a PPG pulse detection algorithm proposed by Lazaro et al. was used to annotate the BVP data for this thesis [41]. Lazaro's MATLAB implementation of this algorithm is available in STADIUS repositories. Their algorithm first uses a linear filtering transformation to enhance the targeted PPG pulses, after which an adaptive thresholding method is used to detect the enhanced peaks. For the data studied in this thesis, however, their method is unable to reliably differentiate the target pulse from the secondary peaks caused by pronounced dicrotic notches (cfr. Section 1.2.2). Despite many segments being annotated as intended, an impractical amount of manual corrections would be needed to obtain the desired results.

Various alternatives can be found in literature on R-peak detection in electrocardiography (ECG). Given that both signal modalities measure cardiac activity, the main difference between them lies in morphology of the pulse complex. If this difference in waveform is accounted for by selection of the appropriate enhancement technique, the same two-step methods (QRS/BVP enhancement, followed by peak detection) can be applied. Elgendi et al. published a comprehensive comparison of popular methods that follow this design approach [42]. After consideration of these options, the algorithm included in the R-DECO package was found to provide the desired performance.

Alternative approach using R-DECO

R-DECO is an open-source toolkit for robust R-peak detection in ECG data [43, 44]. The Matlab package was developed by J. Moeyersons et al. and is publicly available on Physionet.[2] A screenshot of the GUI is shown in Figure 2.1, where it can be seen that it provides a user-friendly peak detection tool, with additional options for filtering and (semi)automated or manual correction of peak annotations.

The main peak detection algorithm in R-DECO is designed to reliably annotate R-peaks in ECG data [43, 44]. Commonly used methods for this application use filters to enhance the amplitude of the R-peaks and attenuate other waves. In contrast, the algorithm used in R-deco uses the lower and upper envelopes for this task. After detecting those enhanced R-peaks, a composite peak detection step with adaptive thresholding is used to further refine the selection of valid peaks. This method was first proposed by Varon et al. in [45].

For the BVP data studied in this thesis, the R-DECO tool was found to achieve much more consistent pulse annotations despite the challenging waveform in some signals. Similar to its function in ECG, the envelope-based pulse enhancement suppresses the problematic dicrotic notches in the BVP signals as if they were P-waves following the QRS-complex of ECG. In BVP signals with a "normal" dicrotic

[2] https://physionet.org/content/r-deco/1.0.0/

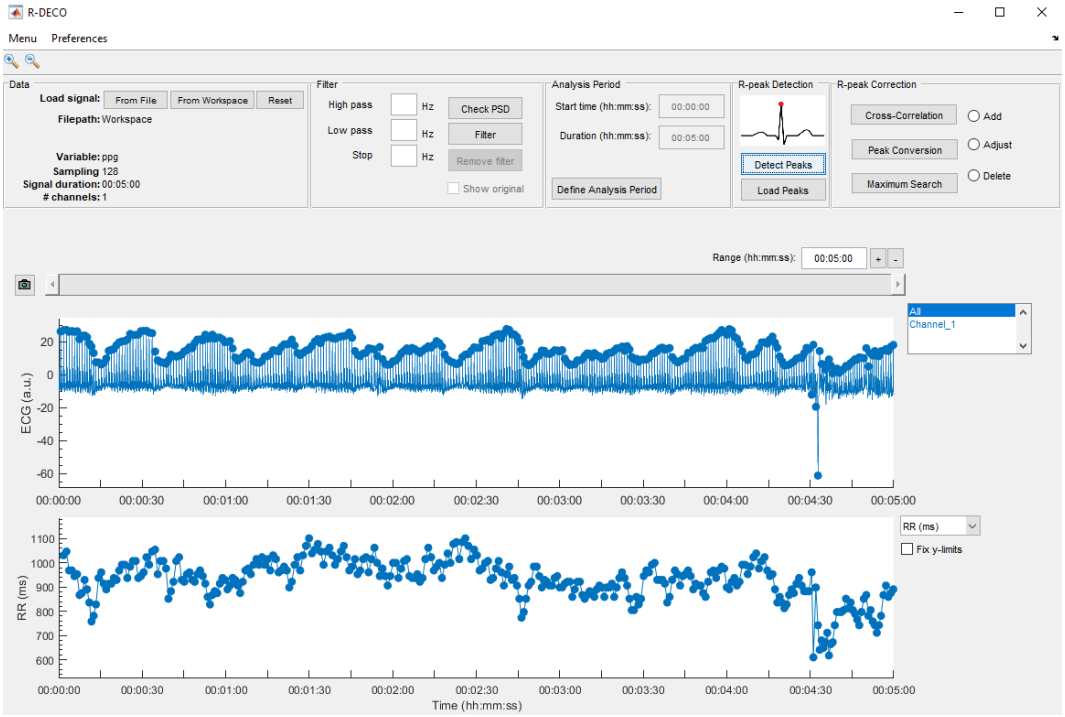

FIGURE 2.1: Graphical user interface of R-DECO with BVP data

notch, the detection algorithm in R-DECO achieves similar performance as the previous algorithm suggested by Lazaro et al. As such, R-DECO is chosen for further processing to achieve a significant reduction in false pulse detections.

Additionally, the built-in add/adjust/delete buttons in R-DECO were used to apply manual corrections to the automatic pulse annotations. Despite the good detection performance, a small number of signals contain motion artefacts or excessive background noise such that automated annotation becomes unreliable. Manual interpolation of short motion artefacts was used to reduce errors in extracted features, while larger noisy segments were removed from further analysis.

Robust fiducial points

Before moving on to the feature extraction, it is important to consider the accuracy of the obtained BVP peaks. As is often the case with peak detection, small noise-related variations in signal amplitude can lead to disproportionate errors in temporal accuracy of the estimated pulse timestamps. The selection of a more robust fiducial point can help to reduce error accumulation in any features derived from these peaks, contributing to more reliable findings in later analysis.

Besides using the BVP peaks (also referred to as apex points, n_A) directly, there are two commonly used fiducial points that are suitable for the purposes of this thesis. Following the definitions of Lazaro et al. in [41], the base point (n_B) represents the

minimum BVP amplitude before onset of the next pulse. The midpoint (n_M) is then defined as the BVP sample closest to 50% amplitude of the upslope between n_B and n_A. Both can be derived automatically from a BVP signal given clean peak annotations n_A. Their implementation is best explained by equations 2.1 and 2.2.

$$n_B = \underset{n \in [n_A - T, n_A]}{\operatorname{argmin}} \; x_{BVP}(n) \quad s.t. \quad T = 150ms \tag{2.1}$$

$$n_M = \underset{n \in [n_B, n_A]}{\operatorname{argmin}} \left| x_{BVP}(n) - \frac{x_{BVP}(n_A) + x_{BVP}(n_B)}{2} \right| \tag{2.2}$$

Where $x_{BVP}(n)$ represents the BVP signal sampled at 128 Hz, such that $n \in [n_A - T, n_A]$ lists the samples in a 150ms segment before each apex point (n_A). Using the same definitions, Peralta et al. tested the performance of these fiducial points for the analysis of pulse rate variability (PRV, cfr. Section 3.1). According to them, the midpoints are typically more robust than apex or base points because of their position on the steep upslope of the BVP pulse [46]. For this reason, n_B is chosen as fiducial point for further feature extraction.

Note that the search windows for base points n_B were reduced to 150ms in order to avoid false annotation of the deep dicrotic notches in some segments of the studied dataset. Although Lazaro and Peralta suggest windows of 250-300ms [41, 46], this modification was found to reduce the error rate in base points without further issues. By extension, this improves the quality of the corresponding midpoints. Figure 2.2 illustrates the results for a short BVP segment.

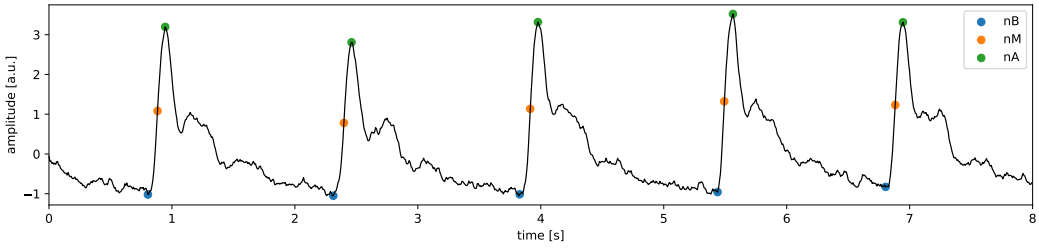

Figure 2.2: Fiducial point annotations obtained for BVP/PPG data: apex points (nA) as found by pulse detection, base points (nB, local minimum in 150ms segment before each nA) and midpoints (nM, 50% amplitude between nB and nA).

2.2 Electrodermal activity

2.2.1 Preprocessing

Section 1.3 specified the EDA data provided by N. Daniels. By visual inspection, the majority of these signals appear to be of sufficient quality to be used with minimal intervention. Later inspection of the extracted features, cfr. Section 5.1.3, confirmed

this assumption. As suggested by Benedek et al. in [1, 2], no filters[3] or amplitude standardization are applied to maintain the original values (in μS). Finally, all signals are downsampled from 128 Hz to 32 Hz to reduce computational load in the decomposition methods described below. This sample rate is still well above the minimum of 10 Hz recommended in the Ledalab documentation (cfr. below).

2.2.2 Signal analysis with Ledalab

As introduced in Section 1.2.3, EDA signals consist of a tonic and phasic component which are both related to (social) stress regulation in the ANS [23, 25, 26]. With the appropriate methods, these components can be separated from each other for more refined feature extraction. The simplest method is to detect local extrema in the signal to identify individual skin conductance responses (SCRs) and describe their relative positions in time and amplitude. However, this trough-to-peak (TTP) method is very prone to errors when subsequent SCRs are too close together [47].

To address this issue, M. Benedek and C. Kaernbach developed two convolution-based methods: continuous (CDA) and discrete (DDA) decomposition analysis. Both were proven to significantly improve the consistency and accuracy of EDA analysis compared to simple TTP methods [1, 2]. Their MATLAB package Ledalab, which provides both algorithms in a user-friendly format, is publicly available on Github. The following sections first explain the main concepts of these algorithms, and then discuss how each method is used to provide relevant insights for this thesis.

General concepts of EDA decomposition

Both algorithms available in Ledalab fundamentally follow the same approach. The common goal is to separate tonic from phasic activity given a continuous EDA signal (referred to as skin conductance $SC(t)$ in descriptions of the algorithms). Knowing that phasic activity in EDA is linked to discrete sweat gland activations by the SNS, the phasic signal ($SC_{phasic}(t)$) is modelled as the superposition of M consecutive SCRs, as shown in equation 2.3.

$$SC_{phasic}(t) \approx \sum_{m=1}^{M} (SCR_m(t)) \tag{2.3}$$

Each SCR is then interpreted as the response to a discrete impulse from the nervous system. Since the true impulse response is unknown in this theoretical model, the impulse response function $IRF(t)$ must be estimated to fit the signal. Deconvolution of $SC_{phasic}(t)$ with this $IRF(t)$ yields a phasic "driver" signal $\delta_{phasic}(t)$, such that distinct peaks in the phasic driver represent the activation pulses sent by the SNS to the sweat glands. A similar approach can be used for the tonic driver $\delta_{tonic}(t)$, to obtain equation 2.4 as proposed in [1].

[3] Note that the Ledalab package does provide preprocessing functions if needed, including low-pass filtering, down-sampling, cutting, smoothing, and artifact-correction. (`http://ledalab.de`)

$$\begin{aligned} SC(t) &= SC_{tonic}(t) + SC_{phasic}(t) \\ &= SC_{tonic}(t) + [\delta_{phasic}(t) * IRF(t)] + e_1(t) \\ &= [\delta_{tonic}(t) + \delta_{phasic}(t)] * IRF(t) + e_2(t) \end{aligned} \tag{2.4}$$

Note the residual errors $e_1(t)$ and $e_1(t)$ representing the difference between the reconstructed SC signal and the original. Depending on the specific algorithm this can be quasi-zero (machine resolution), but this is not always the case. This residual error can be used in combination with more advanced metrics to iteratively improve the initial estimation of $IRF(t)$. Benedek et al. model this IRF with a Bateman function with two parameters, τ_1 and τ_2, representing the "steepness of onset" and "recovery rate" respectively [1].

If this $IRF(t)$ estimate fits the signal well, this method behaves as a template matching filter. The resulting phasic driver will contain a distinct peak wherever the template ($IRF(t)$) matches a SCR in the original signal. Both CDA and DDA algorithms use this information from δ_{phasic} to identify and reconstruct (only in DDA) individual SCRs.

Starting from these general concepts, the difference between CDA and DDA is best explained by their design intentions. According to Benedek et al., the CDA algorithm is designed for accurate separation of phasic and tonic components as continuous signals [1], while the DDA algorithm focuses on the accurate identification of individual SCRs [2]. The next sections clarify how these goals are achieved.

Continuous decomposition analysis

The first method starts from the sudomotor nerve activation (SMNA) model to obtain continuous measures of phasic activity in skin conductance data. In rest, skin conductance is measured as a stable baseline of tonic activity. As response to nervous activation, in this case social stress triggers, the sudomotor nerves activate sweat glands [48, 49, 30]. The SCRs resulting from these activations are modelled as "discrete and compact phasic deflections", i.e. the recovery rate of the IRF is assumed to be high (small time constant τ_2) [1].

The block schematic in Figure 2.3 summarizes how this interpretation is translated to a functional decomposition algorithm. A visual reference from [1] can be found in Figure 2.4. First, deconvolution with an initial estimate of the IRF acts as a template matching filter on the EDA signal to obtain the driver signal $\delta_{SC}(t)$. Note that this deconvolution operation is referred to as "standard" in the diagram. Next, the tonic driver $\delta_{tonic}(t)$ is estimated by interpolation of low-activity segments in the smoothed driver signal. Using equation 2.4, the phasic driver $\delta_{phasic}(t)$ is then computed by subtraction of the tonic and SC driver signals. Finally, reconvolution with the IRF estimate returns the desired continuous signals for both components.

To find a good fit for the IRF estimate, Ledalab initializes $IRF_1(t)$ automatically, and then runs K iterations to improve the parametrization of $IRF_k(t)$ for a given signal. To evaluate the corresponding decomposition results on each iteration, a performance criterion c_1 is defined in equation 2.5.

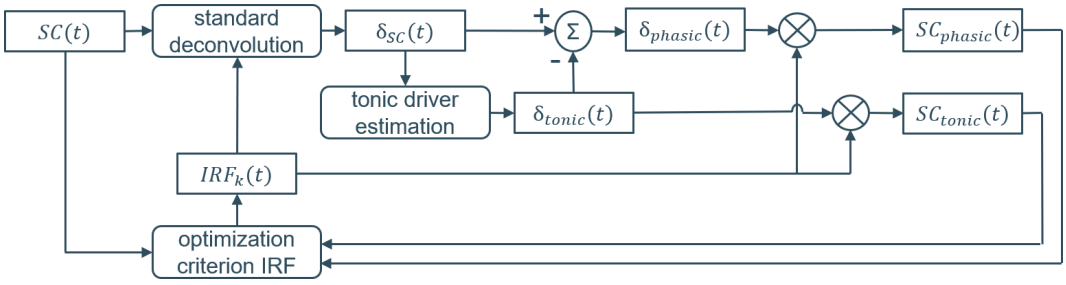

FIGURE 2.3: Block schematic of CDA algorithm based on Benedek et al. [1]

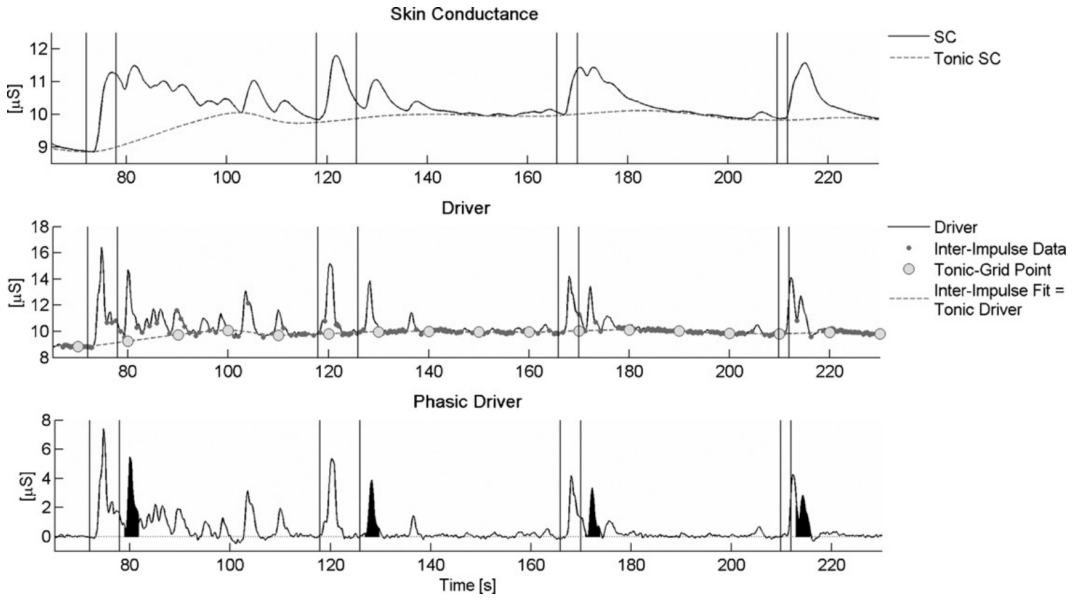

FIGURE 2.4: Illustration of CDA components. Figure from Benedek et al., 2010 [1]

$$c_1 = indist + w_1 \times neg \tag{2.5}$$

Where indist corresponds to the indistinctness criterion, neg to the negativity of the phasic driver, and w_1 is a fixed weight. The indistinctness criterion counts samples of the phasic driver above threshold[4] as a rough estimate of the duration of nonzero segments. Phasic deflections in CDA should be compact and distinct, so excessive durations should be minimized. The negativity of the phasic driver must be minimized to confirm the fit of IRF parametrization. This is quantified via RMSE of the negative samples of the phasic driver. A fixed weight w_1 is then used to combine both metrics into the minimization criterion c_1.

[4]Only phasic driver peaks above $.05\mu S$ are considered valid SCRs, cfr Section 5.1

Discrete decomposition analysis

The second method proposed by Benedek and Kaernbach is of similar design as CDA, but approaches the decomposition problem from a different physiological model in order to focus on accurate separation of individual SCRs. Instead of the SMNA interpretation, DDA is inspired by the poral valve model (PVM), as documented by Edelberg et al. [50]. They found that, when using IRF estimates with large τ_2 (slow recovery), EDA signals occasionally contain an additional small overshoot in certain SCRs that translates to a negative peak in the driver signal. These secondary components, which can be seen in Figure 2.5 (in brown), are interpreted as pore openings (PO). Based on these findings, the DDA algorithm uses a second deconvolution step to extract these PO components from the phasic signal. Note that DDA continues with a high τ_2 parameter for the estimation of "normal" SCRs [2].

Figures 2.5 and 2.6 provide a summarized overview of this algorithm. The initial use of a deconvolution to estimate the tonic driver is similar to CDA, but this time the interpolation points are set much further apart, resulting in a near-constant tonic baseline. The tonic and phasic signals are then reconstructed identically to CDA, after which nonnegative deconvolution (as defined in [2]) is applied to the phasic signal. In summary, the "normal" SCRs are identified as peaks in the obtained phasic driver, while peaks in the additional nonnegative component are annotated separately as PO overshoots. Final reconstruction is then done by adding all components.

Lastly, a different minimization criterion c_2 is used to iteratively improve the IRF estimate. DDA uses three subcriteria: (1) The reconstruction error is indicated with the RMSE of the difference between the original SC data and the reconstruction from phasic and tonic drivers. Due to the nonnegative deconvolution method this can deviate from zero, while CDA should always reconstruct perfectly. (2) The number of phasic responses (m_{ph}) is included, since excessive segmentation is undesirable for consistent results. (3) The duration of SCR impulses in samples is used to construct an indicator for discreteness (d) of the results. These three are then combined to evaluate equation 2.6 on each iteration of $IRF_k(t)$.

$$c_2 = RMSE \times (m_{ph} + 1) \times d \tag{2.6}$$

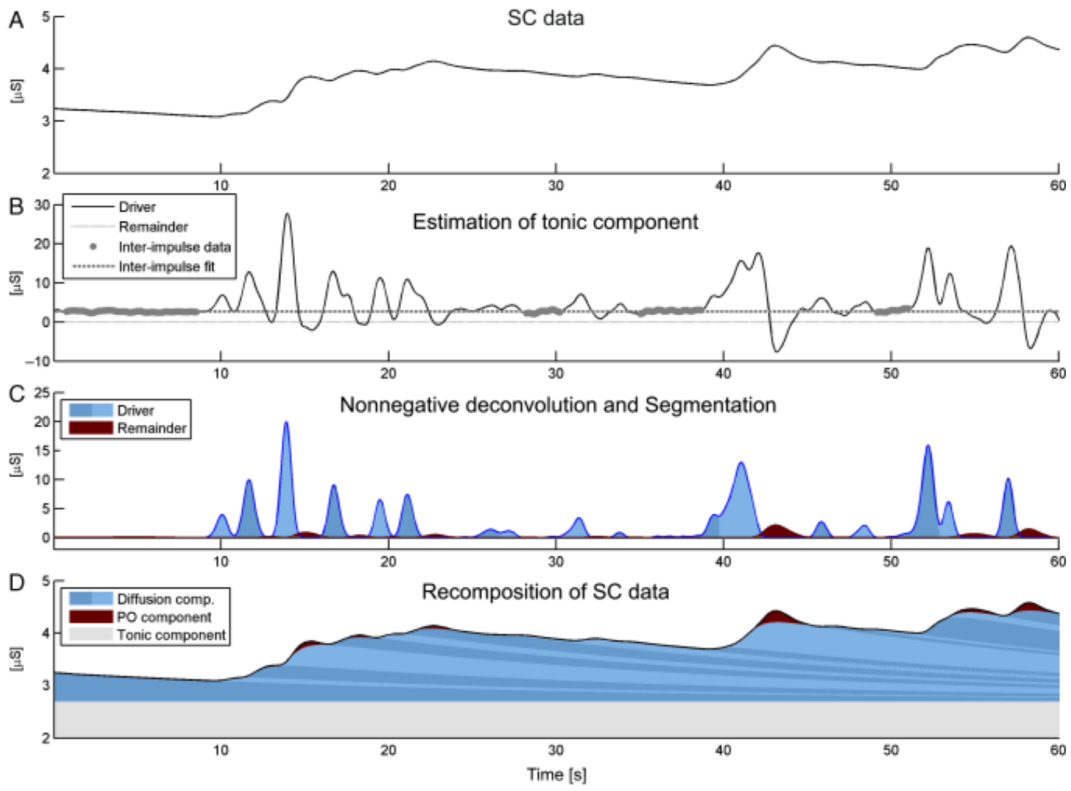

FIGURE 2.5: Illustration of DDA components. Figure from Benedek et al., 2010 [2]

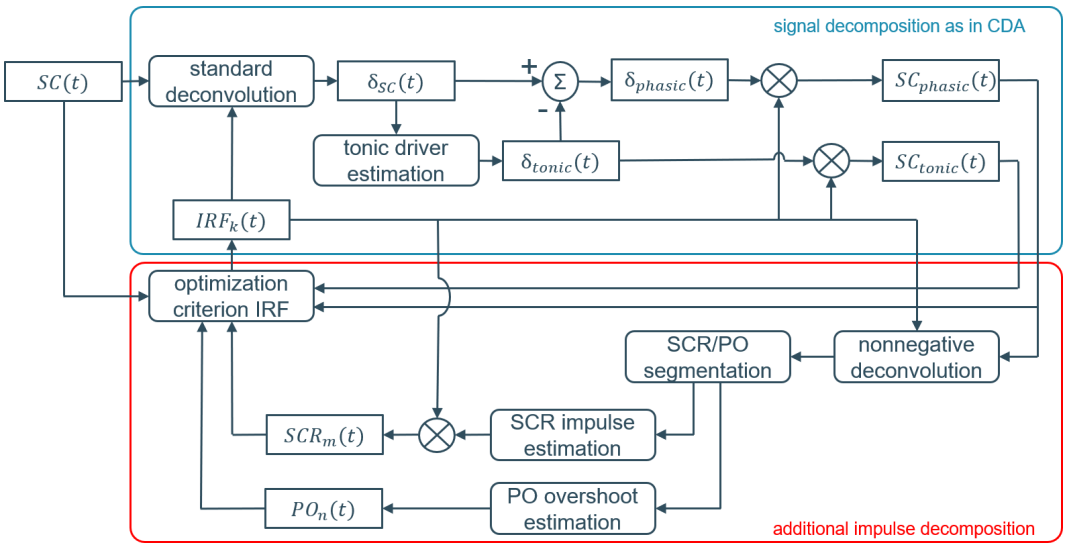

FIGURE 2.6: Block schematic of DDA algorithm based on Benedek et al. [2]

2.3 Respiration

2.3.1 Preprocessing

All RSP signals were bandpass filtered before further analysis, using a zero-phase third order Butterworth filter. To select the cutoff frequencies, the characteristics of the respiratory signals in the dataset were visually inspected. By rough estimation, the breathing rate was found to range between 5-25 breaths per minute. These values translate to 0.083 - 0.416 Hz respectively. To account for the characteristics of the respiration waveform, the cutoff frequencies are set at 0.08 - 0.6 Hz instead. Finally, the signals were not standardized to avoid biasing the amplitude features between observations (cfr. relative amplitude of single-belt RIP signals, in section 1.2.4).

2.3.2 Breath annotations

Inivididual breaths in the RSP signals are annotated to allow later feature extraction. Generally, the inhale and exhale onsets of each breath are represented by distinct valleys and peaks in the RSP signal. However, this distinction can be hard to make in many cases. Small local fluctuations, as well as the physiological patterns shown earlier in figure 1.3 are not always considered "valid" breaths. The challenge in this section is to obtain annotations for all RSP signals that are consistent with the physiological interpretation introduced in section 1.2.4 [34, 35, 51, 39].

Three different approaches were considered in this thesis: first, the peak detection algorithms included in R-DECO (cfr. section 2.1) were found to be unsuitable for respiratory signals. Although it still offers a good tool for manual corrections, alterations of the underlying algorithms were not explored further. Next, the BreathMetrics[5] package published by Noto et al. [51] was also found to return inconsistent results for breathing belt data. Finally, a robust adaptive thresholding method was implemented to annotate peaks based on their prominence value. Compared to the previous alternatives, this simple algorithm is much more robust to variations in peak amplitude, while still providing good control over the inclusion threshold for shallow breaths. The following section explains this last approach in more detail.

Robust peak prominence thresholding

The MATLAB function `findpeaks()` provides a simple tool to compute the prominence of all peaks in a given time signal. First, all local maxima are calculated by comparing the amplitude of consecutive samples. Then, prominence of each peak is calculated based on relative location and amplitude between peaks in the signal. The main advantage of this metric is that it is is independent of absolute amplitudes, such that a prominence threshold can be defined to consistently reject low-prominence peaks across signals with different amplitude distributions.

After obtaining an initial set of local maxima, these are pruned so that only the peaks remain which correspond to "true" breaths. Note, however, that the validation

[5] https://github.com/zelanolab/breathmetrics

of results is based on visual inspection, since no reliable ground truth was available for this dataset (cfr. section 5.1.4). For respiratory signals with low variability across their full duration, a fixed threshold on minimal peak prominence worked relatively well. However, if any of the large amplitude variations shown in figure 1.3 are present, the detection parameters need to be adjusted to that specific signal.

A simple but effective way to account for these amplitude variations is to apply adaptive thresholding, such that the minimum peak prominence is adjusted automatically in function of the variability of each signal. The quantile range (QR) of a respiratory signal provides a robust estimate of the amplitude range of the majority of its samples. By defining two parameters, "robustness" $r \in [0, 0.5]$ and a prominence factor $f \in [0, 1]$, equation 2.7 defines a robust adaptive threshold: it is less susceptible to amplitude outliers in a given signal, while maintaining the ability to finetune the inclusion of shallow breaths. Note that e.g. $r = .10$ represents the 10% quantile range QR.

$$MinPeakProminence = f \times QR(rsp, [r, 1 - r]) \tag{2.7}$$

The combination of these two parameters provides good control over the robustness to amplitude variability (reduce r to ignore more outliers) and inclusion of shallow breaths (reduce f to include smaller peaks). With this method, the RSP signals can be annotated in batch by selecting the appropriate values for r and f. The peak prominence threshold then automatically adjusts for each signal by evaluating equation 2.7 with r and f fixed.

Using a small set of control signals for validation of the results (visual interpretation), this approach was found to perform well for a 30% quantile range ($r = 0.30$) in combination with a prominence factor $f = 0.40$. Finally, the inhale onsets (valleys) corresponding to the annotated exhale onset (peaks) are found as the minimum between consecutive peaks.

2.4 Conclusion

This chapter discussed all signal processing steps for BVP, EDA and RSP signals required for feature extraction. For each BVP signal, the peaks were annotated with the R-DECO package in order to derive reliable midpoints n_M. The EDA signals were deconstructed into tonic and phasic components using the CDA and DDA algorithms by Benedek et al. Finally, annotations of inhale and exhale onsets in the RSP data were obtained based on adaptive peak prominence thresholding after consideration of two alternative approaches (using R-DECO and BreathMetrics).

Certain physiological variations in the observed signals required specific attention to process correctly: the pronounced dicrotic notches in many BVP signals, sparse cases of non- and hyperresponders as well as signals with low recovery rates in the EDA data, and the natural breathing patterns observed in the RSP signals. Although these exceptions remain challenging, the selected methods were found to be consistent for the majority of signals in the studied dataset.

To assess the validity of features derived from these processed signals, Section 5.1 reports the relevant intermediate results. The discussion of the strengths and limitations of these methods follows later as part of chapter 6. Feature extraction methods for each signal modality are discussed in the next chapter.

Chapter 3

Feature extraction methods

This chapter describes the feature extraction for all three signal modalities. The filtered and processed signals, as discussed in chapter 2, provide fiducial points for BVP and RSP data, while the EDA signals were deconstructed into phasic, tonic and SCR components. The aim in the following sections is to select and implement a feature set that is able to capture the targeted regulation mechanisms. For BVP, Section 3.1 focuses on pulse rate variability. Section 3.2 discusses the features derived from CDA- and DDA-based signal components. Finally, Section 3.3 discusses features of respiratory variability.

3.1 Blood volume pulse

The methods described in Section 2.1 provide clean BVP data with corresponding annotations for apex, base, and midpoints. These can be used for the extraction of a wide range of cardiovascular features. The feature set included in this thesis is documented below.

3.1.1 Direct BVP features

The first feature of interest is pulse rate. A simple instantaneous metric provides insight in pulse rate variability between subjects and conditions. The pulse rate is calculated in beats per minute using equation 3.1, for each pair of consecutive midpoints n_M (where R is the total number of fiducial points annotated in a given signal).

$$PR[r] = \frac{60}{n_M[r+1] - n_M[r]} \quad \forall r \in [1, R-1] \tag{3.1}$$

Statistical metrics of these pulse rate series quantify basic cardiac activity: average pulse rate is simply the mean of $PR[r]$ for the duration of the observed signal. Additionally, standard deviation and mean absolute deviation of the pulse rate provide basic metrics for the variability of a subject's pulse rate [29].

3.1.2 Classical PRV analysis

To extract more refined features from this instantaneous pulse rate $PR[r]$, it is converted to a pulse rate variability (PRV) signal in accordance with HRV Task Force standards [40]. These standards provide a practical reference for the implementation of "classical" PRV analysis as described in the next paragraphs. Although more advanced methods have been published after them, the focus in this thesis remains on this classical approach.

The features in this approach are based on power spectral density (PSD) of the PRV, using two standard frequency ranges: the low frequency band (LF) in 0.04-0.15 Hz, and the high frequency band (HF) in 0.15-0.40 Hz. Both bands are typically found to contain a spectral peak, although exceptions can occur where these peaks leave the standard frequency ranges. A shift in (relative) power from the HF band to the LF band is interpreted as a shift from parasympathetic to sympathetic tone in cardiac activity, and vice versa [40].

The construction of the required PRV signals, given an annotated BVP signal, is implemented in four steps: (1) Pulse intervals are calculated as $60/PR[r]$, using the pulse rate from equation 3.1. (2) These pulse interval values are interpolated and resampled to $f_s = 4Hz$. This yields a unfiltered PRV signal at regular sample distances (while the pulse intervals from the previous step corresponded to the fiducial points at irregular timestamps). (3) This unfiltered PRV signal typically contains excessive amplitude deviations near both ends of the signal, due to the interpolation step. To prevent these artefacts from causing spectral leakage in further steps, the ends of the PRV signal are flattened before the first and after the last fiducial point respectively. (4) Finally, a bandpass filter is applied to enhance the spectral content of interest. In this study, a third-order Butterworth filter is applied with cutoff frequencies at 0.03 Hz and 1.00 Hz. These cutoffs remove power contents outside the range of interest (cfr. normalization step below), without affecting the LF and HF bands.

The power spectral density for each PRV signal is then calculated using the Welch method, with 40s sliding windows and 50% overlap. This results in a smooth PSD from which the LF and HF features are extracted by numerical integration of the power bands. Using the total power (TOT) in 0.0 to 1.0 Hz., the normalized power features are calculated as LFn = LF/TOT and HFn = HF/TOT. Finally, the ratio of LF to HF power is calculated to create a combined feature LH_ratio. Table 3.1 summarizes these six PRV features [40, 52].

3.2 Electrodermal activity

Both decomposition methods (CDA and DDA, cfr. Section 2.2) are applied to the EDA dataset with the batch-mode processing tool provided by Ledalab. Features of interest are extracted from these decomposition results based on prior literature, as described in this section.

Abbreviation	band (Hz)	Description
LF	$0.04 - 0.15$	Power in low frequency range
LFn	$0.04 - 0.15$	Normalized power in low frequency range
HF	$0.15 - 0.40$	Power in low frequency range
HFn	$0.15 - 0.40$	Normalized power in low frequency range
TOT	$0.00 - 1.00$	Total power in spectral range of interest
LH_ratio	n.a.	Low/High power ratio

TABLE 3.1: Features extracted with classical PRV analysis

3.2.1 Prior literature

N. Daniels et al. already published early indications of significant ANS response in the EDA data used for this thesis [5]. They compared pre- to post-treatment differences in response to eye-contact for OT and placebo, with a focus on the amplitude (Amp) and half-recovery time (R50_time) of SCR impulses. Although the observed changes in amplitude were not treatment-specific, the half-recovery time did show significant OT-specific treatment effects. Additionally, this treatment effect was found to be more prominent in subjects with lower self-reported social responsiveness scores (SRS).

Starting from these exploratory findings, the goal in this section is to construct a feature set which provides more detailed insights in the underlying regulation mechanisms. Various studies were consulted to support this selection process: Benedek et al. discuss general measures of tonic and phasic activity in [1], as well as more refined features of variability in the response time, amplitude and recovery in SCRs and PO components in [2]. These features were found to cover the majority of features used in other studies related to mental or social stress. [21, 23, 29, 24, 25, 26]

3.2.2 Implemented features

Based on the mentioned sources, the following paragraphs explain the implementation and interpretation of each feature used in this thesis. This feature set fits roughly into two categories: measures of electrodermal activity for the signal as a whole, resulting in single value features directly, and measures of individual skin conductance responses in the signal. The second category results in arrays of M values (one per identified SCR), which are then collected into a smaller set of statistical aggregates to obtain single-value features of variability (cfr. last paragraph in this section).

Signal activity Grundlehner et al. suggest the "energy of the second difference signal", i.e. the signal power of the second derivative of the EDA signal (D2SC_power), to obtain a general measure of electrodermal activity [21]. To complement this feature with a baseline-corrected value, the power of the estimated phasic signal (PhSC_power) is calculated as well [26]. Both are implemented in MATLAB with

the `bandpower()` function. Additionally, the linear regression slope (SC_slope) of the full-length signal was added to provide a measure of long-term recovery [23, 26].

SCR features (both methods) More refined information can be obtained by quantifying the properties of each SCR. For both decomposition methods, the SCR onsets are used to calculate the time intervals between consecutive SCRs (inter-response intervals, IRI) and the peak amplitudes (Amp) of each SCR [23, 26, 2].

SCR features (DDA only) Since DDA provides full reconstructions of each SCR, this method is used to calculate additional SCR shape features. The area-under-curve (Area) of the SCR provides a mixed measure of recovery time and amplitude. For each $SCR_k(t)$, the Area is calculated as $\int SCR_k(t)$ by numerical integration. Other measures of recovery time are obtained through the half-recovery time of the SCR (R50_time), the SCR width at percentile 50 of peak height (P50_width), and the linear regression (R50_slope) and interpolation (P50_slope) slopes between peak amplitude and 50% amplitude of each SCR [2, 26, 25, 5].

The resulting feature set is summarized in Table 3.2. Note that the SCR features are calculated for each individual SCR. To aggregate this information into single value features, the statistics listed in Table 3.3 were calculated per feature for each full-length EDA signal. Finally, for the Amp and the Area, the values per subject were summed up to create an additional measure of phasic activity. In total, 36 features are calculated for each EDA signal.

Method(s)	Abbreviation	Feature description
CDA,DDA	PhSC_power	power of phasic SC signal
CDA,DDA	D2SC_power	power of second derivative of full SC signal
CDA,DDA	SC_slope	linear regression slope of full SC signal
CDA,DDA	Amp	peak amplitude of SCR
CDA,DDA	IRI	inter-response interval for subsequent SCRs
DDA	Area	numerical integration of impulse reconstructions
DDA	R50_time	half-recovery time of discrete impulses
DDA	P50_width	impulse width at 50% of peak amplitude
DDA	R50_slope	regression slope of impulse recovery
DDA	P50_slope	interpolation slope of impulse recovery

Table 3.2: Features extracted from decomposition results

3.3 Respiration

Prior works regarding the effects of mental stress use interesting features of respiratory variability to asses the ANS response to stressors. This section discusses the clinical relevance and practical implementation of the selected features, using the RSP signals and their breath annotations as obtained in the previous chapter.

	Method(s)		Statistics			
Abbreviation	DDA	CDA	Avg	Std	MAD	Sum
Amp	x	x	x	x	x	x
IRI	x	x	x	x	x	
Area	x		x	x	x	x
R50_time	x		x	x	x	
P50_width	x		x	x	x	
R50_slope	x		x	x	x	
P50_slope	x		x	x	x	

TABLE 3.3: Statistical features extracted from SCR series

3.3.1 Prior literature

Vlemincx et al. studied the effects of emotional stress in respiratory signals, acquired with respiratory inductive plethysmography (RIP) [22]. In their study, however, two belts (thoracic and abdominal) were used instead of one. If the respiratory signals are calibrated correctly before each recording, the summation of signal amplitudes of these two belts provides a good estimation of tidal volume [53, 38]. The data in this thesis, however, was recorded with only one breathing belt. Any amplitude-based features are only reliable for variations within the same recording, and their absolute values should be compared between signals directly. Because of this, any amplitude features discussed below are implemented with a focus on respiratory variability across consecutive breaths.

Despite these limitations, tidal volume (Vt) and minute volume (MV) remain interesting amplitude-based features. The tidal volume feature estimates the volume of air displaced by individual breaths, while the minute volume estimates the average air displacement per minute. Although the RSP signals in this thesis only provide a relative measure of chest displacement, the variability of these features across the signal are still expected to provide relevant insights. Other studies such as the ones by Wientjes et al. support this expectation [34, 22].

Additional temporal features in [22] are based on the prior work by Wilhelm, Sackner and Roth. Their measures of inhale time (Ti), exhale time (Te) and respiratory rate (RR) are directly applicable to the breath annotations obtained in Section 2.3 [35].

3.3.2 Implemented features

To make full use of the RSP signals, these five metrics are calculated for each valley-to-peak (VTP) pair in the breath annotations. Their implementations are listed in Table 3.4. Knowing that the VTP pairs represent inhale and exhale onsets (cfr. Section 1.2.4), the implementation of the temporal features is very straightforward. The tidal volume is approximated as the amplitude difference from inhale to exhale onset in VTP pairs as suggested in [32].

Feature	Implementation (per breath b)	Feature description
Ti	$t_{exh}(b) - t_{inh}(b)$	Inhale duration
Te	$t_{inh}(b+1) - t_{exh}(b)$	Exhale duration
RR	$60/(T_i(b) + T_e(b))$	Respiratory rate
Vt	$x_{rsp}(t_{exh}(b)) - x_{rsp}(t_{inh}(b))$	Tidal volume (relative)
MV	$RR(b) \cdot V_t(b)$	Minute volume (relative)

Table 3.4: Quantification of discrete breaths in respiratory data

With these values, the final features are calculated as statistical aggregates of the breath-to-breath series. The means (Avg), standard deviation (Std), mean absolute deviation (MAD) and coefficient of variation ($CV = Std/Avg$) were used per signal. Additionally, the autocorrelation (AR) at one breath lag of a series quantifies the regularity of a measurement [22]. This results in a set of 25 feature values for each RSP signal, representing relevant measures of respiratory variability.

3.4 Conclusion

The feature extraction methods discussed in this chapter result in a set of 70 candidate features. The basic analysis of spectral power in BVP-derived heart rate variability (PRV) returns six features as listed in Table 3.1. The direct BVP features are limited to 3 statistical aggregates of pulse rate. Tables 3.2 and 3.3 provide an overview of the 36 EDA-derived features: 2×10 for trend metrics from both CDA and DDA, as well as 16 SCR-related metrics. Finally, respiratory variability is quantified with 25 features as listed in Table 3.4. A complete list of these 70 features with reference name and description can be found in appendix A.

Chapter 4

Feature analysis methods

The next challenge is to assess whether any of the features extracted in chapter 3 capture significant treatment effects for the studied dataset, and to obtain quantitative insights in these regulation mechanisms. The relevant statistical methods are introduced in this chapter: first, Section 4.1 covers all the important technical considerations. With those concepts established, Section 4.2 specifies all the tests required to answer the initial problem statement from chapter 1.

4.1 Fundamental methods

4.1.1 Data considerations

To obtain reliable results, the distributions of the complete dataset must be taken into account before applying any tests. The following section covers the basic assumptions and which normality transformations can be applied to fit them better in some cases. A short comparison of parametric and non-parametric methods supports the test selection once the parametric assumptions are tested for each feature. Lastly, the effects of sample size and multiple testing are discussed to address the risk of false detections.

Basic and parametric assumptions

The most basic assumption of any statistical analysis is that the selected sample is representative of the target population. The studied dataset is considered representative for the restricted target population of right-handed NT adult men according to the study in [5]. However, the limited sample size must be taken into account when interpreting further results.

The sample must be of sufficient size to have enough testing power. A large sample will better capture the true distribution of the target population, leading to acceptable error rates when applying hypothesis tests. When that true distribution deviates from the expected (e.g. normal) distribution, a larger sample will be needed to obtain equally reliable test results from it. It is here that the assumptions of parametric models become important. With additional information (or assumptions)

regarding the distribution, estimations of the testing power in function of sample size or vice versa become more accurate. The MATLAB function `sampsizepwr()` is used in this study to estimate the power of parametric tests when appropriate. Testing power becomes especially important when the available data (56 subjects) is split further into of 28 datapoints each, or even less after accounting for missing data, noise and outliers.

Non-parametric methods typically require no more than the two basic assumptions listed above to return valid results [54, 55]. Wilcoxon's signed rank or rank sum tests, for example, do not use the values of each sample, but only compare signs or ranks to enable tests on any distribution. However, due to this high tolerance for distributional variations, the obtained results might have less power or even miss significant information, in particular when applied in small samples. It can be beneficial to apply more distribution-specific methods depending on the use case. For this, parametric approaches can be used.

Parametric methods pose further assumptions to obtain more powerful results by leveraging properties of a specific distribution in their test design. If the sample under analysis fits a certain parametric distribution, applying the corresponding parametric test will return much more accurate results than any non-parametric test. However, if the sample does not exactly fit the assumed distribution, it is more correct to apply a non-parametric alternative of that test instead. In this study, the relevant parametric options require normally distributed data and homoskedastic samples (equal variances between two samples). These assumptions are tested in MATLAB with the `jbtest()` and `vartestn()` functions respectively.

Normality transformations

When parametric methods are desired but the required assumptions are not met, it can be valuable to try transformations that bring the sample closer to normality [56]. In order to avoid affecting the relative distribution of split samples (e.g. when testing the difference between treatment groups), all of these transformations should only be applied to the entire sample and never to separate groups. For the dataset in this thesis, a three-step transformation was applied.

First, the Box-Cox power transformation [57] is used to center skewed distributions, which can be useful if a feature is not normally distributed in its extracted form but happens to have an underlying distribution that does fit normality. The MATLAB function `boxcox()` automatically finds the power value λ that brings the transformed distribution closest to normality. Note that a temporary offset value is needed when the original distribution has negative values, which is not supported by the standard transformation. When needed, an offset is added to the original values of a feature sample such that the smallest datapoint is one integer above zero. Although this step does affect the natural shape of the original data, it maintains the relative positions of datapoints. As such, the more powerful parametric tests can be applied for any features that fit the required normality assumptions after transformation.

A second important factor that affects normality is the presence of outliers. Removing them can bring a sample closer to normal distribution, but removing too many datapoints would render the test results useless. Given the already small sample size, the "information per datapoint" is relatively high. When possible, the context of each feature is taken into account to avoid dropping excessive amounts of valuable information.

Lastly, some tests require the sample(s) to fit standard normal distribution ($\mu = 0$, $\sigma = 1$). If the sample under analysis fits basic normality (for any $\mu, \sigma \in R$), univariate z-score standardization provides the required transformation by centering the mean and scaling the standard deviation. This is applied to the complete sample of each individual feature to avoid artificial changes in the relative distribution between groups.

Multiple testing

The effects of multiple testing must be taken into account for correct interpretation of the collective results of repeated tests. In this study, univariate tests are applied to each feature individually (cfr. Section 4.2). Since every test includes the risk of false rejection (type I error) and this risk accumulates for a larger number of parallel tests, the set of features that return significant p-values will likely contain "false detections". The severity of this risk depends on the number of repeated tests, the type of test, and the degree of dependency between tested features. Traditional methods have long been established in statistics, but research fields such as bio-informatics have introduced new ways to manage these effects more efficiently [58].

The most straightforward approach is to adjust the original p-values in accordance with their relative distribution across features. The Bonferroni method is the most basic form of this approach, but it is very strict and often results in excessive information loss. Recent work such as published by Garcia et al. compares more refined correction methods for applications in data mining [59]. Within the context of this thesis, these methods are considered impractical, especially compared to the following approach.

A good alternative is to control the false discovery rate (FDR) as initially proposed by Benjamini and Hochberg [60]. Recent studies, such as the one by Christ et al., demonstrate the practicality of such methods in modern data mining applications [61]. The MATLAB function `mafdr()` supports the estimation of FDR (Benjamini-Hochberg, 1995) and the positive FDR (pFDR), which are both sufficiently powerful for the purposes of this study. The benefit of pFDR-control as proposed by Storey et al. is that it can provide reliable control of multiple testing effects with minimal sample requirements [3]. This implementation is used to gain insight in the risks of false detection when discussing the results in chapter 6.

The interpretation is simple: given a set of features and their independent p-values from the same statistical test, the pFDR algorithm by Storey et al. returns one q-value for each of those p-values. If the rejection threshold is then adjusted to β such that all features for which $q \leq \beta$ are considered significant rejections, that threshold ensures that the expected value of pFDR is β. Typically, this reduces the

final set of "significant" features to a more practical size, while the ability to adjust the pFDR to acceptable levels guarantees more informed conclusions [3, 62].

4.1.2 Difference tests

Statistical difference tests estimate the probability that two or more samples are taken from the same distribution (parametric), or from distributions with similar properties (non-parametric). This is their null hypothesis, and the p-value represents the corresponding probability. The null hypothesis is rejected when $p < \alpha$. In this thesis, a significance level of $\alpha = 0.05$ is used for all tests. A rejection of the null hypothesis is interpreted as a statistically significant indication that the tested samples are taken from different distributions.

This section lists the difference tests that are relevant for the context of this thesis. It is important to take into account that some hypotheses imply paired samples. Samples are paired when each datapoint in sample 1 is contextually linked to a specific datapoint in sample 2, e.g. in the analysis of treatment effects in a patient. Otherwise, the samples are said to be independent. Additionally, not all features fit parametric assumptions (cfr. Section 4.1.1). These two considerations are easily addressed for univariate difference tests.

Testing the difference between two samples for individual features (univariate) is typically done with alternatives of the parametric t-test [63]. If the parametric assumptions are not met, the alternative is to use Wilcoxon's non-parametric methods [64]. In both cases, it is important to specify whether the samples are paired or independent. Table 4.1 lists the MATLAB functions used in this study.

Within the constraints of the given dataset it is difficult to correctly implement multivariate difference tests, especially when testing paired samples. For the purposes of this thesis, the focus will remain on univariate tests. pFDR estimates as explained in the previous section still provide a reliable solution to adjust for multiple testing effects across features. Further discussion of results and context-related limitations can be found in chapter 6.

Name	Samples	Type	MATLAB
Paired-sample t-test	paired	parametric	`ttest()`
Two-sample t-test	independent	parametric	`ttest2()`
Wilcoxon signed rank test	paired	non-parametric	`signrank()`
Wilcoxon rank sum test	independent	non-parametric	`ranksum()`

Table 4.1: Overview of relevant two-sample difference tests

4.1.3 Correlation tests

In addition to various difference tests, Section 4.2.4 requires methods for the assessment of correlation between paired samples. A common method for univariate samples is Pearson's linear correlation, which returns a correlation coefficient ρ

between -1 (perfect anti-correlation) and $+1$ (perfect correlation). The samples are independent when $\rho = 0$. A p-value is then computed for the null hypothesis H0 that the samples are independent: H0 is rejected for $p < 0.05$ when ρ significantly deviates from zero.

The non-parametric alternative is Spearman's correlation method, which poses less strict requirements on the samples by computing Pearson's linear correlation coefficient for the rankings of the sample columns [64]. Both methods are compatible with univariate and multivariate samples, although the multivariate samples might have more difficulty to comply with the model assumptions. The MATLAB function `corr()` is used to compute ρ and the corresponding p-value for both methods. As with difference tests, pFDR estimates provide a reliable correction for multiple testing effects.

4.2 Test specifications

Now that the technical concepts are introduced, this section formulates the tests required to answer the problem statement from chapter 1. After a general assessment of the extracted dataset, the feature analysis is performed in three parts. First, the extracted features are tested for their ability to capture the targeted changes in measured ANS activity. Once validated, the features can be used to quantify the measured treatment effects. Lastly, the prior attachment scores are tested for potential predictive value regarding inter-subject variations. All tests specified below are implemented with a significance threshold $\alpha = 0.05$ and are followed up with pFDR correction across features to obtain consistent results for later interpretation.

4.2.1 Sample preparation and inspection

Before defining further tests, it is important to inspect the available data and perform the relevant preparations. As discussed previously in Section 4.1.1, parametric assumptions are tested for each individual feature. This is tested before and after applying a Box-Cox power transformation, followed by conservative outlier removal. The resulting feature sample is only used if it fits a normal distribution, otherwise the original values are used. Finally, z-score standardization is applied to each feature for consistency.

With that primary dataset (*ftr*) ready for use, three alternative formats are constructed to simplify later tests. The pre- to post-treatment difference (*dftr*) for each subject in rest and stress condition enables direct comparison of treatment effects on those values. In extension to that, the baseline-corrected stress response values *bcsr* quantify the response to eye-contact for each subject. Lastly, a dataset *dbcsr* with the pre- to post-treatment differences in *bcsr* is constructed to complement the analysis of treatment effects in *dftr*. Their mathematical definitions are listed in Table 4.2.

These datasets can be used to confirm whether the assumption that the sampled subjects are representative of the same population holds. Rest and stress conditions in *ftr*, as well as direct *dbcsr* values, are tested for pre-treatment differences. If a

Abbreviations	Construction	Description
ftr	as is	direct feature values
dftr	$= ftr(post) - ftr(pre)$	treatment effect on ftr
bcsr	$= ftr(stress) - ftr(rest)$	baseline-corrected stress response
dbcsr	$= bcsr(post) - bcsr(pre)$	treatment effect on bcsr

Table 4.2: Overview of sample constructions used for further testing

feature is rejected by one or more of these tests, it is likely less reliable in further tests. If many features are rejected, this would be an indication that the sample is too small to reliably capture all variations in the target population. If only a minority of tests are rejected, the pre-treatment similarity assumption holds and further results can be interpreted without reduced confidence.

4.2.2 Stress response

The next question to ask is whether the extracted feature set sufficiently captures the expected physiological responses. If a feature measures significant changes in the subjects' response to stress stimulation, that confirms its relevance for quantification of the social stress induced in this experiment. Such a feature will provide useful information when making further assessments regarding potential treatment effects. The features that do not capture any significant changes in stress response provide some indication of the ANS mechanisms that are likely unaffected by the induced social stress. Additionally, they might still contain relevant pre- to post-treatment changes in fixed condition (rest or stress), which is discussed in the next section.

This response is tested for each feature individually by applying a difference test between resting and stress condition, such that the features which correspond with affected ANS response mechanisms reject the null hypothesis. Note how the samples are paired: each feature value of a subject's activity at rest corresponds to a specific value representing that subject's stress activity.

Three different hypotheses are tested in this first assessment. First, pre-treatment data from all subjects is tested for significant stress responses at baseline conditions. Additionally, the post-treatment stress responses are tested for each group separately to compare changes with the collective baseline response. For this, a two-sample paired test is applied to each feature, for each of the three stress response hypotheses. pFDR estimates for each test are used to take multiple testing into account and select the most informative features for the quantification of stress responses in the studied dataset.

4.2.3 Treatment effect

Now that the informational value of the available features is known, the data can be tested for potential treatment effects. Chapter 1 formulated two hypotheses based on prior OT studies: (1) "OT treatment facilitates parasympathetic tone at rest."

and (2) "OT treatment facilitates sympathetic tone under social stress stimulation." In practical terms, this translates to hypothesized changes in baseline and stress response values after OT treatment. The aim of the following tests is to assess these hypotheses in the available data, and the extent to which these changes can be attributed to OT treatment.

To confirm the posed hypotheses, two sets of difference tests are applied. First, paired tests quantify the significance of treatment effects for each group and condition separately. Second, independent difference tests confirm whether the treatment effects of OT are different from the effects in the placebo group. Note that these tests are applied to samples from the rest and stress conditions (*ftr*) as well as the baseline-corrected stress response (*bcsr*).

General assumption checks for the full dataset will tell whether the parametric or non-parametric method should be preferred. The resulting p-values of each test set (across features) are then used to estimate the pFDR, which reduces the risk of false detections due to multiple testing.

4.2.4 Attachment score interactions

The secondary goal of this thesis is to assess whether a subjects' attachment scores are predictive of their response to social stress stimulation and/or OT treatment. As introduced in chapter 1, the SAAM scores of each subject were collected prior to the experiment to obtain measures of their attachment style at baseline. Prior work suggests that these are good indicators of the social dysregulation patterns targeted by OT treatment [5].

Two interactions are tested. First, the interaction between SAAM scores and pre-treatment response values provides insight in potential baseline variations between subjects. Secondly, interaction tests between SAAM scores and treatment effect values can indicate potential variations in the effectiveness of OT treatment. Both test constructions are applied to samples from the rest condition, stress condition and baseline-corrected stress response (*bcsr*) to cover all potential changes in the experiment.

The practical implementation follows a similar approach to the previous two sections. After initial analysis of the full dataset, the specified samples are constructed and reviewed. Note that this includes the samples constructed from SAAM score data, which are ordinal variables in a range of one to seven. Non-parametric tests are more appropriate to assess these correlations [65]. Each of the three SAAM scores (anxiety, avoidance, security) are tested separately to maintain the benefits of univariate testing. The Spearman correlation coefficient is used for univariate interaction tests with the `corr()` function in MATLAB. The resulting p-values are collected for pFDR estimation across the feature set for each test to adjust for multiple testing in interpretation.

4.3 Conclusion

This chapter presented the statistical tests applied in this thesis to analyze the effects of OT treatment on social stress regulation. After focusing on the fundamental concepts in the first section, the problem statement introduced in chapter 1 was translated to a set of tests that cover the relevant thesis questions and underlying assumptions. In order to maintain reliable results despite the limited sample size and large feature set, multivariate models are avoided. Instead, univariate tests applied to each feature separately provide stable results with minimal influence of feature interactions or problematic sample distributions. Finally, the remaining risk of false detections due to multiple testing is controlled through pFDR correction across the tested features. With these tests fully specified, the corresponding results for the studied dataset are reported in chapters 5-6.

Chapter 5

Results

The methods described in chapters 2-4 were applied to the data provided by N. Daniels et al. in order to obtain quantitative answers to the research questions formulated in Section 1.4. The first section in this chapter reports the relevant intermediate results, from inspection of the raw signals to feature extraction. Then, the second section provides the important findings from the feature analysis.

5.1 Intermediate results

5.1.1 General data inspection

The dataset analyzed in this thesis was introduced in Section 1.3. Daniels et al. collected BVP, EDA and RSP signals from 56 subjects during a four-phasic experiment (rest and stress conditions, before and after treatment with either oxytocin or placebo). Note that, although the original signals were longer, this thesis focused on the first 300 seconds of each recording in order to obtain comparable values [5, 6].

Before elaborating on the separate signal modalities, it should be noted that data from four subjects were removed from the analyses due to acquisition errors. For one of the subjects, there were no recorded signals while for the other three some of the signals were missing. The signals from the remaining 215 subjects were used for the processing and analyses described in the following sections.

5.1.2 Blood volume pulse

A first inspection of the BVP data found that 15 recordings from 5 subjects were missing. Additionally, the recordings from one other subject were removed due to excessive background noise. The remaining 198 observations were used for further processing.

As indicated in Sections 1.2.2 and 2.1, the morphological variations in some BVP signals were challenging to process consistently. As an example, Figure 5.1 illustrates the fiducial points for the BVP segments that were used to illustrate these patterns in the introductory chapter. Although segments with large amounts of background noise were challenging to annotate, signals with low noise levels were annotated

consistently regardless of morphology. Despite a small number of outliers (cfr. Table C.1), the features could be derived from all the signals.

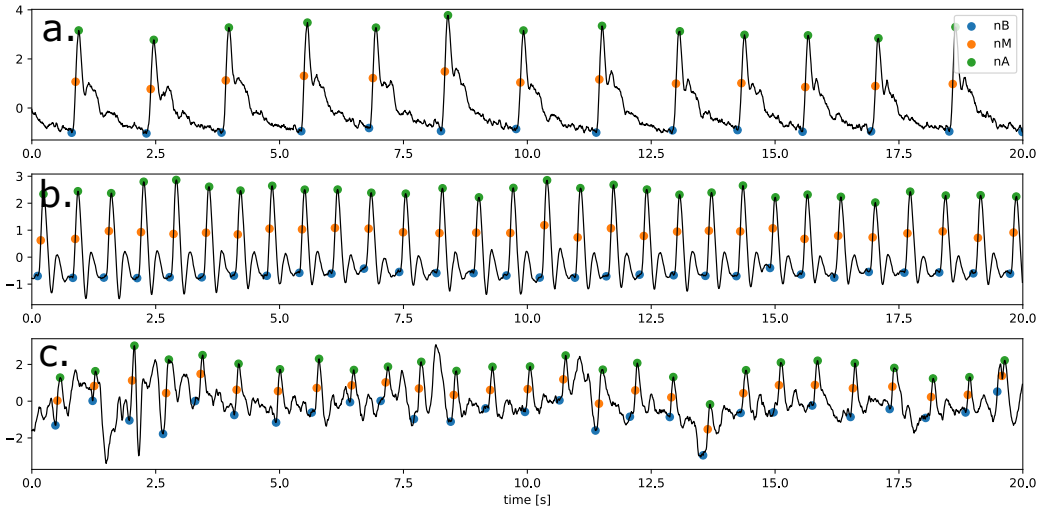

Figure 5.1: Fiducial point annotations obtained for BVP/PPG data. Apex points (nA) in green, base points (nB) in blue and midpoints (nM) in red. Examples of conventional PPG waveforms in (a), deep dicrotic notch waveforms in (b) and a segment with background noise in (c)

5.1.3 Electrodermal activity

All 215 EDA signals were analyzed with the Ledalab[1] tool. Both CDA and DDA algorithms were applied to obtain the tonic, phasic and SCR components. The amplitude threshold for SCR detection in the phasic driver was set at $.05\mu S$ as recommended in the package documentation. Downsampling to 32 HZ reduced the cost of the computations[2] to approximately 20 and 60 minutes, for CDA and DDA respectively.

Although most of the decomposition results were consistent, the fitting of the SCR to the phasic response was not correctly done in cases where a signal was characterized by a reduced phasic activity, often accompanied by slow onsets and long recovery times. An example of this is shown in Figure 5.2. When this occurs, the tail of the SCR component (dotted line) drops abruptly instead of slowly returning to the tonic baseline. This pattern was found in 25 signals in the EDA dataset, and was more prevalent in subjects with low phasic activity. Figure 5.3 shows the decomposition of the EDA from the same subject before treatment, in which the decomposition was correctly done. Both signals were measured under resting

[1] http://ledalab.de

[2] On an Intel core i7 (8th gen), 8GB RAM, no parallelisation available in Ledalab

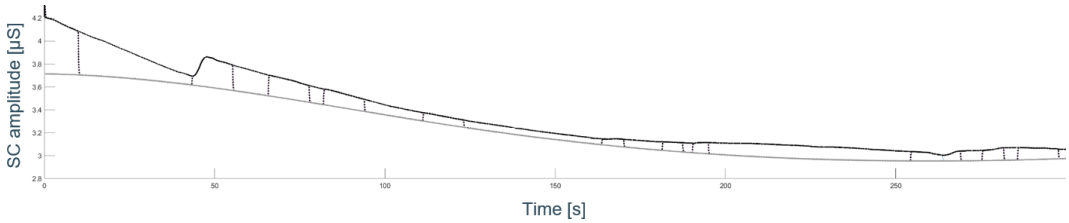

FIGURE 5.2: Example of faulty decomposition by the DDA algorithm. The vertical lines are sudden drops in the recovery slope SCRs that were fitted poorly to the given EDA signal. Note the valid tonic baseline with only one SCR with a very slow onset and recovery

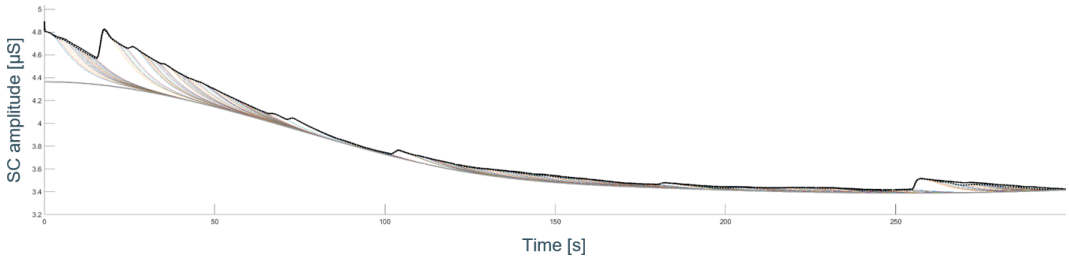

FIGURE 5.3: Example of correct decomposition by the DDA algorithm. The SCRs in this signal are closer to the expected waveform, making it easier for DDA to find a well-fitting $IRF(t)$

conditions. In extreme cases where a signal contained no differentiable SCRs, the IRI-based features (inverse response rate) of this signal were removed from further analyses since no valid value could be assigned.

Additionally, some signals with high noise or strong motion artefacts resulted in extreme values for phasic activity features such as Amp or Area. These were succesfully detected in the outlier removal step. However, visual inspection did show that it remains challenging to differentiate hyperresponders (high phasic excitability) from background noise in some cases.

These exceptional feature values were observed for nine subjects (cfr. Table C.1). A visual inspection confirmed that these values were caused by excessive noise, artefacts or near-zero activity. Also, some of these nine subjects were found to be consistent nonresponders or hyperresponders (cfr. Section 1.2.3) in multiple recordings, resulting in a total of 17 corrupted signals. Since no further reference was available, all observations with numerically valid features were included for further analysis to assess their role in the targeted response mechanisms. High-phasic outliers were removed in the normalization step (cfr. Section 5.2.1).

5.1.4 Respiration

All RSP signals were bandpass filtered and annotated as discussed in Section 2.3. Although no problematic noise was encountered in the annotation process, the natural breathing patterns introduced in Section 1.2.4 were challenging to annotate consistently. In cases of strong motion artefacts, thought to be caused by sudden displacement of the breathing belt, the bandpass filter was unable to remove the transient peak completely. These artefacts were included in the breath annotations due to lack of alternatives within the scope of this thesis and, given their limited occurrence, this issue was not addressed any further.

Figure 5.4 shows the resulting annotations for the segments from Figure 1.3. For some of these breathing patterns, the annotations remained doubtful. However, due to the lack of ground truth, and after visual inspection, it was decided to work with these results and, in any case, the breaths were annotated consistently in most of the cases. Feature extraction produced numerically valid results for all 215 signals.

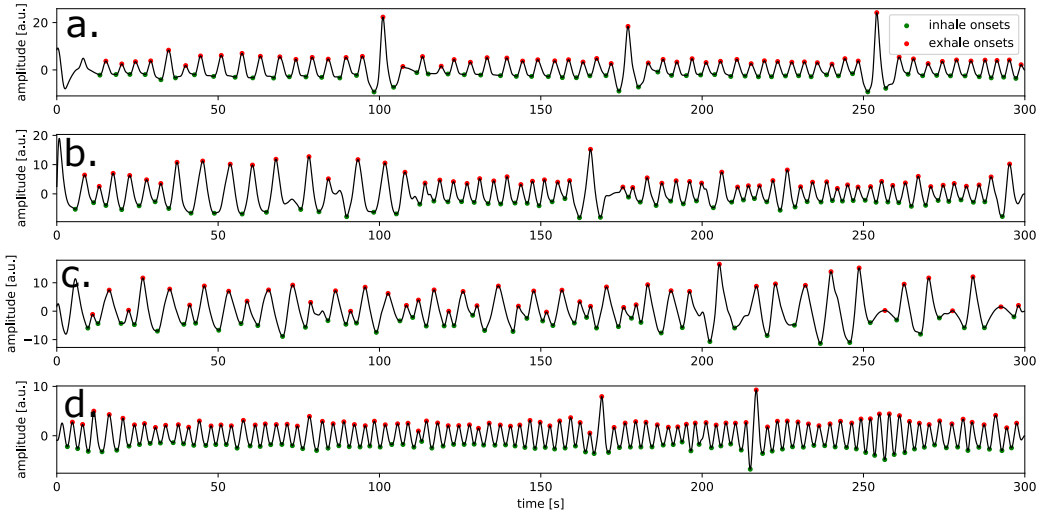

Figure 5.4: Breath annotation results with robust peak prominence thresholding

5.2 Feature analysis

The following section discusses the analysis of the different extracted features with the methods explained in chapter 4, for which a technical summary is presented in appendix B. After a general inspection and normalization of the features, they were used to construct different samples in which different tests were performed. The results provided measures of stress response and treatment effects in the analyzed signal modalities, as well as their dependencies on social conditions and inter-personal variations in attachment style. Additionally, the relevant assumptions made by each test were assessed to validate the results.

5.2.1 Sample preparation and inspection

The MATLAB function `rmoutliers()` was used to identify datapoints in the features that were more than three standard deviations away from the mean. This outlier removal method was only applied if the resulting sample fit a normal distribution well ($p < 0.05$ for Jarque-Bera test). If not, the original sample was used to avoid excessive loss of valid observations. Note that by choosing this approach, the DDA features discussed in Section 5.1.3 were not interpreted as outliers but as relevant datapoints.

After applying the Box-Cox transformation, 44 out of 70 features passed the Jarque-Bera normality test. The remaining features were left unaltered to maintain their original distribution. Z-score standardization was then used to center the means ($\mu = 0$) and scale the variances ($\sigma = 1$) for each feature. Complete numerical results for the transformed samples, as well as the applied Box-Cox λ and offset values, can be found in Table C.1.

To illustrate the general variations in outcome across these normalized features, Figure 5.5 compares the sample distribution before and after applying the boxcox transformation to three of the features. The first sample shows how the long asymmetrical tail in the original distribution of DDA_Amp_avg was removed. Since the applied power transformation was unable to bring the complete sample close enough to a normal distribution, the furthest four observations were removed as outliers. For features such as RSP_Vt_MN, however, the power transformation did manage to bring all datapoints under a more normalized distribution, such that all datapoints were included as non-outliers. Lastly, there were a few cases such as RSP_Te_AR where the original data already followed a normal distribution. In those cases the transformed data was used since this was thought to bring the distribution closer to a normal one.

The pre-treatment similarity assumptions were then verified. Eight of the features displayed significant differences in stationary conditions (rest and/or stress). These are listed in Table 5.1. The same test was performed on the baseline-corrected stress response values (*bcsr*, cfr. Table 4.2), resulting in eight significant features. These are listed in Table 5.2. Such differences between groups, prior to receiving treatment, indicate a risk of false conclusions from later tests. Either the rejected features are insufficient to capture a stable distribution for the small samples, or the

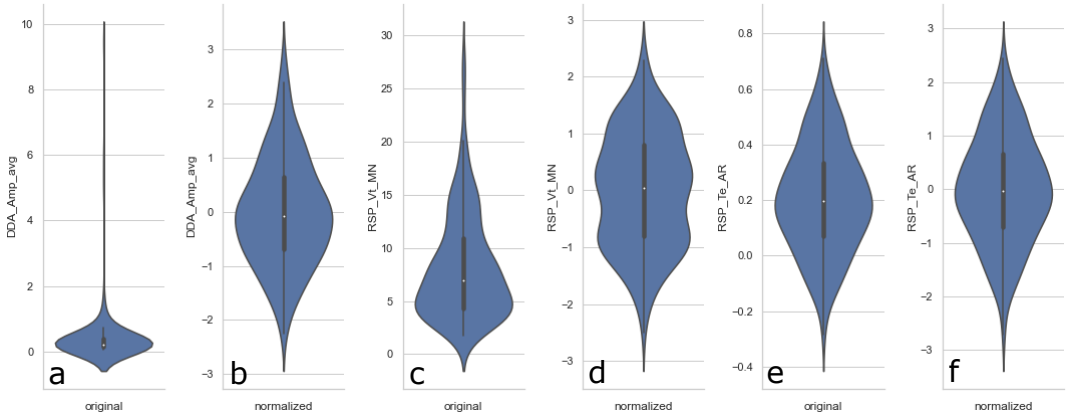

Figure 5.5: Results of the application of the boxcox transformation and outlier removal to three of the features derived from the signals. In a and b, the mean SCR amplitude feature (DDA) and its tranformation, respectively. In c and d, the mean tidal volume (RSP) and its tranformation. In e and f, the autocorrelation at one breath lag of the expiratory times (RSP) and its transformation.

sampled treatment groups are not statistically representative of the same population (at least for these features). However, the pFDR corrections suggest a high risk of false detection in these tests (up to 90%). Although further results should be interpreted with care, the pre-treatment similarity between groups was considered to be sufficiently valid for the purposes of this thesis.

	rest		stress	
Feature	p	q	p	q
RSP RR MAD	**0.0190**	0.0942	0.5893	0.5318
RSP RR SD	**0.0190**	0.0942	0.6407	0.5318
RSP Te MN	**0.0268**	0.0942	**0.0119**	0.5044
DDA Area sum	**0.0460**	0.0942	0.3321	0.5129
CDA Phasic SC Power	**0.0460**	0.0942	0.6277	0.5318
RSP RR MN	0.0732	0.0942	**0.0227**	0.5044
CDA SC slope	0.0906	0.0942	**0.0451**	0.5044
DDA SC slope	0.0906	0.0942	**0.0451**	0.5044

Table 5.1: Significant differences in pre-treatment activity between treatment groups (OT and placebo) under conditions of rest and stress

	pre		post	
Feature	p	q	p	q
PRV LFn	**0.0160**	0.6050	0.4295	0.9933
PRV LH ratio	**0.0271**	0.6050	0.6101	0.9933
PRV HFn	**0.0271**	0.6050	0.6818	0.9933
DDA HalfRec time mad	0.0615	0.6559	**0.0379**	0.6585
DDA HalfRec time std	0.0703	0.6559	**0.0313**	0.6585
DDA HalfRec time avg	0.1770	0.8541	**0.0074**	0.5156
DDA P50 width avg	0.3735	0.8952	**0.0232**	0.6585

TABLE 5.2: Significant differences in baseline-corrected stress response (*bcsr*) between treatment groups, before (pre) and after (post) treatment

5.2.2 Stress response

After doing the data preparation and checking the assumptions, the features were tested for significant differences when quantifying the induced social stress response. First, differences between rest and stress conditions were tested for the complete dataset before treatment, as well as for individual treatment groups after administration. Complete numerical results are listed in Table C.2.

Figure 5.6 provides a visual overview of the features which were found to capture a significant response, in increasing order of p-values and with their corresponding q-values. While the stress response before treatment showed measurable changes for 26 features, the same tests after treatment captured significant differences only for 8 and 20 features in the OT and placebo groups respectively. The q-values did not indicate significant risks of false detection.

To complement these findings, features with significant post-treatment differences in the baseline-corrected stress response were added to Table 5.2 in the previous section. According to that test, only the half-recovery time features from DDA measured significant differences in the stress responses of both treatment groups. However, pFDR corrections indicated a high risk of false detection in these results (cfr. q-values). Note how this does not fully explain the findings from Figure 5.6, where 12 features captured a significant stress response for only the placebo group. The following tests were used to assess the statistical significance of these treatment effects.

5.2.3 Treatment effects

Two sets of tests were applied to assess the effects of OT treatment, by breaking the analysis down into two steps: the identification of features measuring significant differences between treatment groups, followed by an assessment of the corresponding treatment effects in each group. Note that the effect samples (*dftr, bcsr, dbcsr*) were constructed as explained in Table 4.2 to enable direct two-sample comparison of feature values.

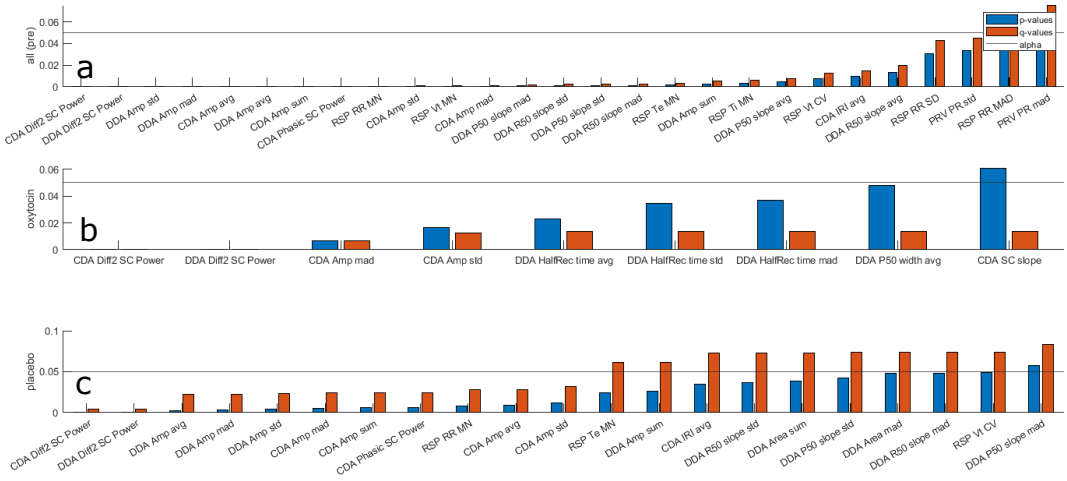

Figure 5.6: Significant stress response features, ranked by p-values. In a, the results of testing the difference between activity levels in rest and stress before treatment, for all subjects. In b and c, the results of testing the same difference after administration of oxytocin and placebo respectively.

Table 5.3 shows the significant results from the first tests. Differences in treatment effect on resting activity were found for seven features. Although no significant differences were found in effects on activity under social stress, six features did show significant differences in treatment effect on the baseline-corrected stress response (*bcsr*). pFDR corrections estimated the risk of false detections in these results at 20% and 35% for rest and *bcsr* samples, respectively (cfr. q-values).

	rest		stress		*dbcsr*	
feature	p	q	p	q	p	q
DDA HalfRec time avg	**0.008**	0.174	1.000	1.000	**0.014**	0.349
DDA P50 width avg	**0.013**	0.174	0.757	1.000	**0.017**	0.349
DDA P50 width std	**0.020**	0.174	0.542	1.000	0.052	0.349
DDA P50 width mad	**0.021**	0.174	0.342	1.000	0.052	0.349
DDA HalfRec time mad	**0.031**	0.192	0.596	1.000	0.060	0.349
DDA Amp sum	**0.035**	0.192	0.749	1.000	**0.032**	0.349
DDA HalfRec time std	**0.045**	0.212	0.849	1.000	**0.037**	0.349
PRV HFn	0.653	0.456	0.151	1.000	**0.027**	0.349
PRV HF	0.865	0.456	0.285	1.000	**0.042**	0.349

Table 5.3: Features which measured significant differences in treatment effects on activity levels in rest, stress and baseline-corrected stress response (*dbcsr*) between treatment groups

To distinguish OT-specific effects from secondary effects (placebo, habituation or other), the differences in treatment effects (*dftr*) were tested for each treatment group separately. The significant differences in stationary conditions (rest and stress) for each group are shown in Figure 5.7. Complete numerical results can be found in Table C.3 in the appendix. Table 5.4 complements these findings with features containing significant treatment effects on baseline-corrected stress response.

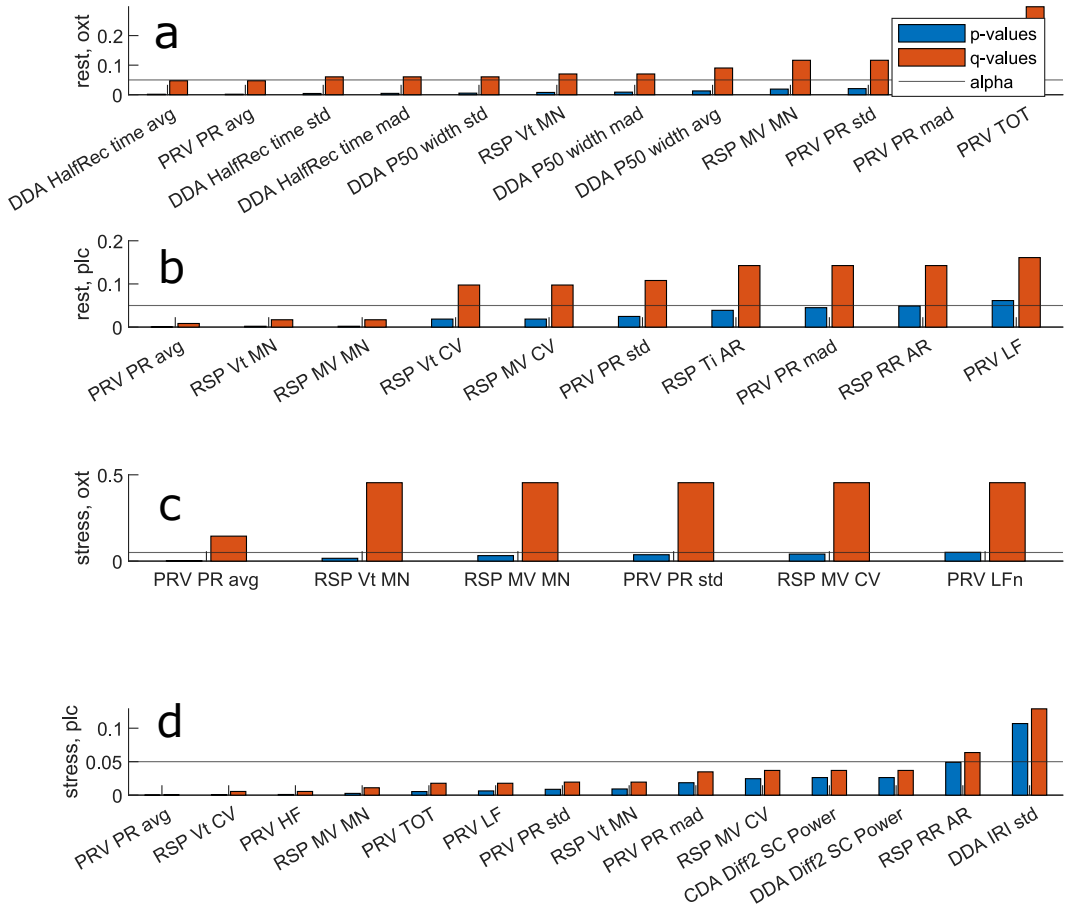

FIGURE 5.7: Significant treatment effects on resting activity for the OT group (a) and the placebo group (b), as well as the effects on activity under social stress for the OT group (c) and the placebo group (d)

Finally, the results from Table 5.3 were plotted in Figure 5.8 to provide a simplified overview of the direction of treatment effects. These are used for further interpretation of physiological responses in the next section. Note that the DDA-based features showed very similar patterns, since they all measure the recovery rate in EDA.

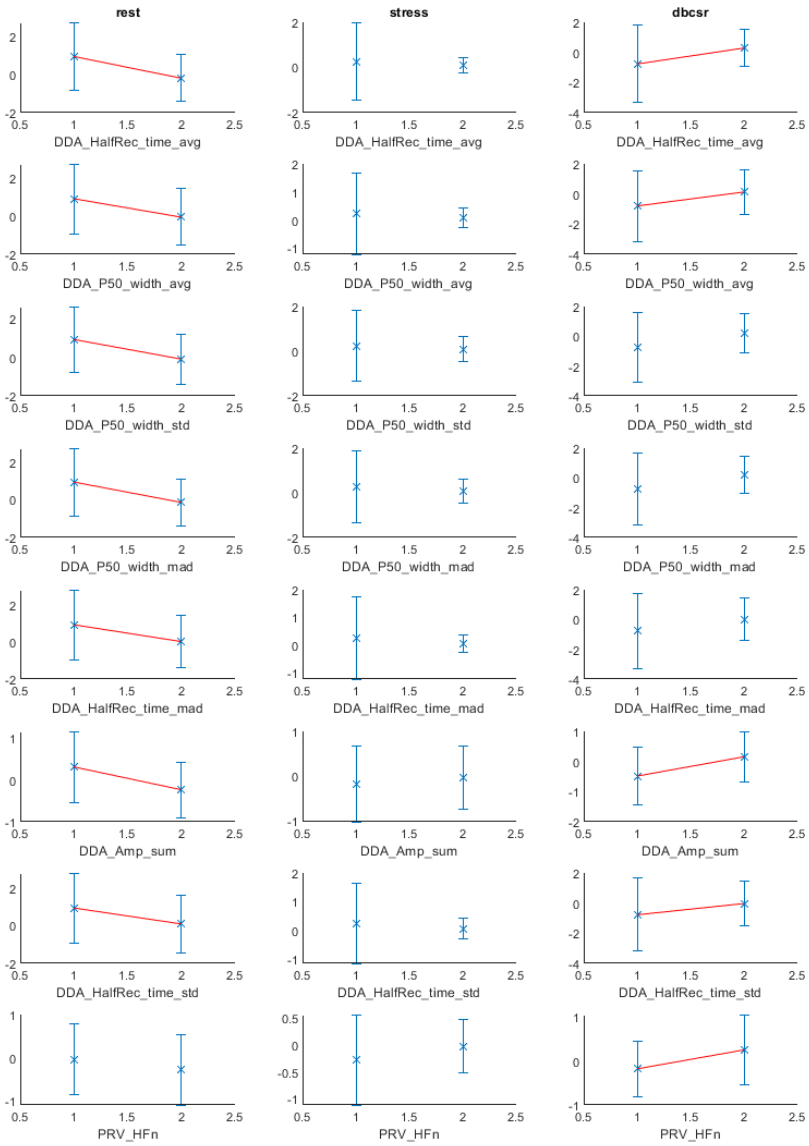

Figure 5.8: Comparison of treatment effects as were tested in Table 5.3. Each plot shows the treatment effects of OT (x=1) and placebo (x=2) by their mean and standard deviation, to indicate relative differences and the direction of effects on features from pre- to post-treatment. The three columns provide these results for effects on activity at rest and stress, as well as the baseline-corrected values. The red lines indicate a significance of $p < 0.05$ as obtained with Wilcoxon rank sum tests.

	oxytocin		placebo	
feature	p	q	p	q
DDA HalfRec time std	**0.0225**	0.1748	0.7570	0.6943
DDA HalfRec time mad	**0.0308**	0.1748	0.8612	0.7370
DDA Amp sum	**0.0422**	0.1748	0.3809	0.6724
RSP RR MN	**0.0435**	0.1748	0.2312	0.6724
DDA HalfRec time avg	**0.0447**	0.1748	0.2012	0.6724

TABLE 5.4: Significant pre- to post-treatment differences in baseline-corrected stress response ($bcsr$), for a side-by-side comparison of the effects of OT and placebo.

5.2.4 Attachment score interactions

Figure 5.9 shows the distribution of the SAAM scores for the 56 participants of this dataset. The histograms of each SAAM score are shown on the main diagonal of the grid, while the pairwise scatterplots visualize interactions between the three scales.

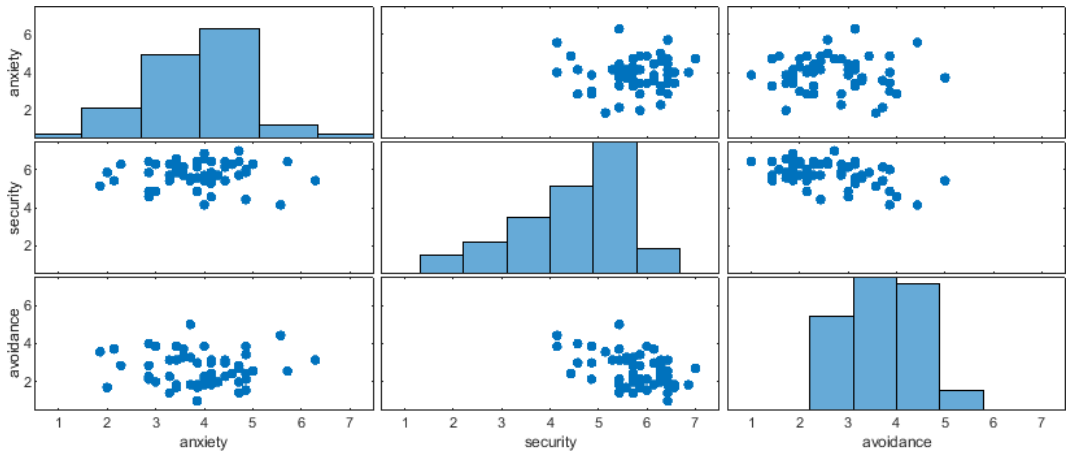

FIGURE 5.9: Attachment score (SAAM) distributions

Given these SAAM scores, the Spearman correlation tests found a relatively large amount of interactions with significantly non-zero correlation coefficients (Spearman's ρ). The corresponding p-values[3] were used to apply pFDR correction across all features, such that the q-values could be used to extract a compact overview of results: Table 5.5 lists the features which were found to interact with SAAM scores in pre-treatment conditions, while Tables 5.6 and 5.7 provide the features treatment effect interactions. Note that these tables, based on $q < 0.05$, do not cover the full

[3] Although Spearman's ρ could be used directly, the p-values also take the sample size of each feature into account. As such, features with low sample size, e.g. due to outliers, are given less weight in the comparison of "strongest" interactions.

results. Tables C.4 - C.6 list all features for which $p < .05$. In all tables, the sign of ρ indicates the direction of the correlation.

Feature	sample	SAAM	p-value	q-value	ρ
RSP Te MN	bcsr	anxiety	0.0071	0.0239	0.3688
RSP Te MAD	bcsr	anxiety	0.0243	0.0239	0.3120
RSP Te SD	bcsr	anxiety	0.0420	0.0239	0.2831
RSP RR MN	bcsr	anxiety	0.0460	0.0239	-0.2807
CDA IRI avg	bcsr	anxiety	0.0461	0.0239	0.2924

Table 5.5: Significant correlations between SAAM scores and pre-treatment data after pFDR correction, only included findings for which $q < 0.05$

Feature	sample	SAAM	p-value	q-value	ρ
CDA SC slope	stress	security	0.0081	0.0121	-0.5076
DDA SC slope	stress	security	0.0081	0.0121	-0.5076
DDA HalfRec time avg	stress	security	0.0219	0.0211	0.4655
DDA HalfRec time mad	stress	security	0.0283	0.0211	0.4475
DDA P50 width avg	stress	security	0.0410	0.0211	0.4200
DDA HalfRec time std	stress	security	0.0485	0.0211	0.4068

Table 5.6: Significant correlations between SAAM scores and OT effects after pFDR correction, only included findings for which $q < 0.05$

Feature	sample	SAAM	p-value	q-value	ρ
CDA Amp avg	stress	anxiety	0.0002	0.0019	0.7211
CDA Amp mad	stress	anxiety	0.0182	0.0328	0.4779
DDA Area sum	stress	anxiety	0.0276	0.0328	0.4317
DDA Amp sum	stress	anxiety	0.0288	0.0328	0.4290
CDA Amp std	stress	anxiety	0.0288	0.0328	0.4464
CDA Amp sum	stress	anxiety	0.0304	0.0328	0.4252
CDA IRI mad	stress	anxiety	0.0327	0.0328	-0.4465
CDA IRI avg	stress	anxiety	0.0338	0.0328	-0.4440
CDA Phasic SC Power	stress	anxiety	0.0343	0.0328	0.4166
PRV HFn	stress	anxiety	0.0483	0.0416	-0.3988

Table 5.7: Significant correlations between SAAM scores and placebo effects after pFDR correction, only included findings for which $q < 0.05$

However, the p-values alone are typically not sufficient to validate the physiological interaction. The figures below illustrate this issue with a small selection of correlation pairs. Figure 5.10 shows two features from pre-treatment conditions, which were selected based on their relatively high correlation coefficient. Similarly, Figure 5.11 shows four feature correlations from the treatment effects sample (*dftr*). Although the CDA_Amp_avg (cfr. scatterplot e) explains its high correlation value ($\rho = 0.72$) well, the other features remain somewhat doubtful.

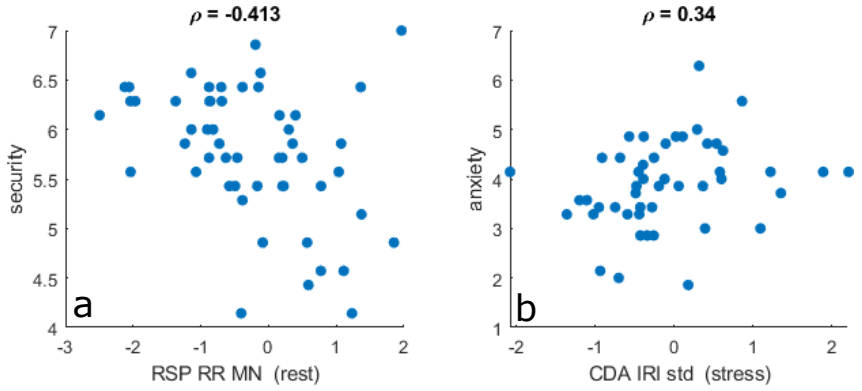

FIGURE 5.10: Two examples of SAAM score interactions with pre-treatment activity levels (rest and stress), with high Spearman's rank correlation coefficient (ρ). The scatterplot in a shows the correlation between mean respiratory rates in rest and the secure attachment scale. In b, the standard deviation of inter-response intervals (EDA) during conditions of social stress. Both samples are observed before treatment.

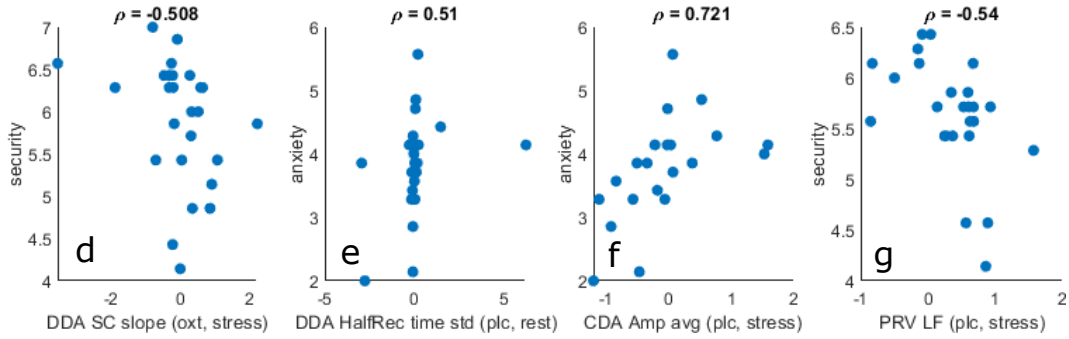

FIGURE 5.11: Four examples of SAAM score interactions with treatment effects, with high Spearman's rank correlation coefficient (ρ). Note the smaller sample size due to separation of treatment groups by oxytocin (oxt) and placebo (plc).

5.3 Conclusion

Section 5.1 reported intermediate results from signal processing and feature extraction, to provide some insight in the technical challenges and exceptions encountered in each signal modality. Taking these findings into account where possible, the features were normalized, inspected and tested according to the specifications in Chapter 4. Section 5.2 then discussed the feature analysis. For the results that were summarized, more extensive numerical results can be found in Appendix C.

Although multiple features were found to capture significant responses to social stress and/or OT treatment, these findings should be interpreted further in their physiological context. The challenging signal patterns and outliers also illustrated strengths and limitations of the technical approach, which should be taken into account when concluding the significance of results. Chapter 6 provides this discussion based on the previously reported results.

Chapter 6

Discussion

This chapter discusses all previous findings to formulate a detailed answer to the problem statement posed in Chapter 1. First, Section 6.1 interprets the results from Chapter 5 to formulate the main conclusions for this thesis. Section 6.2 then situates these findings again in the broader research context introduced in Section 1.1. The strengths and limitations of the analyses are briefly discussed in Section 6.3.

6.1 Main findings

Section 1.4 formulated three primary research questions for this thesis: (1) How is social stress regulation expressed in noninvasive measures of the ANS response? (2) Do the observations contain OT-specific treatment effects? (3) Which contextual factors influence these responses and/or effects? This section summarizes the contributions to answer these questions based on the results presented in Chapter 5.

6.1.1 Noninvasive measures of social stress regulation

A significant stress response was found in 26 features prior to treatment, which confirms that the eye-contact does indeed stimulate a response, and that these responses are succesfully captured by the selected feature set. In these results, EDA-derived features had more discriminative power between stress and stress free periods, although a smaller amount of respiration and pulse rate features were also found to capture significant differences between the two conditions.

After treatment, the stress responses for each group were noticeably different from those in pre-treatment conditions. Note how these differences can be interpreted as a combination of feature-specific facilitations (increased response) and inhibitions (reduced response). For example, multiple half-recovery time features from DDA were found to measure a significant stress response only for the OT group after treatment. Reversely, the RSP_Vt_CV measured a significant response in the pre-treatment group, while only reaching $p = 0.049$ in the placebo-group and $p > 0.05$ for the OT group (cfr. Table C.2). Although less pronounced in this case, this suggests a treatment-specific inhibition of respiratory variability. The relatively

better performance of DDA features can likely be attributed to the higher uncertainty in respiratory features (cfr. Section 6.3), complemented by the fact that the used decomposition methods were validated in prior work while a more simplistic approach was used for the breath annotations.

6.1.2 Identification of treatment effects

The assessment of treatment effects was approach in two steps. First, direct comparison of effects between treatment groups indicated the features capturing treatment-specific effects (as opposed to shared effects due to secondary factors). Second, a side-by-side (indirect) comparison of tests in separate treatment groups allowed these treatment-specific effects to be broken down into its components: changes in features measuring activity in rest and/or stress due to one or both treatments were identified to explain the results from direct comparison in more detail. Additionally, the baseline-corrected stress response (*bcsr*) was calculated for each feature, as the difference between its values in rest and under stress. These alternative samples complement the findings from direct assessment of activity in rest and under stress.

Direct comparison of treatment groups

The results presented in Chapter 5 confirmed that treatment-specific effects on social stress regulation did occur within the context of the experiment. Remarkably, multiple features captured significant changes at rest when compared before and after treatment, while these differences were not detected during eye-contact (i.e. under social stress). It was also found that the *bcsr* found more significant differences due to the fact that it assesses the activity levels under stress relative to the subjects' resting activity.

Despite the risk of false detections ($q \approx 0.20$), seven DDA-based features indicated the strongest treatment-specific effects on resting activity. Their *bcsr* values remained significant for four features. Given the non-significance of these features in stress response, this effect was likely due to the effect in resting activity (carry-over from baseline correction).

Additionally, the high frequency power of the PRV was also affected by treatment, alhough this effect was only significant for the *bcsr*. In combination with the prior tests on rest and stress conditions separately (no significant treatment effect in either condition), this suggests a shared effect between both. Alternatively, this could be a false detection: pFDR correction estimated the risk of false detections to be 30% when including these features as significant ($p = 0.04$ and $q = 0.35$). However, given that a similar effect on the PRV_HF was reported in a prior study of this dataset by Daniels et al., this is likely a valid detection [6].

Regarding the direction of effects, the main findings were that OT appears to result in an increase in EDA recovery time and variability in resting conditions. The effects on *bcsr* showed a decrease in the same features compared to placebo treatment. These differences were not observed in activity under stress, which is thought to be related to the dependencies of OT treatment on social context (cfr.

Section 6.2). Despite this lack of statistical differences in the activity under stress, the results did show a remarkable difference in inter-subject variability between treatment groups. This indicates a context-dependency of participants' response to OT treatment, which was much less pronounced in the placebo group.

Indirect comparison of treatment groups

By comparing the features before and after treatment for each group, the results from direct comparison (cfr. Table 5.3) can be further broken down into contributions from OT treatment, placebo, social context, habituation and/or unknown factors.

Such OT-specific treatment effects were first suggested when testing stress responses: out of the 26 features that measured significant responses in pre-treatment, 20 features showed a reduced response after OT treatment, while only 6 features showed a similar reduction in the placebo group. Note how this leads to a categorization of response properties: some responses are actively affected by OT treatment, while others are influenced by a secondary factor (placebo or other) or even not affected at all.

A similar conclusion can be made for treatment effects on rest and stress activity. Features such as DDA_HalfRec_time_avg only measured significant treatment effects in the OT group. Such OT-specific treatment effects were found for 11 features in activity levels at rest, and for 5 features during eye-contact. Reversely, some features were affected by secondary (non-treatment-specific) factors. For example, all four tests measured a significant treatment effect on PRV_PR_avg, i.e. in rest and stress for both treatment groups. This indicates that the differences in that feature were significant, but independent from the controlled social stress and treatment factors.

Finally, statistical evidence of OT-specific treatment effects on the baseline-corrected stress response was found for four DDA-based features, as well as for the mean respiratory rate. These five features measured significant changes in stress response from pre- to post-treatment in the OT-group, albeit with a 17% risk of false detections, while no significant treatment effects on *bcsr* were found in the placebo group.

6.1.3 Contextual dependencies

So far, it is clear that not all features responded consistently to the controlled factors of treatment and social stimuli. The previous tests indicated that, while some features did measure only OT-specific effects, other features responded to only placebo treatment. Some features even responded independently of treatment, i.e. they changed or remained stable in all tested conditions. Additionally, some features contained strong variations between subjects, which made it more difficult to differentiate the targeted effects.

As was introduced in Section 1.1, these variations are thought to be, at least partially, influenced by variations in the participants' attachment style (SAAM). The correlations between SAAM scores and differences in stress response before

treatment, as well as the correlations with treatment effects, were evaluated to verify this.

The strongest pre-treatment interactions were found between anxious attachment and variations in the baseline-corrected stress response. According to the correlation coefficients ρ, participants with higher anxiety scores had a stronger tendency towards slower and more irregular breathing in response to social stress (relative to resting activity). Similarly, more anxious participants were found to respond with a stronger increase in inter-response-intervals (lower response rate in EDA). Remarkably, participants with higher secure attachment showed a similar decrease in respiratory rate. Although the expectation from prior works is that secure and anxious attachment scales are negatively correlated, this was not evident in this dataset, leading to the detection of unexpected interactions.

There were also strong correlations between attachment scores and treatment effects: the OT group showed strong interactions between DDA-based features and the secure attachment scale (under stress), while treatment effects in the placebo group were more prominently correlated with the anxious attachment scale. Although not all correlations should be considered significant, interactions such as the one observed between the participants' anxiety scores and their mean SCR amplitude (CDA_Amp_avg), which is an indication of phasic activity, appeared to be relatively linear for the 28 participants in the placebo group. Since this increased stress response was only detected in the placebo group, this is an additional indicator of OT-specific anxiolytic effects being more pronounced in people with higher anxiety levels at baseline.

Despite the large amount of statistically significant correlations, it must be taken into account that these interactions might not extrapolate well to other participant groups. As was shown in the previous chapter, the 56 participants are not distributed evenly across the full range (1-7) of the three SAAM scales. The correlations reported in this thesis should therefore not be expected to directly agree with findings for other populations, especially when different SAAM score distributions are expected/observed.

6.2 Comparison to prior studies

This section briefly compares the conclusions from this thesis to the prior works introduced in Section 1.1. The interactionist approach suggested in [11] is followed in order to constructively discuss the observed context dependencies of OT.

The effects of OT treatment were already explored by Daniels et al. using the same dataset as in this thesis [5, 6]. They reported an OT-specific reduction in EDA recovery time during eye-contact, as well as a facilitation of parasympathetic tone in PRV during rest, i.e. a relative increase in PRV_HF. The results in Figure 5.8 agreed with these findings, although the effect on PRV_HF was not statistically significant in this thesis.

More generally, Bartz and Graustella formulated the effects of OT treatment as a combination of anxiolytic and prosocial effects [11, 12]. Bakermans et al. noted

additional findings of defensive aggressive behavior and increased alertness towards people outside the close social circle, supporting the role of OT in "tend and defend" mechanisms [9]. Beyond the findings in EDA recovery time and cardiovascular tone, additional OT-specific effects were found in the mean respiratory rate, confirming the relevance of all three signal modalities in this study.

Finally, some features showed clear context-dependencies, matching the hypotheses in [9]. Bakermans et al. proposed three contextual categories: social context, personal history, and personality traits. In this thesis, the social stress induced by intermittent eye-contact provided a well-controlled parameter to compare effects, showing clear responses in the feature set (cfr. Section 5.2.2). The attachment scores provided a measure of inter-personal variations, and this information was found to reach relatively strong correlations (up to $\rho = 0.71$) with the treatment effects captured by the extracted features (cfr. Section 5.2.4). More detailed inspection of the treatment effects in Section 6.1 found additional context-dependencies of lesser known origin. These are thought to be related to feature-specific response mechanisms, placebo effects, habituation to the context of the experiment, and/or unknown variations in external factors.

6.3 Strengths and limitations

The findings from previous sections do have strengths and limitations that should be taken into account. Due to various challenges in the signal processing steps, the approach in this thesis required certain trade-offs to be made. The following section goes over the important considerations for each signal modality.

6.3.1 Blood volume pulse

As was introduced in Section 3.1, the BVP waveform contained strong variations in the depth of dicrotic notches. Section 2.1 discussed the typical factors of morphological differences, with reference to prior literature. Although this was likely caused by a combination of sensor placement and person-dependent cardiovascular properties, this thesis was unable to verify the true cause(s) for this dataset. As such, the resulting BVP features were considered valid for the purposes of this thesis, under the assumption that the morphological variations are "true" information, and not noise accumulated during acquisition.

In order to obtain more consistent annotations, it was decided to use the peak detection algorithm as implemented in the R-DECO tool instead of the PPG-oriented pulse detection method by J. Lazaro. Although the resulting annotations were found to be reliable for this dataset, it should be noted that the used algorithm, which was originally intended for ECG processing, has not been formally validated for use with PPG signals. It is recommended to remain critical of the results when using this method in later work [43, 41].

6.3.2 Electrodermal activity

While the raw EDA signals were generally of very good quality, the few cases of low and high phasic activity were challenging to address correctly. Section 1.2.3 introduced these variations, with reference to physiological interpretations by Boucsein et al. [30]. Most likely, these low/high activity levels correspond to subjects with high/low response thresholds ("nonresponders" and "hyperresponders") respectively. However, these signals can be difficult to differentiate from noise caused by poor electrode contact or motion artefacts without further context.

The decomposion methods, especially DDA given its focus on SCR separation, mostly had difficulty fitting the low activity signals, which resulted in challenging outliers in some cases. These technical issues were reported earlier in Section 5.1.3. However, the features extracted from these rough SCRs estimates were still found to be representative of the signal properties (e.g. very low values for the recovery rate features). The opposite cases of high activity signals simply contained more phasic activity, which was less problematic for the used decomposition methods.

Because of the physiological relevance of these variations in activity levels, all signals were included in the feature analysis, with the exception of invalidated response rate features (due to lack of SCRs in purely tonic signals). Using the non-parametric tests, features containing extreme cases of high phasic activity could be included without excessive risk of false results, in order to assess their role in the targeted response mechanisms. Four signals with high frequency noise were succesfully detected during outlier removal, leaving the remaining sample safe for use.

6.3.3 Respiration

The various breathing patterns introduced in Section 1.2.4 were the primary challenge in this signal modality. Since the participants' breathing was not explicitly controlled, these signals were observations of natural breathing at rest and under social stress. A breath annotation method based on peak prominence thresholding was eventually implemented to capture the respiratory variations as consistently as possible (cfr Section 2.3).

It remains difficult to obtain breath annotations that are fully aligned with a physiological "ground truth". Studies in prior literature typically assumed the validity of third-party algorithms, or used expert annotations of their specific data as ground truth, which was not applicable for this thesis.

Additionally, the relative amplitudes of a single breathing belt limit the use of amplitude-based respiratory features. Features such as minute volume and tidal volume were often reported to be relevant for the assessment of psychological stress in reference literature (cfr. Section 3.3). However, this information would only be available by adding a second breathing belt (abdominal) and calibrating both belts as in [38]. If the absolute amplitudes were available, it would be possible to define a more consistent threshold of "valid breath amplitude" compared to the current peak prominence thresholding method.

6.3.4 Feature analysis

There are certain limitations to the findings discussed earlier in Sections 6.1 and 6.2. Generally, this comes down to the differentiative power of the applied tests, and the validity of the underlying assumptions for the dataset under analysis.

In chapter 5, the decision was made to apply univariate tests to each feature individually, instead of using a multivariate approach. The features were then compared using pFDR correction. The benefits of this approach are minimal assumption requirements, a high tolerance for feature interaction, and the ability to select any univariate test statistic. As a trade-off, this approach does not provide a direct statistical metric for the significance of treatment effects across all features in the sample. However, the sample size (56 subjects) was considerably small compared to the extracted feature set (70 features), which would require careful dimensionality reduction (e.g. principal component analysis or stepwise regression) before reliable conclusions can be made.

Finally, it should be noted that the results of this thesis do not necessarily extrapolate well to all groups in the broader human population. As explained in Section 1.3, the reported findings represent a population of healthy, neurotypical, right-handed adult men. Figure 5.9 clearly illustrated that the SAAM attachment scores of the 56 participants in the studied sample were not uniformly distributed across the entire scale (1-7). This must be taken into account when studying context-dependencies of OT treatment in other populations.

Chapter 7

Conclusion

7.1 Conclusion

This thesis found significant OT-specific treatment effects on social stress regulation in adult NT men, based on a combined analysis of noninvasive measurements of blood volume pulse, electrodermal activity and respiration. These findings contribute to the research regarding exogenous oxytocin as a medical treatment for social disorders, by offering a broad overview of physiological responses to social stress in a reference group of healthy, adult NT men. New studies on different populations or physiological pathways can then be interpreted in comparison to the results reported in this thesis.

To enable the assessment of social stress regulation, a total of 70 features were extracted from each signal modality. First, classical PRV features quantified cardiovascular activity. Second, features of respiratory variability were extracted from annotations of the RSP signals. Third, two decomposition algorithms provided features of tonic and phasic activity in the EDA signals. The research questions were then answered by testing the effects of social stress and OT treatment on these features. Because of the limited sample size and large feature set, the feature analyses consisted of univariate tests, in combination with a pFDR correction. This provided an estimate of the risk of false detections in each test set, while keeping the statistical assumption requirements to a minimum.

This approach provided relevant insights in the targeted response mechanisms. OT-specific treatment effects on recovery features in the EDA aligned with prior findings reported by Daniels et al. Additionally, a new treatment effect was found in the mean respiratory rate. PRV-based power features, however, were not found to be significant, although visual inspection did suggest a tendency towards the facilitative effect expected from prior works. DDA-based features of recovery time and variability were found to confirm the anxiolytic role of OT in resting conditions, while prosocial effects on stress response were found in a smaller number of EDA and PRV features.

Additionally, these findings supported the hypothesized context-dependencies of OT-treatment and social stress regulation. Inter-subject variations in attachment style (SAAM scores) were found to be highly correlated with variations in pre-

treatment stress response and with treatment effects for some features. Despite the strong indications of such SAAM-dependent responses in this study, it was noted that these findings might not extrapolate well to other studies. Finally, some unexpected effects were thought to be caused by placebo effects, habituation to the context of the experiment, and/or unknown variations in external factors.

7.2 Future work

The findings in this thesis, as well as their limitations, indicate that there are still interesting opportunities for future research. This section suggests new or alternative approaches to address the technical challenges discussed in section 6.3, which might help to clarify the remaining uncertainties in this thesis.

Blood volume pulse Features related to PPG wave morphology could provide additional insight in the subject's cardiovascular condition. As discussed in chapter one, psychological stress can significantly affect PPG morphology. However, one must be careful to correctly distinguish acute mental stress from chronic stress patterns. S. Kontaxis proposed a model-based pulse decomposition analysis (PDA) that supports detailed quantification of PPG morphology for the assessment of ANS responses to acute stress [66]. Such an addition to the feature set could be provide more detailed insights in cardiovascular activity under influence of OT treatment and social stress.

Cardiorespiratory coupling Features of cardiorespiratory coupling should be considered given that the respiratory data is available in this dataset. The interactions between heart rate variability and respiration have been known to be good indicators of mental stress, and can be quantified according to studies such as by Morales et al. [33, 67]. Although this was already explored by J. Morales and A. Rozo in the past year, their results were not included in the feature analysis in this thesis due to time constraints.

Electrodermal activity The main challenge with features of EDA was in the strong variations of response excitability between participants. While very low phasic activity (nonresponders) can lead to invalid measures of response rate, very high phasic activity (hyperresponders) results in SCRs which can be difficult to distinguish from noise or artefacts. Since no new studies addressing this issue were found, it remains a potential point of improvement. Additionally, the inclusion of PO components as obtained by the DDA algorithm could provide relevant information.

Respiration The lack of ground truth is a clear limitation on the performance of breath annotations. With only visual inspection, it is difficult to validate the used methods. Additionally, the single breathing belt limited the use of amplitude based features. If more consistent RSP data is required in future studies, it is recommended to use a second breathing belt to record both thoracic and abdominal breathing. Additionally, if the context of the experiment allows this, it would be beneficial for

the quality of the features if the participants could control their breathing. As for alternative approaches to respiratory analysis, one recent study by Harrison et al. proposed a robust algorithm for the estimation of respiratory volume per time unit (RVT) based on the Hilbert-transform of breathing belt signals [68]. This might be an interesting alternative to the approach in this thesis, since no breath annotations are required.

Feature analysis It is important to note that the decision to only analyze the first 300 seconds of each signal has certain trade-offs, despite the benefit of allowing direct comparison between conditions of rest and stress. The main cost is the loss of information regarding the dynamics of social stress response during consecutive screen openings (intermittent eye-contact, cfr. Section 1.3). An alternative approach to the testing framework could provide more insight in these mechanisms. Similarly, the decision to apply univariate tests, albeit with pFDR correction, comes at the cost of potentially missing feature interactions. However, given the limited sample size, a multivariate testing approach could be challenging to apply correctly. Dimensionality reduction methods, such as principal components analysis (PCA), would be needed to remove excessive multicollinearities from the 70 features.

Appendices

Appendix A

Implemented features

ID	feature name	description
1	PRV PR avg	Mean pulse rate, cfr. equation 3.1
2	PRV PR std	Standard deviation of the pulse rate
3	PRV PR mad	Mean absolute deviation of the pulse rate
4	PRV TOT	Total signal power in 0-1 Hz of PRV
5	PRV LF	Low-frequency power of PRV (0.04-0.15 Hz)
6	PRV HF	High-frequency power of PRV (0.15H-0.40 Hz)
7	PRV LFn	normalized LF power of PRV (LF/TOT)
8	PRV HFn	normalized HF power of PRV (HF/TOT)
9	PRV LH ratio	ratio of LF/HF PRV power

Table A.1: BVP features, cfr. Section 3.1

ID	feature name	description
10	CDA Phasic SC Power	Power of the estimated phasic signal
11	CDA Diff2 SC Power	"Power of second difference" of SC(t) [21]
12	CDA SC slope	linear regression slope of full-length SC(t)
13	CDA Amp sum	Sum of SCR peak amplitudes, as identified by CDA
14	CDA Amp avg	Mean of SCR peak amplitudes
15	CDA Amp std	Standard deviation of SCR peak amplitudes
16	CDA Amp mad	Mean absolute deviation of SCR peak amplitudes
17	CDA IRI avg	Mean inter-response interval (time between SCRs)
18	CDA IRI std	Standard deviation of IRI
19	CDA IRI mad	Mean absolute deviation of IRI

Table A.2: CDA features, cfr. Section 3.2

ID	feature name	description
20	DDA Phasic SC Power	Power of the estimated phasic signal
21	DDA Diff2 SC Power	"Power of second difference" of SC(t) [21]
22	DDA SC slope	linear regression slope of full-length SC(t)
23	DDA Amp sum	Sum of SCR peak amplitudes
24	DDA Area sum	Sum of SCR area-under-curve values
25	DDA Amp avg	Mean of SCR peak amplitudes
26	DDA Amp std	Standard deviation of SCR peak amplitudes
27	DDA Amp mad	Mean absolute deviation of SCR peak amplitudes
28	DDA IRI avg	Mean inter-response interval (time between SCRs)
29	DDA IRI std	Standard deviation of IRI
30	DDA IRI mad	Mean absolute deviation of IRI
31	DDA Area avg	Mean SCR Area (area-under-curve)
32	DDA Area std	Standard deviation of SCR Area
33	DDA Area mad	Mean absolute deviation of SCR Area
34	DDA HalfRec time avg	Mean half-recovery time of SCRs
35	DDA HalfRec time std	Standard deviation of half-recovery time
36	DDA HalfRec time mad	Mean absolute deviation of half-recovery time
37	DDA P50 width avg	Mean width of SCRs at 50% of peak amplitude
38	DDA P50 width std	Standard deviation of P50 width
39	DDA P50 width mad	Mean absolute deviation of P50 width
40	DDA R50 slope avg	Mean linear regression slope of SCRs (peak to 50%)
41	DDA R50 slope std	Standard deviation of R50 slope
42	DDA R50 slope mad	Mean absolute deviation of R50 slope
43	DDA P50 slope avg	Mean interpolation slope of SCR (peak to 50%)
44	DDA P50 slope std	Standard deviation of P50 slope
45	DDA P50 slope mad	Mean absolute deviation of P50 slope

Table A.3: DDA features, cfr. Section 3.2

ID	feature name	description
46	RSP Ti MN	Mean inhale time (valley to peak)
47	RSP Ti SD	Standard deviation of inhale time
48	RSP Ti MAD	Mean absolute deviation of inhale time
49	RSP Ti CV	Coefficient of variation of inhale time (SD/MN)
50	RSP Ti AR	Autocorrelation at one breath lag (AR) of Ti
51	RSP Te MN	Mean exhale time (peak to valley)
52	RSP Te SD	Standard deviation of exhale time
53	RSP Te MAD	Mean absolute deviation of exhale time
54	RSP Te CV	Coefficient of variation of exhale time
55	RSP Te AR	Autocorrelation at one breath lag of Te
56	RSP RR MN	Mean respiratory rate (breaths per minute)
57	RSP RR SD	Standard deviation of respiratory rate
58	RSP RR MAD	Mean absolute deviation of respiratory rate
59	RSP RR CV	Coefficient of variation of respiratory rate
60	RSP RR AR	Autocorrelation at one breath lag of RR
61	RSP Vt MN	Mean relative tidal volume (not calibrated)
62	RSP Vt SD	Standard deviation of relative tidal volume
63	RSP Vt MAD	Mean absolute deviation of relative tidal volume
64	RSP Vt CV	Coefficient of variation of relative tidal volume
65	RSP Vt AR	Autocorrelation at one breath lag of Vt
66	RSP MV MN	Mean relative minute volume (not calibrated)
67	RSP MV SD	Standard deviation of relative minute volume
68	RSP MV MAD	Mean absolute deviation of relative minute volume
69	RSP MV CV	Coefficient of variation of relative minute volume
70	RSP MV AR	Autocorrelation at one breath lag of MV

TABLE A.4: RSP features, cfr. Section 3.3

Appendix B

Test specifications

The following tables provide a detailed overview of the univariate tests specified in chapter 4. A short description of each test is given in Table B.3. A technical overview of the corresponding implementations can then be found in Table B.4 by matching the ID numbers. A few abbreviations were used to fit the Table layout, these are provided in Tables 4.2 and B.2. The forward slash in the "Compare A/B" column defines the construction of the compared samples A and B, while the "Fixed conditions" column gives additional info on shared properties of the compared samples. For example, test 1 tests the difference between both treatment groups for the pre-treatment feature values in resting conditions.

Abbreviations	Construction	Description
ftr	as is	direct feature values
dftr	$= ftr(post) - ftr(pre)$	treatment effect on ftr
bcsr	$= ftr(stress) - ftr(rest)$	baseline-corrected stress response
dbcsr	$= bcsr(post) - bcsr(pre)$	treatment effect on bcsr

Table B.1: Overview of sample abbreviations used for Table B.4

Abbreviations	Explanation
all	all subjects included
oxt	oxytocin group
plc	placebo group
pre	pre-treatment
post	post-treatment
rest	screen closed, no eye-contact over entire recording
stress	eye-contact on intermittent screen openings

Table B.2: Overview of indexing abbreviations used for Table B.4

Description	ID
Pre-treatment similarity assumption checks	
in resting conditions	1
in stress conditions	2
in baseline-corrected stress response (*bcsr*)	3
in post-treatment *bcsr* values for comparison	4
Differentiative value of features when measuring stress response	
at baseline (pre-treatment) for all subjects	5
after oxytocin treatment	6
after placebo treatment	7
Significance of measured treatment effects on ANS activity	
effects of oxytocin treatment on activity at rest	8
effects of placebo treatment on activity at rest	9
effects of oxytocin treatment on activity under stress	10
effects of placebo treatment on activity under stress	11
effects of oxytocin treatment on *bcsr*	12
effects of placebo treatment on *bcsr*	13
Compare treatment effects of oxytocin and placebo	
different effects on activity in resting state?	14
different effects on activity in stress state?	15
different effects on *bcsr*?	16
Do the SAAM scores correlate with pre-treatment response variations?	
in rest conditions	17
in stress conditions	18
in *bcsr*	19
Do the SAAM scores correlate with variations in treatment effects?	
for oxytocin group, in rest conditions	20
for oxytocin group, in stress conditions	21
for oxytocin group, in *bcsr*	22
for placebo group, in rest conditions	23
for placebo group, in stress conditions	24
for placebo group, in *bcsr*	25

Table B.3: Descriptive summary of all univariate tests

ID	Dataset	Fixed conditions	Compare A/B	Test(A/B)
1	ftr	pre, rest	oxt/plc	ranksum
2	ftr	pre, stress	oxt/plc	ranksum
3	bcsr	pre	oxt/plc	ranksum
4	bcsr	post	oxt/plc	ranksum
5	ftr	all, pre	rest/stress	signrank
6	ftr	oxt, post	rest/stress	signrank
7	ftr	plc, post	rest/stress	signrank
8	ftr	oxt, rest	pre/post	signrank
9	ftr	plc, rest	pre/post	signrank
10	ftr	oxt, stress	pre/post	signrank
11	ftr	plc, stress	pre/post	signrank
12	bcsr	oxt	pre/post	signrank
13	bcsr	plc	pre/post	signrank
14	dftr	rest	oxt/plc	ranksum
15	dftr	stress	oxt/plc	ranksum
16	dbcsr	-	oxt/plc	ranksum
17	ftr	rest	pre/SAAM	Spearman
18	ftr	stress	pre/SAAM	Spearman
19	bcsr	-	pre/SAAM	Spearman
20	dftr	rest	oxt/SAAM	Spearman
21	dftr	stress	oxt/SAAM	Spearman
22	dbcsr	-	oxt/SAAM	Spearman
23	dftr	rest	plc/SAAM	Spearman
24	dftr	stress	plc/SAAM	Spearman
25	dbcsr	-	plc/SAAM	Spearman

TABLE B.4: Technical summary of all univariate test implementations

Appendix C

Numerical test results

	normality (p-value)		missing values		Box-Cox transformation	
Feature	original	transformed	invalid	outliers	λ	offset (+)
PRV PR avg	0.0010	0.1788	17	2	0.0000	0
PRV PR std	0.0010	0.4618	17	1	0.0000	0
PRV PR mad	0.0010	0.5000	17	1	0.0000	0
PRV TOT	0.0010	0.1266	17	1	0.0000	0
PRV LF	0.0010	0.1380	17	0	0.0000	0
PRV HF	0.0010	0.4171	17	1	0.0000	0
PRV LH ratio	0.0010	0.5000	17	0	0.0000	0
CDA Phasic SC Power	0.0010	0.5000	0	0	0.0863	0
CDA Diff2 SC Power	0.0010	0.0525	0	2	-0.1518	0
CDA Amp sum	0.0010	0.0717	0	0	-0.1596	1
CDA Amp avg	0.0010	0.5000	8	3	0.0000	0
CDA IRI avg	0.0010	0.0502	14	1	0.0000	0
CDA IRI std	0.0010	0.4673	14	0	0.0000	1
CDA IRI mad	0.0010	0.2030	14	0	0.0000	1
DDA Phasic SC Power	0.0010	0.5000	0	0	0.1029	0
DDA Diff2 SC Power	0.0010	0.0525	0	2	-0.1518	0
DDA Amp sum	0.0010	0.3280	0	0	-0.1819	1
DDA Area sum	0.0010	0.5000	0	0	0.2058	1
DDA Amp avg	0.0010	0.0997	9	4	0.0000	0
DDA IRI avg	0.0010	0.5000	17	1	0.0000	0
DDA IRI std	0.0010	0.5000	17	1	0.0000	1
DDA IRI mad	0.0010	0.5000	17	2	0.0000	1
DDA Area std	0.0010	0.0906	9	0	0.0000	1
RSP Ti MN	0.0010	0.5000	0	0	-2.3529	0
RSP Ti MAD	0.0010	0.0783	0	0	0.0672	0
RSP Ti AR	0.0081	0.5000	0	0	-1.4193	2
RSP Te MN	0.0010	0.5000	0	0	-0.9275	0
RSP Te SD	0.0010	0.5000	0	0	0.0623	0
RSP Te MAD	0.0010	0.5000	0	0	0.0292	0
RSP Te CV	0.0010	0.1449	0	0	-0.0723	0
RSP Te AR	0.3041	0.4407	0	0	0.3563	2
RSP RR MN	0.0109	0.5000	0	1	1.7134	0
RSP RR SD	0.0012	0.1264	0	1	0.1591	0
RSP RR MAD	0.0010	0.3031	0	1	0.0339	0
RSP RR CV	0.0010	0.3484	0	0	0.0386	0
RSP RR AR	0.5000	0.5000	0	0	0.8359	2
RSP Vt MN	0.0010	0.0564	0	0	-0.1021	0
RSP Vt SD	0.0010	0.0604	0	0	0.1493	0
RSP Vt MAD	0.0010	0.1034	0	0	0.0362	0
RSP Vt AR	0.0261	0.5000	0	0	-0.8860	2
RSP MV MN	0.0010	0.2017	0	0	-0.0658	0
RSP MV SD	0.0010	0.1251	0	0	0.0343	0
RSP MV MAD	0.0010	0.1530	0	0	-0.0671	0
RSP MV AR	0.3719	0.5000	0	1	0.3333	2

Table C.1: Normalized features (before and after transformations)

	all		oxytocin		placebo	
feature	p	q	p	q	p	q
CDA Diff2 SC Power	**0.0000**	0.0002	**0.0003**	0.0000	**0.0003**	0.0045
DDA Diff2 SC Power	**0.0000**	0.0002	**0.0003**	0.0000	**0.0003**	0.0045
DDA Amp std	**0.0000**	0.0693	**0.0038**	0.0001	**0.0139**	0.0234
DDA Amp mad	**0.0000**	0.1500	**0.0029**	0.0001	**0.0188**	0.0225
CDA Amp avg	**0.0000**	0.0633	**0.0093**	0.0002	**0.0139**	0.0285
DDA Amp avg	**0.0001**	0.1888	**0.0022**	0.0002	**0.0206**	0.0225
CDA Amp sum	**0.0001**	0.0694	**0.0059**	0.0002	**0.0139**	0.0242
CDA Phasic SC Power	**0.0001**	0.6309	**0.0063**	0.0005	**0.0372**	0.0242
RSP RR MN	**0.0002**	0.1494	**0.0086**	0.0006	**0.0188**	0.0285
CDA Amp std	**0.0003**	0.0164	**0.0115**	0.0011	**0.0123**	0.0320
RSP Vt MN	**0.0004**	0.8664	0.1014	0.0012	**0.0402**	0.1287
CDA Amp mad	**0.0005**	0.0066	**0.0050**	0.0014	**0.0067**	0.0242
DDA P50 slope mad	**0.0008**	0.0827	0.0578	0.0021	**0.0155**	0.0843
DDA R50 slope std	**0.0009**	0.0980	**0.0370**	0.0022	**0.0164**	0.0736
DDA P50 slope std	**0.0011**	0.0693	**0.0422**	0.0024	**0.0139**	0.0750
DDA R50 slope mad	**0.0012**	0.1425	**0.0480**	0.0024	**0.0188**	0.0750
RSP Te MN	**0.0017**	0.1635	**0.0246**	0.0034	**0.0189**	0.0618
DDA Amp sum	**0.0028**	0.9250	**0.0263**	0.0052	**0.0403**	0.0618
RSP Ti MN	**0.0033**	0.4860	0.2692	0.0058	**0.0353**	0.1962
DDA P50 slope avg	**0.0044**	0.1218	0.1829	0.0073	**0.0178**	0.1646
RSP Vt CV	**0.0078**	0.4860	**0.0490**	0.0125	**0.0353**	0.0750
CDA IRI avg	**0.0095**	0.6265	**0.0345**	0.0146	**0.0372**	0.0736
DDA R50 slope avg	**0.0134**	0.1919	0.2418	0.0196	**0.0206**	0.1928
RSP RR SD	**0.0304**	0.7186	0.1183	0.0425	**0.0389**	0.1341
PRV PR std	**0.0337**	0.7317	0.2418	0.0453	**0.0389**	0.1928
RSP RR MAD	**0.0392**	0.5642	0.1513	0.0507	**0.0361**	0.1543
PRV PR mad	0.0608	0.6071	0.2758	0.0757	**0.0372**	0.1963
RSP MV MN	0.0645	0.5642	0.2478	0.0774	**0.0361**	0.1928
RSP Te SD	0.0714	0.9044	0.6938	0.0827	**0.0402**	0.3612
RSP Vt AR	0.1329	0.3740	0.4692	0.1489	**0.0315**	0.2709
RSP Te MAD	0.1528	0.9044	0.8093	0.1642	**0.0402**	0.3994
DDA HalfRec time std	0.1563	0.0347	0.8824	0.1642	**0.0139**	0.4154
DDA Area std	0.1683	0.2528	0.0875	0.1714	**0.0254**	0.1217
RSP RR AR	0.1777	0.1184	0.5506	0.1737	**0.0178**	0.3009
DDA Area mad	0.1809	0.2758	**0.0480**	0.1737	**0.0268**	0.0750
PRV PR avg	0.1977	0.2087	0.1374	0.1744	**0.0216**	0.1501
PRV HFn	0.1977	0.9544	0.3674	0.1744	**0.0410**	0.2331
RSP Te AR	0.2023	0.1241	0.1068	0.1744	**0.0178**	0.1287
RSP MV CV	0.2023	0.3871	0.2584	0.1744	**0.0315**	0.1928
DDA Area sum	0.2091	0.8612	**0.0385**	0.1757	**0.0402**	0.0736
DDA Area avg	0.2346	0.3819	0.1094	0.1923	**0.0315**	0.1287
PRV TOT	0.2805	0.5485	0.2312	0.2227	**0.0361**	0.1928
PRV LF	0.2849	0.9090	0.1425	0.2227	**0.0402**	0.1504
DDA HalfRec time mad	0.2940	0.0370	0.7570	0.2246	**0.0139**	0.3861
CDA IRI mad	0.3250	0.2871	0.3758	0.2387	**0.0270**	0.2331
RSP Vt MAD	0.3298	0.7186	0.2376	0.2387	**0.0389**	0.1928
RSP RR CV	0.3390	0.8854	0.3809	0.2387	**0.0402**	0.2331
CDA IRI std	0.3409	0.1618	0.4237	0.2387	**0.0189**	0.2542
DDA P50 width std	0.4006	0.0875	0.9464	0.2748	**0.0155**	0.4197
DDA P50 width mad	0.4526	0.0615	0.9250	0.3043	**0.0139**	0.4181
PRV LH ratio	0.4831	0.9090	0.5098	0.3184	**0.0402**	0.2836
DDA IRI std	0.5443	0.8552	0.7971	0.3518	**0.0402**	0.3994
DDA P50 width avg	0.5946	0.0480	0.4926	0.3771	**0.0139**	0.2791
RSP Ti SD	0.6165	0.7366	0.9292	0.3837	**0.0389**	0.4181
RSP MV SD	0.6358	0.7366	0.3037	0.3886	**0.0389**	0.2112
RSP Ti AR	0.6954	0.5165	0.4386	0.4174	**0.0361**	0.2581
RSP Te CV	0.7430	0.3246	0.1663	0.4272	**0.0287**	0.1615
DDA IRI mad	0.7477	0.6265	0.9090	0.4272	**0.0372**	0.4181
RSP MV MAD	0.7499	0.7186	0.2584	0.4272	**0.0389**	0.1928
RSP Ti CV	0.7638	0.5642	0.3539	0.4279	**0.0361**	0.2304
PRV LFn	0.8152	0.7533	0.3533	0.4492	**0.0391**	0.2304
PRV HF	0.8307	0.4929	0.1094	0.4503	**0.0353**	0.1287
CDA SC slope	0.8626	0.0609	0.1742	0.4530	**0.0139**	0.1615
DDA SC slope	0.8626	0.0609	0.1742	0.4530	**0.0139**	0.1615
DDA HalfRec time avg	0.8932	0.0230	0.6964	0.4616	**0.0138**	0.3612
DDA IRI avg	0.9087	0.4842	0.8314	0.4616	**0.0353**	0.4038
RSP Ti MAD	0.9202	0.9044	0.5677	0.4616	**0.0402**	0.3048
DDA Phasic SC Power	0.9347	0.3016	0.8689	0.4620	**0.0275**	0.4154
RSP MV AR	0.9701	0.4564	0.9899	0.4726	**0.0353**	0.4327
RSP Vt SD	1.0000	0.8476	0.3282	0.4802	**0.0402**	0.2232

TABLE C.2: Features which measured significant differences between activity levels in rest and stress, for the full sample before treatment (all) and the separate groups after administration of oxytocin and placebo.

	rest, oxt		rest, plc		stress, oxt		stress, plc	
feature	p	q	p	q	p	q	p	q
DDA HalfRec time avg	**0.0016**	0.0472	0.6385	0.3560	0.5485	0.9040	0.5998	0.2143
PRV PR avg	**0.0017**	0.0472	**0.0003**	0.0084	**0.0023**	0.1447	**0.0000**	0.0005
DDA HalfRec time std	**0.0042**	0.0604	0.7127	0.3560	0.8864	0.9040	0.9464	0.2315
DDA HalfRec time mad	**0.0045**	0.0604	0.9899	0.3715	0.7103	0.9040	0.6766	0.2143
DDA P50 width std	**0.0054**	0.0604	0.8489	0.3564	0.5485	0.9040	0.6571	0.2143
RSP Vt MN	**0.0077**	0.0702	**0.0019**	0.0170	**0.0164**	0.4543	**0.0092**	0.0195
DDA P50 width mad	**0.0087**	0.0702	0.5009	0.3557	0.3914	0.9040	0.5449	0.2143
DDA P50 width avg	**0.0128**	0.0904	0.2919	0.3485	0.5677	0.9040	0.7366	0.2143
RSP MV MN	**0.0190**	0.1165	**0.0019**	0.0170	**0.0319**	0.4543	**0.0026**	0.0110
PRV PR std	**0.0207**	0.1165	**0.0247**	0.1080	**0.0370**	0.4543	**0.0087**	0.0195
PRV PR mad	**0.0487**	0.2496	**0.0450**	0.1425	0.0633	0.4543	**0.0186**	0.0348
PRV TOT	0.0633	0.2975	0.0736	0.1611	0.0765	0.4543	**0.0054**	0.0178
PRV LF	0.0865	0.3753	0.0615	0.1611	0.1829	0.8370	**0.0063**	0.0178
RSP Vt CV	0.1943	0.4870	**0.0186**	0.0975	0.2376	0.9040	**0.0007**	0.0055
CDA Diff2 SC Power	0.1943	0.4870	0.8854	0.3564	0.0693	0.4543	**0.0263**	0.0369
DDA Diff2 SC Power	0.1943	0.4870	0.8854	0.3564	0.0693	0.4543	**0.0263**	0.0369
PRV HF	0.2652	0.5337	0.2528	0.3485	0.3606	0.9040	**0.0010**	0.0055
RSP MV CV	0.3389	0.5974	**0.0186**	0.0975	**0.0409**	0.4543	**0.0246**	0.0369
RSP Ti AR	0.8734	0.7892	**0.0388**	0.1425	0.3809	0.9040	0.2277	0.1893
RSP RR AR	0.8913	0.7892	**0.0488**	0.1425	0.3949	0.9040	**0.0490**	0.0636

Table C.3: All features which measured significant treatment effects of oxytocin (oxt) or placebo (plc) on activity levels in rest or stress.

Feature	sample	SAAM	p-value	q-value	ρ
RSP RR MN	rest	security	0.0019	0.1284	-0.4135
RSP Te MN	rest	security	0.0068	0.2299	0.3611
RSP Ti MN	rest	security	0.0126	0.2861	0.3343
CDA IRI std	stress	anxiety	0.0181	0.0756	0.3399
CDA Amp avg	stress	anxiety	0.0182	0.0756	-0.3395
CDA IRI mad	stress	anxiety	0.0243	0.0756	0.3249
CDA IRI avg	stress	anxiety	0.0256	0.0756	0.3221
RSP RR MN	stress	security	0.0292	0.4267	-0.3026
RSP Te MN	stress	security	0.0304	0.4267	0.3005
CDA Amp std	stress	security	0.0307	0.4267	-0.3091
DDA Amp mad	stress	avoidance	0.0270	0.8480	0.3129
DDA Amp avg	stress	avoidance	0.0360	0.8480	0.3004
DDA Amp std	stress	avoidance	0.0365	0.8480	0.2966
RSP Te MN	bcsr	anxiety	0.0071	0.0239	0.3688
RSP Te MAD	bcsr	anxiety	0.0243	0.0239	0.3120
RSP Te SD	bcsr	anxiety	0.0420	0.0239	0.2831
RSP RR MN	bcsr	anxiety	0.0460	0.0239	-0.2807
CDA IRI avg	bcsr	anxiety	0.0461	0.0239	0.2924
RSP Vt CV	bcsr	security	0.0362	0.3834	0.2912
DDA Amp avg	bcsr	avoidance	0.0303	0.9894	0.3129

TABLE C.4: Significant correlations between SAAM scores and pre-treatment data

Feature	sample	SAAM	p-value	q-value	ρ
RSP RR SD	rest	anxiety	0.0174	0.6724	-0.4461
RSP RR CV	rest	anxiety	0.0304	0.6724	-0.4096
RSP Te CV	rest	anxiety	0.0417	0.6724	-0.3873
RSP RR MAD	rest	anxiety	0.0496	0.6724	-0.3744
RSP MV AR	rest	security	0.0354	0.2326	-0.4064
RSP RR AR	rest	avoidance	0.0229	0.9703	-0.4285
CDA IRI std	stress	anxiety	0.0219	0.2601	0.4752
CDA IRI mad	stress	anxiety	0.0280	0.2601	0.4579
CDA SC slope	stress	security	0.0081	0.0121	-0.5076
DDA SC slope	stress	security	0.0081	0.0121	-0.5076
DDA HalfRec time avg	stress	security	0.0219	0.0211	0.4655
DDA HalfRec time mad	stress	security	0.0283	0.0211	0.4475
DDA P50 width avg	stress	security	0.0410	0.0211	0.4200
DDA HalfRec time std	stress	security	0.0485	0.0211	0.4068
CDA SC slope	stress	avoidance	0.0177	0.6149	0.4611
DDA SC slope	stress	avoidance	0.0177	0.6149	0.4611
RSP Te MAD	bcsr	anxiety	0.0281	0.8987	0.4306
DDA IRI std	bcsr	security	0.0381	0.9802	-0.4665
DDA IRI mad	bcsr	security	0.0487	0.9802	-0.4461

Table C.5: Significant correlations between SAAM scores and OT effects

Feature	sample	SAAM	p-value	q-value	ρ
DDA HalfRec time std	rest	anxiety	0.0078	0.4633	0.5102
DDA HalfRec time mad	rest	anxiety	0.0134	0.4633	0.4785
DDA HalfRec time avg	rest	anxiety	0.0387	0.8922	0.4076
DDA R50 slope mad	rest	avoidance	0.0177	0.3711	-0.4613
DDA P50 slope mad	rest	avoidance	0.0213	0.3711	-0.4493
RSP Te MN	rest	avoidance	0.0245	0.3711	-0.4319
DDA P50 slope avg	rest	avoidance	0.0287	0.3711	0.4291
DDA P50 slope std	rest	avoidance	0.0329	0.3711	-0.4195
DDA R50 slope std	rest	avoidance	0.0433	0.4073	-0.3992
CDA Amp avg	stress	anxiety	0.0002	0.0019	0.7211
CDA Amp mad	stress	anxiety	0.0182	0.0328	0.4779
DDA Area sum	stress	anxiety	0.0276	0.0328	0.4317
DDA Amp sum	stress	anxiety	0.0288	0.0328	0.4290
CDA Amp std	stress	anxiety	0.0288	0.0328	0.4464
CDA Amp sum	stress	anxiety	0.0304	0.0328	0.4252
CDA IRI mad	stress	anxiety	0.0327	0.0328	-0.4465
CDA IRI avg	stress	anxiety	0.0338	0.0328	-0.4440
CDA Phasic SC Power	stress	anxiety	0.0343	0.0328	0.4166
PRV HFn	stress	anxiety	0.0483	0.0416	-0.3988
PRV LF	stress	security	0.0054	0.3122	-0.5397
PRV TOT	stress	security	0.0108	0.3151	-0.5005
CDA SC slope	stress	security	0.0247	0.3373	0.4393
DDA SC slope	stress	security	0.0247	0.3373	0.4393
RSP RR MN	stress	security	0.0341	0.3373	0.4169
PRV HF	stress	security	0.0348	0.3373	-0.4237
DDA Amp avg	stress	avoidance	0.0253	0.0789	-0.4555
RSP Vt CV	stress	avoidance	0.0263	0.0789	-0.4351
DDA P50 width avg	stress	avoidance	0.0333	0.0789	0.4269
DDA HalfRec time avg	bc-stress	anxiety	0.0226	0.2864	-0.4542
RSP Vt SD	bc-stress	anxiety	0.0262	0.2864	0.4355
CDA SC slope	bc-stress	anxiety	0.0288	0.2864	0.4290
DDA SC slope	bc-stress	anxiety	0.0288	0.2864	0.4290
CDA Amp sum	bc-stress	anxiety	0.0333	0.2864	0.4187
CDA IRI avg	bc-stress	anxiety	0.0364	0.2864	-0.4385
CDA Amp avg	bc-stress	anxiety	0.0366	0.2864	0.4585
DDA HalfRec time std	bc-stress	anxiety	0.0473	0.3244	-0.4004
CDA SC slope	bc-stress	security	0.0385	0.9839	0.4080
DDA SC slope	bc-stress	security	0.0385	0.9839	0.4080
DDA Amp avg	bc-stress	avoidance	0.0215	0.9854	-0.4668
RSP Vt CV	bc-stress	avoidance	0.0496	0.9854	-0.3888

TABLE C.6: Significant correlations between SAAM scores and placebo effects

Bibliography

[1] Mathias Benedek and Christian Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1):80–91, June 2010.

[2] Mathias Benedek and Christian Kaernbach. Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, March 2010.

[3] John D. Storey. The positive false discovery rate: a Bayesian interpretation and the q-value. *The Annals of Statistics*, 31(6), December 2003.

[4] Jellina Prinsen, Stephanie Brams, and Kaat Alaerts. To mirror or not to mirror upon mutual gaze, oxytocin can pave the way: A cross-over randomized placebo-controlled trial. *Psychoneuroendocrinology*, 90:148–156, April 2018.

[5] Nicky Daniels, Jellina Prinsen, Javier R. Soriano, and Kaat Alaerts. Oxytocin enhances the recovery of eye-contact induced autonomic arousal: A treatment mechanism study with placebo-controlled design. *European Neuropsychopharmacology*, 39:87–98, October 2020.

[6] N. Daniels, J. R. Soriano, J. Prinsen, and K. Alaerts. P.310 Effects of intranasal administration of the neuromodulator oxytocin on autonomic cardiac function. *European Neuropsychopharmacology*, 44:S45–S46, March 2021.

[7] Kaat Alaerts, Sylvie Bernaerts, Bart Vanaudenaerde, Nicky Daniels, and Nicole Wenderoth. Amygdala–Hippocampal Connectivity Is Associated With Endogenous Levels of Oxytocin and Can Be Altered by Exogenously Administered Oxytocin in Adults With Autism. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 4(7):655–663, July 2019.

[8] Sylvie Bernaerts, Bart Boets, Guy Bosmans, Jean Steyaert, and Kaat Alaerts. Behavioral effects of multiple-dose oxytocin treatment in autism: a randomized, placebo-controlled trial with long-term follow-up. *Molecular Autism*, 11(1):6, January 2020.

[9] M. J. Bakermans-Kranenburg and M. H. van IJzendoorn. Sniffing around oxytocin: review and meta-analyses of trials in healthy and clinical groups with implications for pharmacotherapy. *Translational Psychiatry*, 3(5):e258–e258, May 2013.

[10] Marinus H. Van IJzendoorn and Marian J. Bakermans-Kranenburg. A sniff of trust: meta-analysis of the effects of intranasal oxytocin administration on face recognition, trust to in-group, and trust to out-group. *Psychoneuroendocrinology*, 37(3):438–443, March 2012.

[11] Jennifer A. Bartz, Jamil Zaki, Niall Bolger, and Kevin N. Ochsner. Social effects of oxytocin in humans: context and person matter. *Trends in Cognitive Sciences*, 15(7):301–309, July 2011.

[12] Adam J. Graustella and Colin MacLeod. A critical review of the influence of oxytocin nasal spray on social cognition in humans: Evidence and future directions. *Hormones and Behavior*, 61(3):410–418, March 2012.

[13] G.C. Armsden and M.T. Greenberg. The inventory of parent and peer attachment: Individual differences and their relationship to psychological well-being in adolescence. *Journal of Youth and Adolescence*, 16(5):427–454, 1987.

[14] Omri Gillath, Joshua Hart, Erik E. Noftle, and Gary D. Stockdale. Development and validation of a state adult attachment measure (SAAM). *Journal of Research in Personality*, 43(3):362–373, June 2009.

[15] J.N. Constantino, S.A. Davis, R.D. Todd, M.K. Schindler, M.M. Gross, S.L. Brophy, L.M. Metzger, C.S. Shoushtari, R. Splinter, and W. Reich. Validation of a brief quantitative measure of autistic traits: Comparison of the social responsiveness scale with the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders*, 33(4):427–433, 2003.

[16] Kristen S. L. Lam and Michael G. Aman. The Repetitive Behavior Scale-Revised: Independent Validation in Individuals with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 37(5):855–866, May 2007.

[17] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-IV*. American Psychiatric Association, Washington (D.C.), 4th ed., 1st, 3rd print. edition, 1994.

[18] Kei Ohashi, Yoshifumi Mizuno, Taishi Miyachi, Tomoko Asai, Masayuki Imaeda, and Shinji Saitoh. Concordance of DSM-5 and DSM-IV-TR classifications for autism spectrum disorder. *Pediatrics International*, 57(6):1097–1100, 2015.

[19] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5*. American Psychiatric Association, Washington (D.C.), fifth edition, text revision edition, 2013.

[20] Michael G. Ziegler. Psychological Stress and the Autonomic Nervous System. In *Primer on the Autonomic Nervous System*, pages 291–293. Elsevier, 2012.

[21] Bernard Grundlehner, Lindsay Brown, Julien Penders, and Bert Gyselinckx. The Design and Analysis of a Real-Time, Continuous Arousal Monitor. In

2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks, pages 156–161, June 2009. ISSN: 2376-8894.

[22] Elke Vlemincx, Ilse Van Diest, and Omer Van den Bergh. Emotion, sighing, and respiratory variability. *Psychophysiology*, 52(5):657–666, 2015.

[23] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Troster, and U. Ehlert. Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):410–417, March 2010.

[24] Jacqueline Wijsman, Bernard Grundlehner, Hao Liu, Hermie Hermens, and Julien Penders. Towards mental stress detection using wearable physiological sensors. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1798–1801, August 2011. ISSN: 1558-4615.

[25] Elena Smets, Giuseppina Schiavone, Emmanuel Rios Velazquez, Walter De Raedt, Katleen Bogaerts, Ilse Van Diest, and Chris Van Hoof. Comparing task-induced psychophysiological responses between persons with stress-related complaints and healthy controls: A methodological pilot study. *Health Science Reports*, 1(8):e60, August 2018.

[26] Dorien Huysmans, Elena Smets, Walter De Raedt, Chris Van Hoof, Katleen Bogaerts, Ilse Van Diest, and Denis Helic. Unsupervised Learning for Mental Stress Detection:. In *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 26–35, Funchal, Madeira, Portugal, 2018. SCITEPRESS - Science and Technology Publications.

[27] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3):R1, February 2007.

[28] Michelle Hickey, Justin P. Phillips, and Panayiotis A. Kyriacou. Investigation of peripheral photoplethysmographic morphology changes induced during a hand-elevation study. *Journal of Clinical Monitoring and Computing*, 30(5):727–736, October 2016.

[29] Azadeh Kushki, Jillian Fairley, Satyam Merja, Gillian King, and Tom Chau. Comparison of blood volume pulse and skin conductance responses to mental and affective stimuli at different anatomical sites. *Physiological Measurement*, 32(10):1529–1539, October 2011.

[30] Wolfram Boucsein. Principles of Electrodermal Phenomena. In Wolfram Boucsein, editor, *Electrodermal Activity*, pages 1–86. Springer US, Boston, MA, 2012.

[31] Wolfram Boucsein. Methods of Electrodermal Recording. In Wolfram Boucsein, editor, *Electrodermal Activity*, pages 87–258. Springer US, Boston, MA, 2012.

[32] Victoria B. Egizio, Michael Eddy, Matthew Robinson, and J. Richard Jennings. Efficient and Cost-effective Estimation of the Influence of Respiratory Variables on Respiratory Sinus Arrhythmia. *Psychophysiology*, 48(4):488–494, April 2011.

[33] John Morales, Jonathan Moeyersons, Pablo Armanac, Michele Orini, Luca Faes, Sebastiaan Overeem, Merel Van Gilst, Johannes Van Dijk, Sabine Van Huffel, Raquel Bailon, and Carolina Varon. Model-Based Evaluation of Methods for Respiratory Sinus Arrhythmia Estimation. *IEEE Transactions on Biomedical Engineering*, 68(6):1882–1893, June 2021.

[34] Cornelis J. E. Wientjes, Paul Grossman, and Anthony W. K. Gaillard. Influence of drive and timing mechanisms on breathing pattern and ventilation during mental task performance. *Biological Psychology*, 49(1):53–70, September 1998.

[35] Frank Wilhelm, Walton Roth, and Marvin Sackner. The LifeShirt An Advanced System for Ambulatory Measurement of Respiratory and Cardiac Function. *Behavior modification*, 27:671–91, November 2003.

[36] K. Konno and J. Mead. Measurement of the separate volume changes of rib cage and abdomen during breathing. *Journal of Applied Physiology*, 22(3):407–422, March 1967.

[37] Kristofer Montazeri, Sigurdur Aegir Jonsson, Jon Skirnir Agustsson, Marta Serwatko, Thorarinn Gislason, and Erna S. Arnardottir. The design of RIP belts impacts the reliability and quality of the measured respiratory signals. *Sleep and Breathing*, January 2021.

[38] M. A. Sackner, H. Watson, A. S. Belsito, D. Feinerman, M. Suarez, G. Gonzalez, F. Bizousky, and B. Krieger. Calibration of respiratory inductive plethysmograph during natural breathing. *Journal of Applied Physiology*, 66(1):410–420, January 1989.

[39] Jonathan D. Power, Charles J. Lynch, Marc J. Dubin, Benjamin M. Silver, Alex Martin, and Rebecca M. Jones. Characteristics of respiratory measures in young adults scanned at rest, including systematic changes and "missed" deep breaths. *NeuroImage*, 204:116234, January 2020.

[40] Marek Malik, J. Thomas Bigger, A. John Camm, Robert E. Kleiger, Alberto Malliani, Arthur J. Moss, and Peter J. Schwartz. Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17(3):354–381, March 1996.

[41] Jesus Lazaro, Eduardo Gil, Jose Maria Vergara, and Pablo Laguna. Pulse Rate Variability Analysis for Discrimination of Sleep-Apnea-Related Decreases in the Amplitude Fluctuations of Pulse Photoplethysmographic Signal in Children. *IEEE Journal of Biomedical and Health Informatics*, 18(1):240–246, January 2014.

[42] Mohamed Elgendi, Björn Eskofier, Socrates Dokos, and Derek Abbott. Revisiting QRS Detection Methodologies for Portable, Wearable, Battery-Operated, and Wireless ECG Systems. *PLOS ONE*, 9(1):e84018, January 2014.

[43] Jonathan Moeyersons, Matthew Amoni, Sabine Van Huffel, Rik Willems, and Carolina Varon. R-DECO: an open-source Matlab based graphical user interface for the detection and correction of R-peaks. *PeerJ Computer Science*, 5:e226, October 2019.

[44] Goldberger Ary L., Amaral Luis A. N., Glass Leon, Hausdorff Jeffrey M., Ivanov Plamen Ch., Mark Roger G., Mietus Joseph E., Moody George B., Peng Chung-Kang, and Stanley H. Eugene. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, 101(23):e215–e220, June 2000.

[45] Carolina Varon, Alexander Caicedo, Dries Testelmans, Bertien Buyse, and Sabine Van Huffel. A Novel Algorithm for the Automatic Detection of Sleep Apnea From Single-Lead ECG. *IEEE Transactions on Biomedical Engineering*, 62(9):2269–2278, September 2015.

[46] Elena Peralta, Jesus Lazaro, Raquel Bailon, Vaidotas Marozas, and Eduardo Gil. Optimal fiducial points for pulse rate variability analysis from forehead and finger photoplethysmographic signals. *Physiological Measurement*, 40(2):025007, February 2019.

[47] Wolfram Boucsein. Applications of Electrodermal Recording. In Wolfram Boucsein, editor, *Electrodermal Activity*, pages 259–523. Springer US, Boston, MA, 2012.

[48] T. Mano. Microneurography as a tool to investigate sympathetic nerve responses to environmental stress. *Aviakosmicheskaia I Ekologicheskaia Meditsina = Aerospace and Environmental Medicine*, 31(1):8–14, 1997.

[49] V. G. Macefield, B. G. Wallin, and A. B. Vallbo. The discharge behaviour of single vasoconstrictor motoneurones in human muscle nerves. *The Journal of Physiology*, 481(3):799–809, 1994.

[50] Robert Edelberg. Electrodermal Mechanisms: A Critique of the Two-Effector Hypothesis and a Proposed Replacement. In Jean-Claude Roy, Wolfram Boucsein, Don C. Fowles, and John H. Gruzelier, editors, *Progress in Electrodermal Research*, NATO ASI Series, pages 7–29. Springer US, Boston, MA, 1993.

[51] Torben Noto, Guangyu Zhou, Stephan Schuele, Jessica Templer, and Christina Zelano. Automated analysis of breathing waveforms using BreathMetrics: a respiratory signal processing toolbox. *Chemical Senses*, 43(8):583–597, September 2018.

[52] Jesus Lazaro, Eduardo Gil, Margot Deviaene, Raquel Bailon, Dries Testelmans, Bertien Buyse, Carolina Varon, and Sabine Van Huffel. Pulse Photoplethysmography Derived Respiration for Obstructive Sleep Apnea Detection. In *44th annual Computing in Cardiology Conference*, September 2017.

[53] Elvan Tabachnik, Nestor Muller, Baldwin Toye, and Henry Levinson. Measurement of ventilation in children using the respiratory inductive plethysmograph. *The Journal of Pediatrics*, 99(6):895–899, December 1981.

[54] J. P. Verma and Abdel-Salam G. Abdel-Salam. *Testing Statistical Assumptions in Research*. John Wiley & Sons, Incorporated, Newark, UNITED STATES, 2019.

[55] Alfred DeMaris and Steven H. Selman. *Converting Data into Evidence*. Springer New York, New York, NY, 2013.

[56] Gerald Van Belle, Lloyd Fisher, Patrick J. Heagerty, and Thomas Lumley. *Biostatistics: a methodology for the health sciences*. Wiley series in probability and statistics. John Wiley & Sons, Hoboken, NJ, 2nd ed edition, 2004.

[57] G. E. P. Box and D. R. Cox. An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252, 1964.

[58] Jelle J. Goeman and Aldo Solari. Multiple hypothesis testing in genomics. *Statistics in Medicine*, 33(11):1946–1978, 2014.

[59] Salvador García, Alberto Fernández, Julián Luengo, and Francisco Herrera. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences*, 180(10):2044–2064, May 2010.

[60] Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[61] Maximilian Christ, Andreas W. Kempa-Liehr, and Michael Feindt. Distributed and parallel time series feature extraction for industrial big data applications. *CoRR*, abs/1610.07717, 2016.

[62] John D. Storey and Robert Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16):9440–9445, August 2003.

[63] Tae Kyun Kim. T test as a parametric statistic. *Korean Journal of Anesthesiology*, 68(6):540–546, November 2015.

[64] Jean D. Gibbons and Subhabrata Chakraborti. *Nonparametric Statistical Inference*. Chapman & Hall/CRC, 5th edition, 2011.

[65] Steve McKillup. *Non-parametric tests for ratio, interval or ordinal scale data*, page 319–345. Cambridge University Press, 2 edition, 2011.

[66] Spyridon Kontaxis, Eduardo Gil, Vaidotas Marozas, Jesús Lázaro, Esther García, Mar Posadas-de Miguel, Sara Siddi, Maria Luisa Bernal, Jordi Aguiló, Josep Maria Haro, Concepción de la Cámara, Pablo Laguna, and Raquel Bailón. Photoplethysmographic Waveform Analysis for Autonomic Reactivity Assessment in Depression. *IEEE Transactions on Biomedical Engineering*, 68(4):1273–1281, April 2021.

[67] John Morales, Pascal Borzée, Dries Testelmans, Bertien Buyse, Sabine Van Huffel, and Carolina Varon. Linear and Non-linear Quantification of the Respiratory Sinus Arrhythmia Using Support Vector Machines. *Frontiers in Physiology*, 12:623781, February 2021.

[68] Samuel J. Harrison, Samuel Bianchi, Jakob Heinzle, Klaas Enno Stephan, Sandra Iglesias, and Lars Kasper. A Hilbert-based method for processing respiratory timeseries. *NeuroImage*, 230:117787, April 2021.