Machine Learning
Data Analytics

Friedrich-Alexander-Universität
Technische Fakultät

Systematic Benchmarking of Pre-Ejection Period Extraction Algorithms

Bachelor's Thesis in Medical Engineering

submitted
by

Julia Jorkowitz

born 18.12.2001 in Neustadt a.d.Aisch

Written at

Machine Learning and Data Analytics Lab
Department Artificial Intelligence in Biomedical Engineering
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

in Cooperation with

Department of Palliative Medicine, University Hospital Erlangen

Advisors: Robert Richer M.Sc., Luca Abel M.Sc., Prof. Dr. Bjoern Eskofier,
Dipl.-Ing. Stefan Griesshammer, Dr. med. Tobias Steigleder,
Prof. Dr. med. Christoph Ostgathe

Started: 01.02.2024

Finished: 01.07.2024

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.
Die Richtlinien des Lehrstuhls für Bachelor- und Masterarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Erlangen, den 01.07.2024

Übersicht

Benchmarking spielt eine entscheidende Rolle bei der Weiterentwicklung wissenschaftlicher und technologischer Disziplinen, indem es standardisierte Methoden zur Bewertung und zum Vergleich der Leistung verschiedener Algorithmen und Systeme bereitstellt. Es gewährleistet die Zuverlässigkeit, Reproduzierbarkeit und Transparenz von Ergebnissen, die für den wissenschaftlichen Fortschritt und technologische Innovationen unerlässlich sind. Diese Bachelorarbeit konzentriert sich auf die Bedeutung des Benchmarkings im Kontext der Messung der Herzfunktion und zielt auf die Pre Ejection Period (PEP) ab.

Die PEP ist ein vielversprechender Marker für die Aktivität des sympathischen Nervensystems mit großem Potenzial für klinische und Forschungsanwendungen. Dennoch wird seine breite Nutzung durch den komplexen Messprozess und das Fehlen von Extraktionsalgorithmen, die auf gängigen Datensätzen validiert sind, eingeschränkt. Diese Arbeit adressiert diese Herausforderungen durch die Etablierung eines systematischen Benchmarks für PEP-Extraktionsalgorithmen unter Verwendung kombinierter Elektrokardiagramm - und Impedanzkardiogramm -Signale.

Für die Studie wurden zwei unterschiedliche Datensätze herangezogen: ein Datensatz von Teilnehmern während einer akuten Stressaufgabe (Trier Social Stress Test (TSST)) und deren stressfreier Kontrollbedingung (Friendly Trier Social Stress Test (f-TSST)), sowie ein weiterer Datensatz von Teilnehmern, die einer Tilt Table Test (TTT)-Untersuchung unterzogen wurden. Diese Datensätze boten verschiedene Bedingungen für eine robuste Evaluierung der Algorithmusleistung. Der Benchmark umfasste mehrere Methoden zur Detektion des Q-Wellen-Beginns in Elektrokardiogramm-Signalen, welche den Startpunkt des PEP markieren, und des B-Punkts in Impedanzkardiogramm-Signalen, der dem Endpunkt des PEP entspricht.

Insgesamt 11.768 manuell annotierte Markierungen dienten als Goldstandard, der einen umfassenden Vergleich der algorithmischen Leistung.

Für beide Datensätze lieferten unterschiedliche Algorithmuskombinationen die besten Ergebnisse. Für den durch den TSST gewonnenen Datensatz zeigte die Kombination die beste Leistung, bei der 40 ms vom R-Peak subtrahiert wurden, um den Q-Wellen-Beginn zu bestimmen, und lokale Minima der zweiten Ableitung des first derivative of the cardiac impedance (dZ/dt)-Signals des Impedanzkardiogramms zur Identifizierung des B-Punkts verwendet wurden. Hier betrug der mittlere absolute Fehler 14,55 ms ± 19,20 ms. Für den zweiten Datensatz beinhaltete der beste Ansatz, der einen mittleren absoluten Fehler von 18,41 ms $\pm$ 16.32 ms zeigte, die Subtraktion von 32 ms vom R-Peak und die Verwendung einer Methode, die den Punkt mit dem größten vertikalen Abstand vom dZ/dt-Signal zu einer geraden Linie, die zwischen dem C-Punkt und dem Punkt 150 ms vor dem C-Punkt gezogen wurde, zur Bestimmung des B-Punkts identifizierte.

Abstract

Benchmarking plays a crucial role in advancing scientific and technological fields by providing standardized methods for evaluating and comparing the performance of various algorithms and systems. It ensures the reliability, reproducibility, and transparency of results, which are essential for scientific progress and technological innovation. This bachelor's thesis focuses on the importance of benchmarking in the context of cardiac function measurement, targeting the PEP.

The PEP is a promising marker for Sympathetic Nervous System (SNS) activity, with high potential for clinical and research applications. However, its complex measurement process and the lack of extraction algorithms that are validated on common datasets hindered its widespread use. This thesis addresses these challenges by establishing a systematic benchmark for PEP extraction algorithms using combined Electrocardiogram (ECG) and Impedance cardiogram (ICG) signals.

The study utilized two distinct datasets: one from participants during an acute stress task (Trier Social Stress Test (TSST)) and its stress-free control condition (Friendly Trier Social Stress Test (f-TSST)), as well as one dataset from participants undergoing a TTT examination. These datasets provided varied conditions for a robust evaluation of algorithm performance. The benchmark involved multiple methods for detecting the Q-Wave onset in ECG signals, which marks the start point of the PEP, and the B-Point in ICG signals, corresponding to the PEP end point.

A total of $11{,}768$ manual labeled annotations served as the gold standard, enabling a comprehensive comparison of algorithmic performance.

For both datasets, different algorithm combinations emerged as the best-performing. For the dataset obtained through the TSST, the combination that showed the best performance, with a Mean Absolute Error (MAE) of 14.55 ms $\pm$ 19.20 ms, involved subtracting 40 ms from the R-Peak to determine the Q-Wave onset and using local minima of the second derivative of the dZ/dt signal of the ICG for identifying the B-Point. For the second dataset, the best approach, which showed a MAE of 18.41 ms $\pm$ 16.32 ms, involved subtracting 32 ms from the R-Peak and utilizing a method that identifies the point with the greatest vertical distance from the dZ/dt signal to a straight line drawn between the C-point and the point 150 ms before the C-Point for determining the B-Point.

Contents

1 Introduction **1**

2 Related Work **5**

2.1 Pre-Ejection Period as Marker for Sympathetic Activity 5

2.2 Benchmarking and Comparison of Algorithms for Biomedical Signal Processing 7

3 Methods **9**

3.1 Dataset Description 9

3.1.1 TSST Dataset 9

3.1.2 GUARDIAN Dataset 13

3.2 Measurements 15

3.2.1 Electrocardiogram (ECG) 15

3.2.2 Impedance cardiogram (ICG) 16

3.3 Fiducial Point Detection Algorithms 18

3.3.1 Preprocessing 18

3.3.2 Q-Wave onset 19

3.3.3 C-Point 21

3.3.4 B-Point 21

3.3.5 Outlier Detection 26

3.3.6 Outlier-Correction 26

3.4 Manual Event Labeling 27

3.4.1 Q-Wave onset 28

3.4.2 B-Point 29

3.5 Evaluation 29

4 Results & Discussion **31**
4.1 Reference PEP data 31
4.2 Fiducial Point Detection Algorithms 34
4.2.1 Q-Wave onset 40
4.2.2 B-Point 46
4.2.3 Best-performing Algorithms 52
4.3 General Discussions & Limitations 57

5 Conclusion & Outlook **59**

List of Figures **61**

List of Tables **63**

Bibliography **65**

A Additional Tables **75**

B Additional Figures **79**

C Acronyms **89**

Chapter 1

Introduction

As the world becomes increasingly interconnected, the prominence of open science continues to rise. By providing open access to research information and results, open science fosters reproducibility and facilitates the verification of results. Furthermore, this approach allows a broader and more diverse group of people to contribute to scientific discourse, enriching various fields with fresh perspectives and innovative ideas [Sto10; Bor12; Gen23]. In the field of data science, an application scenario of this can be found in the identification of the most efficient and effective algorithm for a specific problem, for which the objective comparison of algorithms is indispensable. This requires well-defined challenges, known as "benchmarks", which consist of standardized datasets and consistent evaluation strategies. These benchmarks allow for the assessment of algorithm performance against common standards [Vol23]. Benchmarking has already been successfully established in a variety of topics, such as sensor-based gait analysis [Küd24] or ECG analysis [Str20].

One application scenario in which systematic benchmarking is missing so far is for the extraction of the Pre Ejection Period (PEP), a time period associated with cardiac function. Defined as the time interval between ventricular depolarization and the onset of blood ejection, PEP has been acknowledged as a promising indicator of the influence of the Sympathetic Nervous System (SNS) on the heart rate [New79; Lar86]. This poses an advantage over conventional Heart Rate Variability (HRV) metrics, which are typically influenced by both sympathetic and parasympathetic activity. The PEP is usually measured using an ECG to extract the Q-Wave onset, which corresponds to the beginning of the PEP, and an ICG to extract the B-point, corresponding to the end of the PEP [For19].

Despite its potential as a promising and rather unbiased marker for sympathetic activity, the PEP is not widely used in research, which is due to various aspects. First, the measurement setup

is more complex compared to solely extracting HRV from ECG or Photoplethysmogram (PPG) recordings, a task which wearable devices such as smartwatches or fitness trackers are already capable of performing. Acquiring the ICG involves placing multiple electrodes on the body to accurately detect impedance changes, which is more invasive and time-consuming than the previously mentioned ECG or PPG recordings. Second, the extraction of both relevant points is prone to error. Especially the B-Point detection has been proven to be particularly difficult due to differences in the waveform between individuals, but also within the same person [Erm12]. This variation in waveform causes difficulties in accurately identifying the end point of the PEP in both manual and automatic detection, leading to errors in determining the duration of the PEP in both cases. This affects the ability to draw reliable conclusions and interpret the results, which therefore results in limited usage.

Lastly, there is a lack of publicly available benchmark datasets on which algorithms for extracting the fiducial points can be validated. *PhysioNet* is a research resource that provides free access to a large and growing collection of physiological signals and clinical data for research in biomedical engineering, computational biology, and medicine [Phy24d]. It also offers software tools and algorithms for analyzing and processing physiological data across a variety of topics. The platform includes a wide range of datasets, such as ECG recordings under different conditions (normal sinus rhythm, arrhythmia databases, and long-term ECG data) [Phy24a], Electroencephalography (EEG) data (including motor movements, imagery, and eye state recordings)[Phy24b], and respiratory signals (including sleep studies and data from patients with chronic obstructive pulmonary disease) [Phy24c]. However, as of now, there are no published results for the extraction of the PEP on *PhysioNet*, emphasizing the issue of missing datasets and research in this area.

The goal of this bachelor's thesis is to address the last two issues by presenting the first systematic benchmark for PEP extraction algorithms from combined ECG and ICG signals. This approach integrates various methods for Q-Wave detection in ECG and B-Point detection in ICG signals, alongside multiple Outlier Correction algorithms, using the `tpcp` Python library to create comprehensive data science pipelines [Küd23]. Manually labeled data served as the gold standard against which the results were evaluated. To achieve this, two distinct datasets were collected, each employing individual study protocols and different measurement systems. One dataset was gathered in a study under the *EmpkinS collaborative research center*, where participants were exposed to acute psychosocial stress through the TSST as well as to a stress-free control condition (f-TSST) since acute stress exposure activates the stress pathways in the body, including an activation of the SNS, leading to changes in the PEP [Emp23]. The data of the other dataset was obtained during a TTT examination performed by participants within the framework of the

GUarded by **A**dvanced **R**adar technology-based **DI**agnostics **A**pplied in palliative and intensive care **N**ursing (GUARDIAN) project at the *Department of palliative Medicine of the University Hospital Erlangen* [Uni23]. Finally, the performances of the pipelines are evaluated separately for each dataset by comparing them to the manually labeled events. Additionally, the differences in algorithm performances between the two datasets were identified and analyzed.

Chapter 2

Related Work

2.1 Pre-Ejection Period as Marker for Sympathetic Activity

The Autonomic Nervous System (ANS) is composed of two main components: the SNS and the Parasympathetic Nervous System (PNS), each having opposing effects on the body [McC07]. These systems operate continuously but with varying intensities depending on the situation, which enables precise regulation of diverse bodily functions [McC07]. The SNS is responsible for triggering "fight-or-flight" responses and is therefore prominent in stressful situations. To prepare the body for such a situation, the blood flow is reallocated to maximize supply to the muscles [McC07]. Conversely, parasympathetic activity predominates in calm moments, facilitating relaxation and energy conservation. As can be seen from the term "rest and digest", which describes the physiological state during parasympathetic activation, the PNS primarily supplies the digestive tract with blood [McC07; Gib19]. The ANS influences various bodily functions, including heart rate, blood pressure, metabolism, body temperature, and pupillary response [McC07; Gib19]. To maintain balance within the body the SNS and PNS work antagonistically. For instance, the SNS dilates the pupils to enhance vision and prepare for potential threats, whereas the PNS constricts the pupils to reduce the amount of light entering the eyes in relaxed situations [Gib19]. Another example is heart rate: the SNS increases heart rate, while the PNS decreases it [Gib19]. Due to the simultaneous influence of both systems, it is challenging to determine the impact of only one system on a specific bodily function [New79]. To measure and understand ANS activity, various physiological parameters are used. A typical parameter is the Heart Rate (HR), which is influenced by both the SNS and the PNS. Generally, the PNS has a stronger effect on the HR, complicating efforts to measure only SNS activity [New79; Cac94]. Similarly, blood pressure (BP) is regulated by both the SNS and the PNS, presenting the same challenge in isolating the effects of

a single system [Gib19]. In contrast to these parameters, studies show that the PEP is influenced only by the SNS [Cac94]. This makes PEP a reliable marker of sympathetic activity, superior to typical parameters like HR [New79]. The term PEP describes the cardiological time period between ventricular depolarization and the opening of the aortic valve, resulting in the onset of blood ejection [Lar86].

The PEP plays a crucial role in stress research by examining the response of the ANS to stress-inducing stimuli. Elevated stress levels are often associated with a shortened PEP, indicating increased sympathetic activation [New79]. Furthermore, PEP serves as a valuable tool in the study of anxiety disorders and other mental health conditions. Individuals with anxiety disorders often exhibit increased sympathetic activity, reflected in shortened PEP times [Fu18]. This measure can be used to quantify the physiological effects of anxiety and monitor its severity. In clinical research and practice, PEP can also be used to investigate cardiovascular diseases. Shortened PEP times have been linked to an increased likelihood of cardiovascular problems such as aortic valve dysfunctions [New79]. Therefore the PEP duration can provide information for a variety of applications.

To be able to determine the PEP, various algorithms have been proposed in related works. Since the Q-Wave is not visible in the ECG of every person, there are approaches to use other points that are easier to detect instead of the Q-Wave Onset as start point of the PEP. For example Bernston et al. propose using the onset of the R-Peak as replacement [Ber04]. However, an even simpler and more reliable detection is possible for the R-Peak itself. Therefore, Seery et al. recommend its use as start point of the PEP to eliminate detection errors [See16]. Additionally, there are algorithms that exploit the ease of detecting the R-Peak by subtracting a fixed time interval from it to locate the position of the Q-Wave Onset [Lie13]. Regarding the B-Point, there are some approaches that utilize the relationships between the degrees of derivatives to identify significant points that indicate a possible morphology of the B-Point. These include the inflection points of the second derivative and the maxima of the third derivative, which are each used in a method by Debski et al [Deb93]. Additionally, algorithms based on only the dZ/dt signal and its zero crossings or peaks have been implemented [For19]. All these methods are based on fixed rules. To adapt the detection to the different waveforms of the dZ/dt signal, Forouzanfar et al. developed an algorithm that consists of several steps to adapt the determination process to the respective morphology of the signal [For19].

2.2 Benchmarking and Comparison of Algorithms for Biomedical Signal Processing

Benchmarking serves as a critical tool in the field of algorithmic evaluation, facilitating the comparison and assessment of various algorithms designed to tackle a specific problem [Hot05]. This method allows to determine how effective various algorithms are by testing them against a set of performance challenges (”benchmarks”) [Hot05]. Ensuring the reproducibility of benchmarking results is crucial for their reliability and integrity [Pre95; Xia24]. Reproducibility embraces the ability for independent parties to replicate and verify experimental findings, confirming the reliability of the conclusions [Pre95]. To achieve this, unrestricted access in the sense of open science to both research methodologies and outcomes is considered essential in related studies [Xia24; Pre95]. This open access allows researchers to find and use the best approaches for specific problems [Hot05; Vol23; Wan23].

Since algorithms may perform differently on various datasets, having a diverse and representative data sample in benchmarking is very important [Hot05; Vol23]. Using benchmarking methods and promoting transparency in research helps the field of algorithm evaluation to grow, leading to new advancements in areas like computer vision, machine learning, and artificial intelligence [Vol23].
In the context of this study, benchmarking serves as the fundamental basis for evaluating the performance of various algorithms developed for automatic PEP detection.

Achievements in benchmarking have already been made in various fields of study. One notable example is the *Gaitmap* project, which focuses on sensor-based gait analysis [Küd24]. As part of the project, 20 algorithms were implemented tailored for the analysis of data obtained from Inertial Measurement Units (IMU) [Küd24]. These algorithms have been compiled from several sources in order to obtain the broadest possible portfolio of methods that can provide insights into the complex dynamics of human gait from IMU data [Küd24]. Moreover, the Gaitmap project goes beyond algorithm implementation by providing access to datasets containing gait-related information. These datasets serve as resources for researchers, offering a diverse array of real-world gait data that can be used for validation, testing and comparison purposes [Küd24]. Additionally, the project offers a platform with benchmark challenges, fostering a collaborative environment where researchers can showcase their algorithms and compete to achieve optimal performance. To enable this, the platform is designed to accommodate the integration of new data [Küd24].

Another topic is ECG analysis, for which a benchmarking approach is given by the PTB-XL dataset [Str20]. This dataset offers a vast library of ECG data, which includes a total of 21,837 ECG

records from 18,885 patients, each lasting ten seconds [Str20]. In addition to the raw ECG data, the PTB-XL dataset includes algorithms designed to identify the critical points and extract relevant information from the ECG signals. These algorithms facilitate the detection of specific features, such as the QRS complex [Str20]. The dataset's diagnostic capabilities are further enhanced by its classification system, which can categorize the ECG signals into one of 71 diagnostic categories. These categories are organized hierarchically, allowing for both coarse and fine classifications depending on the specific requirements of the analysis [Str20]. To enable the assessment of diagnostic accuracy, each diagnosis label within the PTB-XL dataset is accompanied by a likelihood score ranging from 15 to 100. These scores represent the probability of the diagnosis being correct, with higher values indicating greater confidence in the diagnosis [Str20]. By offering both the data and the accompanying diagnostic algorithms, it provides a robust foundation for developing and benchmarking new models for automated ECG analysis. Researchers can use this dataset to train machine learning algorithms, validate their performance, and compare their results against established benchmarks. Additionally, there is the rPPG (remote Photoplethysmography) toolbox, a comprehensive software package designed for the extraction and analysis of physiological signals, specifically aimed at measuring heart rate and related metrics from video recordings [Lab24]. Utilizing advanced computer vision and signal processing techniques, the rPPG toolbox non-invasively detects subtle color changes in human skin caused by blood flow, allowing for the estimation of pulse waveforms without the need for physical contact with the subject.

Achievements regarding benchmarking have also been made in emotional analysis. The Dataset for Emotion Analysis using Physiological Signals (DEAP) is a resource in the field of affective computing and human emotion analysis [Koe12]. The dataset was created to study human emotional responses to multimedia content, particularly music videos. DEAP is designed to facilitate research in emotion recognition and analysis, providing a comprehensive dataset for developing and testing algorithms that can interpret human emotions based on physiological responses. Drawing inspiration from these projects, efforts are now done in this work to extend benchmarking to the realm of PEP. To implement this, this work builds on a previous Bachelor's thesis, which focused on investigating the PEP as a potential stress marker [Stü23]. For this purpose, algorithms for automatic event detection of both PEP start and end points were implemented and their performance was tested and evaluated on one dataset. This approach was adopted and expanded. The data was supplemented by a further dataset on which the algorithms were also tested. Thus, the algorithms have been modified, making them applicable to multiple datasets rather than being tailored to just one. For benchmarking purposes, everything was designed to allow the easy addition of new datasets in the future.

Chapter 3

Methods

To compare the performances of the algorithms on different data, two datasets were used within this work. One of these was recorded in the context of a study investigating the contactless measurement of stress, its determinants, and consequences under the *EmpkinS collaborative research center* at the *EmpkinS Lab*, where the participants performed the TSST and the f-TSST [Kur24]. For the sake of simplicity this dataset will be referred as TSST Dataset in the following. The other dataset was derived from a study, which was conducted in relation to the project GUARDIAN at the *Department of Palliative Medicine of the University Hospital Erlangen*. From now on, this dataset is called GUARDIAN Dataset. In addition to variations in study protocols, the two datasets also employed different measurement systems.

3.1 Dataset Description

In the following section, the two datasets are described in more detail with respect to their study populations and data acquisition.

3.1.1 TSST Dataset

The TSST Dataset was derived from 15 participants (9 female and 6 male), who were divided into two groups: 7 in the sitting group and 8 in the standing group. Depending on the group assignment, the TSST and f-TSST were performed either sitting or standing. The recruitment process took place via mail, social media, flyer, and in person. The exclusion criteria included age below 18 or above 50 years, non-German native language, BMI lower than 18 or higher than 30, drug use, as well as experience with a similar stress test. For participation, either 50 Euros or 5 subject

hours, in the case of psychology students, were received. Table 3.1 shows the demographic and anthropometric data of the participants divided by gender.

Table 3.1: Demographic and anthropometric data of the study participants for the TSST Dataset

	Age [years]	**Height** [cm]	**Weight** [kg]	**BMI** [kg m^{-2}]
Female	24.41$\pm$ 2.51	168.33 $\pm$ 5.50	59.44 $\pm$ 6.06	20.92 $\pm$ 1.14
Male	21.50 $\pm$ 1.61	180.83 $\pm$ 5.43	76.50 $\pm$ 10.69	23.35 $\pm$ 2.79
Total	23.07 $\pm$ 2.54	173.33 $\pm$ 8.21	66.27 $\pm$ 11.73	21.89$\pm$ 2.30

Trier Social Stress Test (TSST)

The TSST is experimental gold standard to induce acute stress, by putting the participants in an interview situation [All17]. In the study at the *EmpkinS lab*, the following protocol was followed to record the data. The test consists of three main parts, each lasting five minutes. The first phase is the preparation phase, during which the participants were instructed to take notes on their personality and complete questionnaires. In the subsequent phase, known as the *Talk* phase, participants were required to deliver a speech about their personality in front of a panel of two individuals dressed in lab coats. This panel was of mixed gender, and the participants were instructed to address only the person of the opposite gender, maintaining eye contact throughout. The panel members were trained in advance to exhibit minimal emotional response and to interact with participants solely using predetermined sentences. The final phase involved a mental arithmetic task where the participants were asked to repeatedly subtract 17 from $2{,}043$. If a mistake was made, the participants had to start over from $2{,}043$. A visualisation of the setup for the standing and sitting group can be found in Figure 3.1. Breaks were implemented between the different phases to provide calmer periods for radar measurements, as it was anticipated that participants might move too much during the TSST for the radar to be effective. Additionally, there were breaks at the halfway points of both the *Talk* and *Math* phases. During these breaks, the participants were instructed to remain as still as possible. The temporal sequence of the TSST is graphically represented in Figure 3.2.

Figure 3.1: Setup of the TSST for standing (a) and sitting (b) group

Friendly Trier Social Stress Test (f-TSST)

For a control condition the participants of TSST Dataset performed the f-TSST, a friendlier version of the TSST, either a day before or after the TSST. The overall time schedule remained unchanged, but the environment was made more pleasant. In this version, the panel members wore casual clothes instead of lab coats, and both panel members were allowed to engage in conversation with the participant. Moreover they were instructed to respond in a friendly and affirmative manner. In addition, the math problem was simplified in the form that 10 and 20 had to be added alternately. Within this dataset, all participants underwent every phase of both the TSST and the f-TSST. Table 3.2 provides the mean duration of the recording length for each phase, categorized by condition.

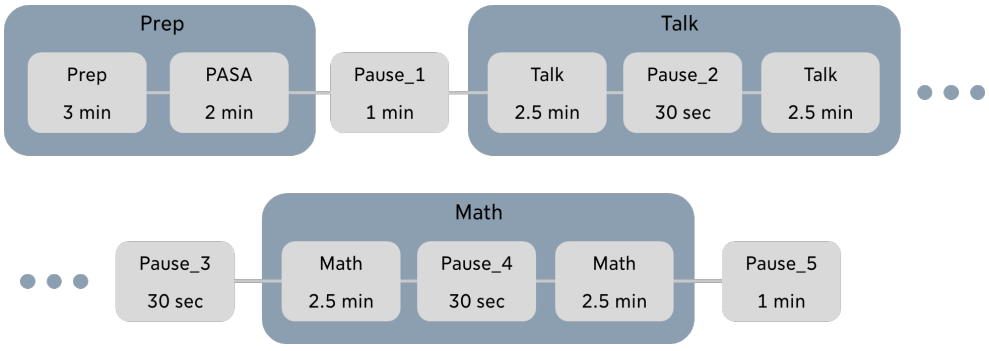

Figure 3.2: Temporal sequence of the (f-) TSST

Table 3.2: Mean duration recordings per phase divided into conditions for the TSST Dataset

		Mean duration [min:sec]
TSST	**Prep**	09:21 ± 04:48
	Pause_1	01:51 ± 00:49
	Talk	09:25 ± 04:56
	Math	09:34 ± 04:54
	Pause_5	01:51 ± 00:49
f-TSST	**Prep**	08:09 ± 04:31
	Pause_1	01:41 ± 00:50
	Talk	08:06 ± 04:49
	Math	08:07 ± 04:38
	Pause_5	01:42 ± 00:50

3.1.2 GUARDIAN Dataset

The GUARDIAN Dataset includes data from 24 participants (12 female and 12 male). An overview of the composition regarding age, height, weight, and BMI can be found in Table 3.3.

The age range of the participants of GUARDIAN Dataset is larger than that of TSST Dataset, especially regarding females, which is also reflected in larger standard deviations of all other values of this subset.

Table 3.3: Demographic and anthropometric data of the study participants for the GUARDIAN Dataset

	Age [years]	**Height** [cm]	**Weight** [kg]	**BMI** [$\mathrm{kg\,m^{-2}}$]
Female	35.67± 13.21	169.67 ± 8.00	68.75 ± 15.16	23.72± 4.04
Male	26.75 ± 4.23	183.42± 7.30	80.00± 10.23	23.76± 2.82
Total	31.21± 10.91	172.79± 10.65	73.96± 13.41	22.83± 4.03

Tilt Table Test (TTT)

To obtain the GUARDIAN Dataset the participants performed the TTT within the framework of the GUARDIAN-Project. The TTT is a non-invasive clinical examination to induce orthostatic stress and is commonly used to examine the causes for syncope [Teo16]. This test involves securing the patient on a tilt couch and moving them to an upright position, which causes blood to flow into the lower extremities and causes a decreased blood volume in the upper body, leading to a drop in blood pressure [Zys24; San13]. While this blood pressure drop is quickly compensated in healthy individuals by vasoconstriction of lower body vessels, this condition can lead to orthostatic hypotension and even to syncope if the blood pressure regulation fails [Zys24]. Figure 3.3 shows a schematic representation of the setup of a TTT.

In the course of collecting the data used in this Bachelor's thesis, the test persons first went through a resting phase, followed by a Valsalva maneuver, and Apnea phase, and the actual Tilt-Up/Tilt-Down maneuvers of the TTT. During the resting phase, the participants were instructed to lay relaxed, allowing for the measurement of baseline vital parameters. In the subsequent phase, the Valsalva maneuver was performed several times in succession, during which an exhalation is performed against closed airways [Pst16]. The back pressure of the inhaled air increases the intrathoracic pressure inside the body, which leads to a drop in blood pressure. The Valsalva

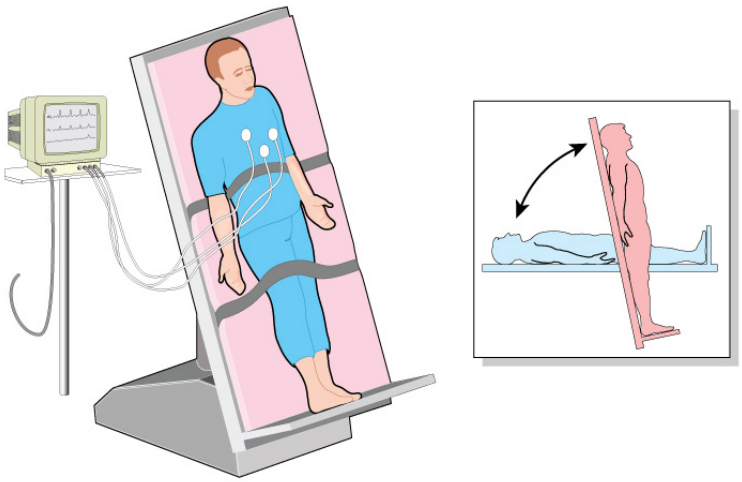

Figure 3.3: Schematic representation of the setup of a TTT [Ban24]

maneuver is therefore a valuable tool for assessing cardiovascular function and diagnosing heart problems, as it tests the body's ability to respond to sudden changes in intrathoracic pressure [Pst16; Sri24]. During the Apnea phase, the participants repeatedly held their breath. In contrast to the Valsalva maneuver, no pressure is generated by forced exhalation. Not every phase was performed by all participants. The number of people that performed each phase, as well as the mean duration with associated standard deviation, is described in Table 3.4.

Table 3.4: Mean duration of recordings and quantity of participants per phase for the GUARDIAN Dataset

	Number participants	**Mean duration** [min:sec]
Resting	22	12:59 $\pm$ 03:28
Valsalva	22	19:03 $\pm$ 01:11
Apnea	22	06:54 $\pm$ 01:55
TiltUp	21	15:37 $\pm$ 05:27
TiltDown	22	11:48 $\pm$ 00:39

3.2 Measurements

Both ECG and ICG signals were recorded in order to obtain the data for both datasets. These two measurement modalities are described below. For data acquisition of the TSST Dataset the Biopac MP160 system along with the AcqKnowledge software package was used for both ECG and ICG measurement [BIO24]. A sampling rate of $1{,}000$ Hz was chosen. For the GUARDIAN Dataset the Task Force Monitor was used for the data monitoring [Wil18; CNS24].

3.2.1 Electrocardiogram (ECG)

The ECG is a measure of the electrical activity of the heart [Sat24; AL-15]. This non-invasive technique uses electrodes on the body surface to measure electrical impulses that precede and trigger the contraction of the heart [AL-15]. The characteristic waveform of an ECG consists of six main components: P, Q, R, S, T and U-wave [Dou24]. The P wave is triggered by depolarization of the muscle cells in the atrium [Bec06; Sat24]. The subsequent depolarization of the cardiac interventricular septum is reflected in the ECG in the form of the onset of the Q wave [Win16]. The depolarization spreads to the ventricles and Purkinje fibers during the R and S waves [Win16]. The cause of the U-wave, which is not always visible, is not yet fully understood [Gir05]. A visualization of a typical ECG-waveform with all relevant points and intervals marked is displayed in Figure 3.4.

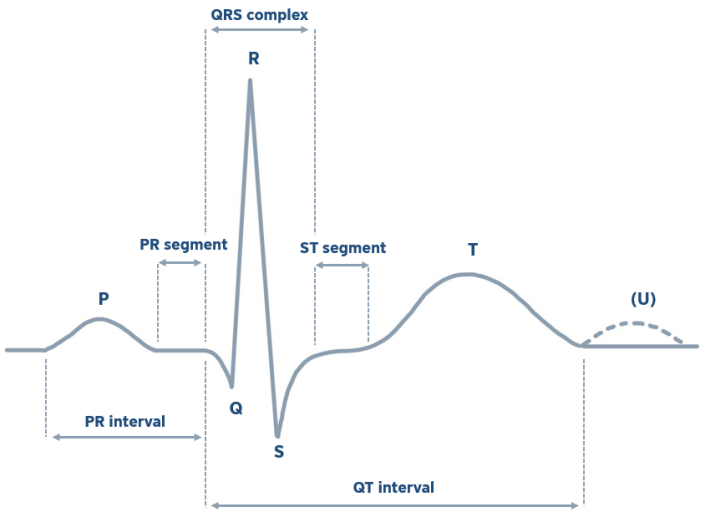

Figure 3.4: Schematic illustration of a ECG waveform with relevant points and intervals (modified from [Dou24])

In both datasets, the ECG was acquired according to Lead II of Einthovens's triangle [Ein12]. In terms of PEP calculation, the Q-Wave onset plays a crucial role as the starting point of this cardiological time period. However, since the definition of this point is vague and the Q-Wave Onset is not visible for every patient, its use is questionable [See16].

3.2.2 Impedance cardiogram (ICG)

While an ECG provides information about the start of the PEP, it can not capture the mechanical beginning of the blood ejection from the ventricles, which is necessary as end point for PEP computation [She90; Pil23]. To detect this point an ICG can be used [She90]. This non-invasive method is based on the relation between blood flow in the aorta and the resistance to electrical current in the thorax [Wan06]. By measuring this resistance, or impedance, and applying Ohm's law, which relates impedance (Z), voltage (U), and current (I), blood flow can be inferred [Wan06]. To measure the impedance, electrodes are placed on the body surface. These electrodes are divided into two groups: current electrodes, which apply a high-frequency, low-amplitude alternating current, and voltage electrodes, which measure the resulting voltage change [Wan06; Man18]. For data recording in this study, four electrodes were placed on the neck and the lower part of the chest [She90]. On both the neck and the chest, one voltage and one current electrode are placed on top of each other per side of the body [She90]. The voltage electrodes are positioned medial at a distance of 5 cm from the current electrodes [She90]. Figure 3.5 shows a visual representation of the electrode placement.

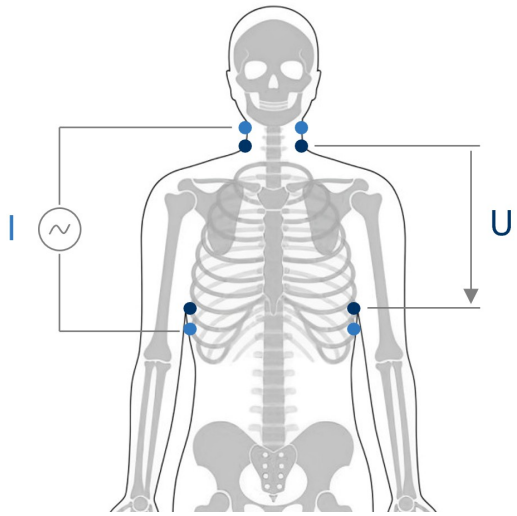

Figure 3.5: Overview of electrode placement for ICG measurement (modified from [Ste23])

Before processing the ICG signal, the first step is typically to compute the first derivative by time dZ/dt, which reflects the change of the thoracic impedance. Similar to the ECG signal, the dZ/dt waveform has characteristic points, which can be automatically detected (Figure 3.6) [Lab70]. Prominent points are A, B, C, X and O point. A marks the start of electromechanical systole caused by contraction of the atrium [Man18]. The time point of the aortic valve opening resulting in start of blood ejection can be observes at the point marked with B. The peak of the dZ/dt signal is represented by the C point and is created by ventricular contraction [Man18]. At the closure of the aortic valve the X-Point can be seen in the signal[Man18]. Lastly, the point O is associated with opening of the mitral valve and the resulting change of volume [Man18].

The most relevant point in an ICG regarding PEP calculating is the B-Point, as it reflects the aortic valve opening, which corresponds to the end of the PEP [She90; Man18]. Unfortunately, the B-Point is challenging to detect as it can be located at any point during the ascending slope between the A-Point and the C-Point [Man18; She90]. Figure 3.6 shows the waveform of dZ/dt signal of the ICG with the described points marked.

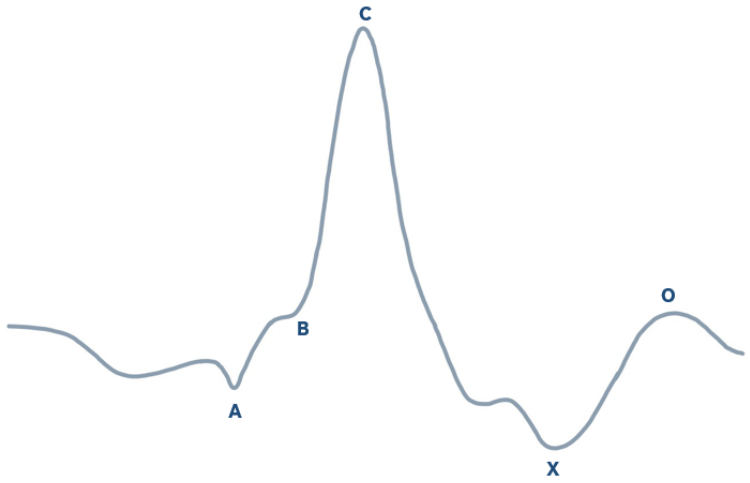

Figure 3.6: Schematic illustration of a dZ/dt waveform with relevant points (modified from [Ulb14])

3.3 Fiducial Point Detection Algorithms

An overview of all the algorithms concerning Q-Wave onset and B-Point detection as well as Outlier Correction mentioned in this section can be found in Figure 3.7.

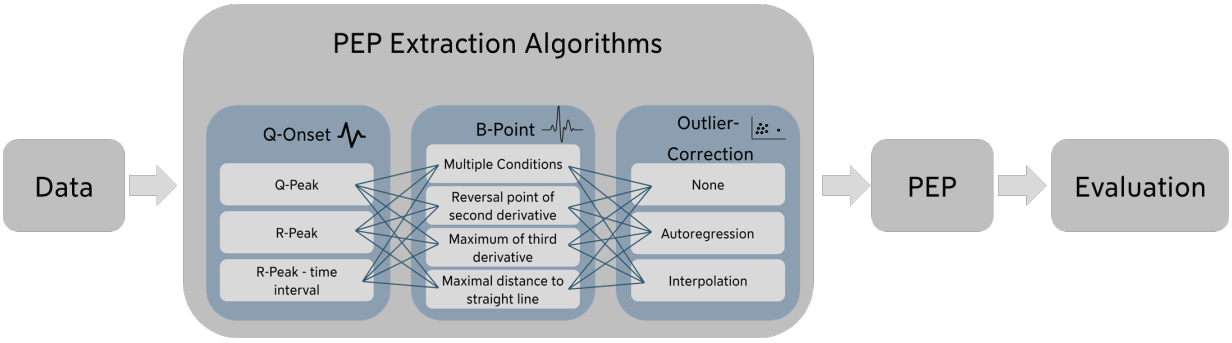

Figure 3.7: Overview of the systematic PEP benchmarking approach in this thesis including all employed algorithms. For comparison, pipelines for each combination of Q-Wave onset, B-Point and Outlier Correction algorithms were build and evaluated against each other.

3.3.1 Preprocessing

To enable the detection of the relevant points for PEP computation the ECG and ICG signals were cleaned in advance, since both signals are prone to baseline drifts, noise and artefacts [For19; Chr04]. For this purpose the ecg_clean method with the *biosppy* method from the *neurokit2* python package was applied to the ECG and a 4th order Butterworth bandpass filter with cutoff frequencies of 0.5 Hz and 25 Hz was used for the ICG signal [For19; Mak21]. To ensure a beat-to-beat PEP computation, the different heartbeats were segmented. For this purpose a method developed by Sternemann et al., which is based on the heartbeat segmentation provided by the neurokit2 library, was used [Mak21; Ste23]. This method uses the time interval between two adjacent R-Peaks to determine the start and end points of a heartbeat [Ste23]. In contrast to the method of the *neurokit2* package, the heartbeat borders are set adaptively rather than by a fixed heartbeat duration [Ste23; Mak21]. Figure 3.8 shows segments of raw ECG and ICG signals from the GUARDIAN Dataset alongside the corresponding cleaned signals for comparison.

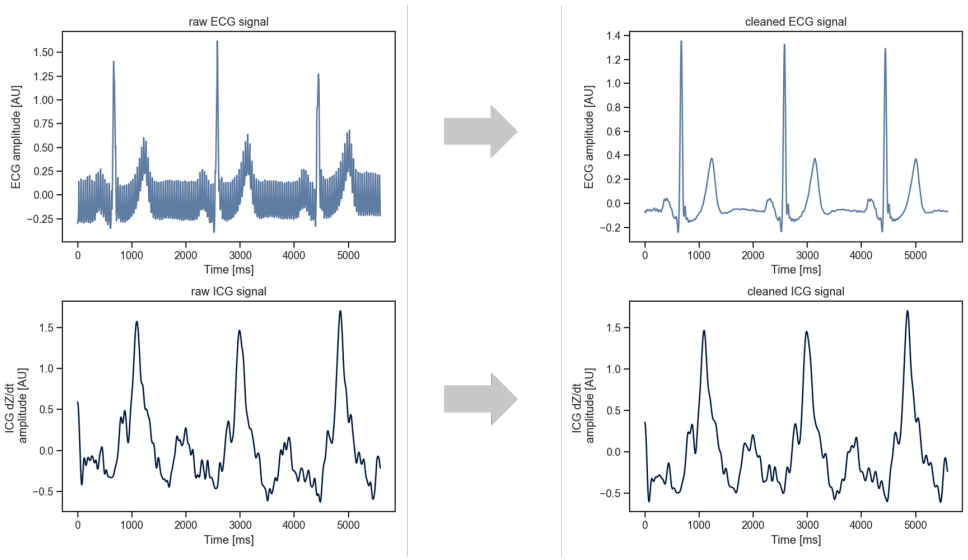

Figure 3.8: ECG and ICG signal from the GUARDIAN Dataset before and after preprocessing

3.3.2 Q-Wave onset

In this work three different methods for determining the Q-Wave onset were compared regarding their accuracy to detect the starting point of the PEP.

Algorithm that uses the Q-Peak as start point (QP)

Within this method, the Q-Peak is used to estimate the Q-Wave onset. Since the Q-Wave is not visible in the ECG of every person, the onset of the R-Peak was used instead, as proposed in related work [Ber04]. The onset of the R-Peak is always present, and it matches the Q-Peak when the Q-Wave is visible. This makes the onset of the R-Peak a good choice for ensuring comparability between all participants [Ber04]. For implementation, a method developed by Sternemann et al. was used [Ste23]. This method utilizes the *ecg_delineate* function from the *neurokit* python package to extract the Q-Peak. Within the *ecg_delineate* function, a discrete wavelet transform is performed to identify the significant points of an ECG [Mak21]. As the R-Peaks are required for this process and have already been determined during preprocessing, this step did not need to be implemented separately. Figure 3.9 displays exemplary heartbeats with Q-Wave onsets labeled by this algorithm. This algorithm will hereafter be represented by the abbreviation QP.

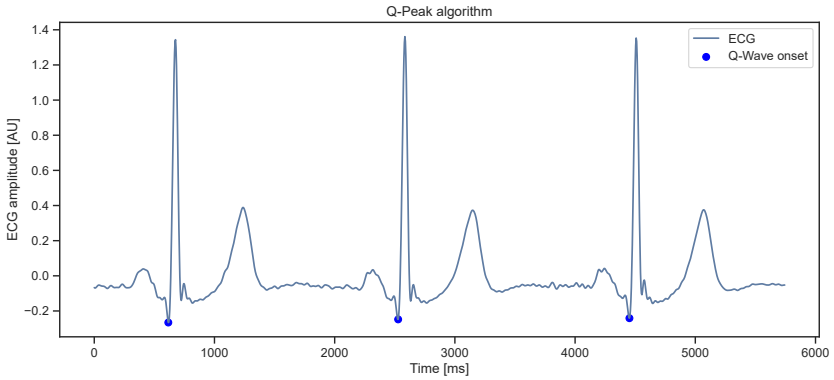

Figure 3.9: Labeled Q-Wave onsets using the QP algorithm

Algorithm that uses the R-Peak as start point (RP)

This algorithm follows the suggestion of previous work by Seery et al. to use the R-peak instead of the Q-Wave onset as starting point of the PEP [See16]. This approach has the advantage that the R-peak is the most significant point in the ECG and can therefore be determined with great reliability [See16]. Some labeled points of this algorithm are provided in Figure 3.10. This method will be referred as RP in the following.

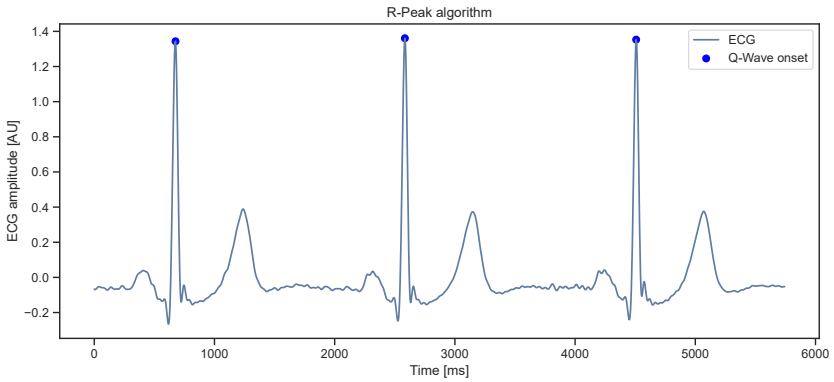

Figure 3.10: [Labeled Q-Wave onsets using the RP algorithm

Algorithm that subtracts 40 ms from the R-Peak to identify the start point (RP-40)

Based on the algorithm above, this approach uses the R-peak and subtracts a fixed time interval to estimate the onset of the Q-Wave. Van Lien et al. determined 40 ms to be the most promising time interval [Lie13]. However, in this thesis different time intervals $\Delta t, t \in [30, 32, 34, 36, 40]ms$

were used and compared. This method is based on the fact that the QR-interval of the ECG is highly stable under all circumstances [Pil23]. The functionality of this method utilizing a time interval of 40 ms is depicted graphically in Figure 3.11. For the sake of simplicity, these algorithms are represented in the rest of this thesis by the abbreviations (RP-40,RP-32,RP-34,RP-36).

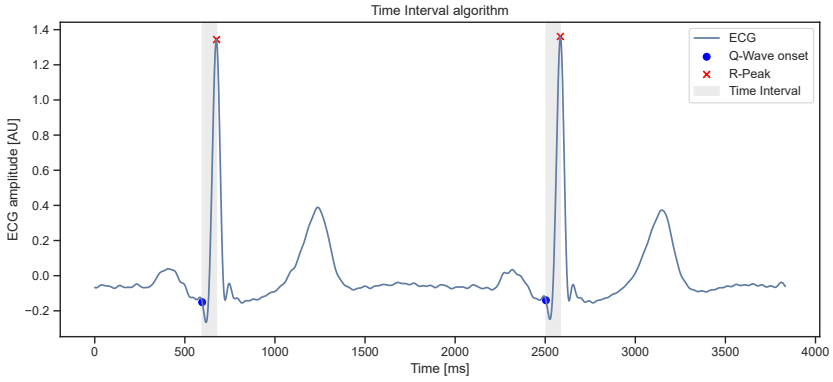

Figure 3.11: Labeled Q-Wave onsets using the RP-40 algorithm

3.3.3 C-Point

In order to limit the search radius of the B-Point algorithms, the C-Point was used as the upper time limit and was provided to the algorithms as a parameter. The algorithm used to determine the C-Points is therefore explained beforehand. Like the R-Peak in the ECG, the C-Peak is the most prominent point in the ICG and is mapped as global maximum within one heartbeat in the dZ/dt signal. For C-Point detection, an algorithm implemented in a previous master thesis [Ste23] was used. This algorithm is an adapted version of a peak-detection method of the *scipy* library [Vir20]. If multiple C-Points were detected, the algorithm selected the one closest to the average distance between the previous three C-Points and their respective R-Peaks [Ste23].

3.3.4 B-Point

In the case of B-Point extraction, four different algorithms were implemented. In contrast to the Q-Wave onset, the B-Point algorithms were given a parameter to determine whether Outlier Correction should be performed afterwards.

Algorithm that uses reversal points of the second derivative to identify the end point (SecDer)

This method is based on the fact that a reversal point of the second derivative corresponds to an inflection point in the dZ/dt signal, which is a possible B-Point location [Árb17; Deb93]. Therefore, all local minima in the second derivative of the cardiac impedance (dZ^2/dt^2) signal between the R-Peak and the C-Point were identified. If several points were found, the point closest to the C-Point was selected. In the case of Outlier Correction, for all heartbeats where a local minimum could not be found, the R-Peak was chosen as the B-Point. The connection between the local minima and the B-Points is illustrated in Figure 3.12. This approach will be referred to as SecDer in the following.

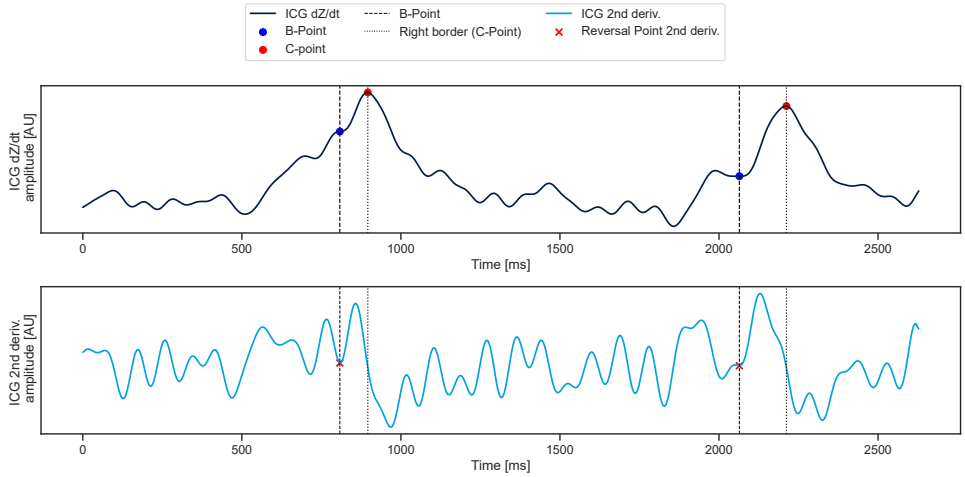

Figure 3.12: Labeled B-Points using the SecDer algorithm

Algorithm that uses local maxima of the third derivative to identify the end point (ThirDer)

This algorithm uses the third derivative of the ICG signal for B-Point extraction. From the third derivative of the cardiac impedance (dZ^3/dt^3) signal, the local maximum was determined as this point is equivalent to the point with the greatest change in slope in the dZ/dt signal, which is another possible morphology in which the B-Point can occur [Árb17]. Following the instructions by Arbol et al. [Árb17], the search interval was limited to 150 ms before the C-Point in order to achieve the highest possible accuracy. Visually, this algorithm is demonstrated in Figure 3.13. For this method, the abbreviation ThirDer will be used.

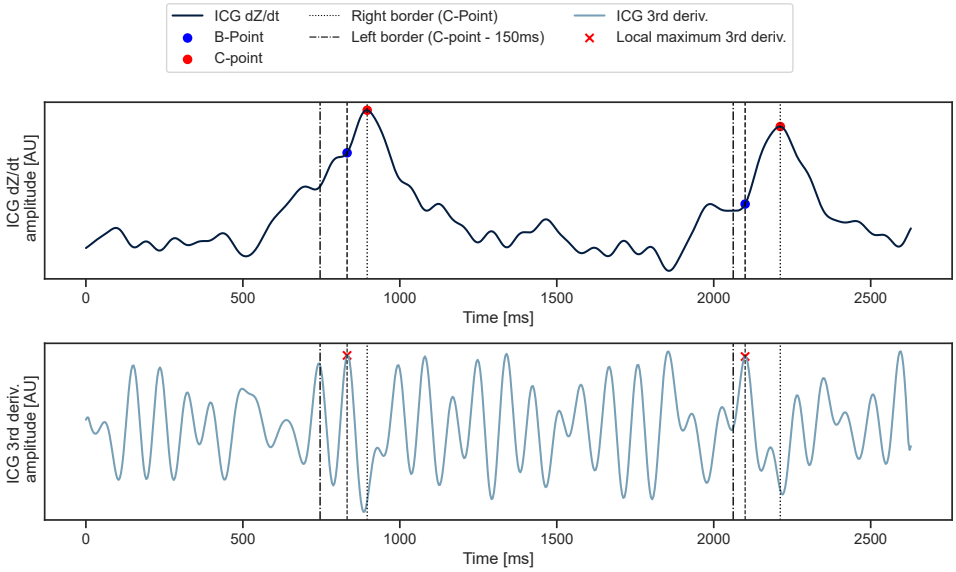

Figure 3.13: Labeled B-Points using the ThirDer algorithm

Algorithm that uses the maximal distance to a straight line between the C-Point and 150 ms before the C-Point to identify the end point (StrLin)

This approach uses the dZ/dt signal, in which a connecting line is drawn between the C-Point and the point of the dZ/dt signal 150ms before [Dro22]. The point with the greatest vertical distance to the line is then selected as the B-Point [Dro22]. The functionality of the algorithm is depicted in Figure 3.14. This algorithm will be abbreviated as StrLin in the subsequent sections.

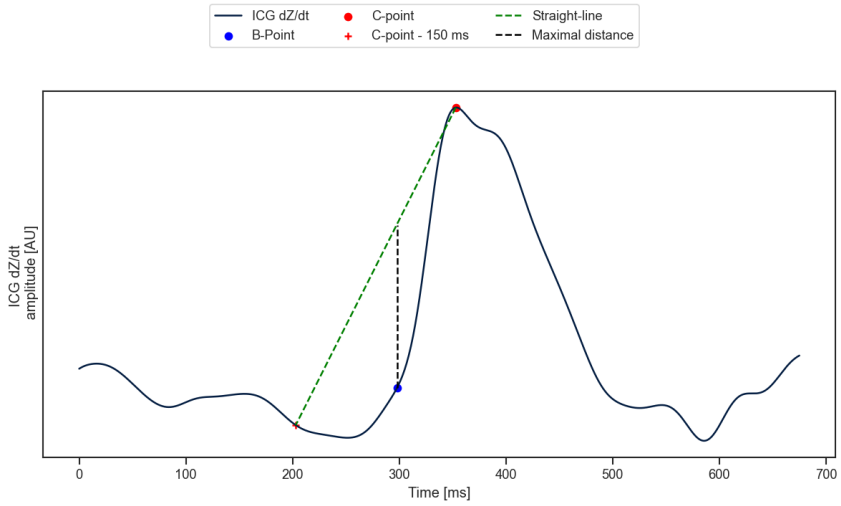

Figure 3.14: Labeled B-Points using the StrLin algorithm

Algorithm that uses multiple conditions to identify the end point (MultCon)

For the last approach, an algorithm developed by Forouzanfar et al. has been implemented [For19]. This method employs a series of steps to refine the search interval, followed by specific decision criteria, to identify the B-Point [For19]. The narrowing down of the search interval occurs in three sequential steps. Firstly, the A-Point associated with the heartbeat is detected [For19]. Subsequently, all monotonically increasing segments between the A- and C-Point are selected for further analysis [For19]. In contrast to the original implementation by Forouzanfar et al., the segment of all the previously detected segments, that extended from at least half of the amplitude of the C-Point to two-thirds of the amplitude was taken for further analysis [For19]. Therefore, the definition for the C-Point amplitude as distance between the zero line and the C-Point was chosen according to Sherwood et al. [She90].

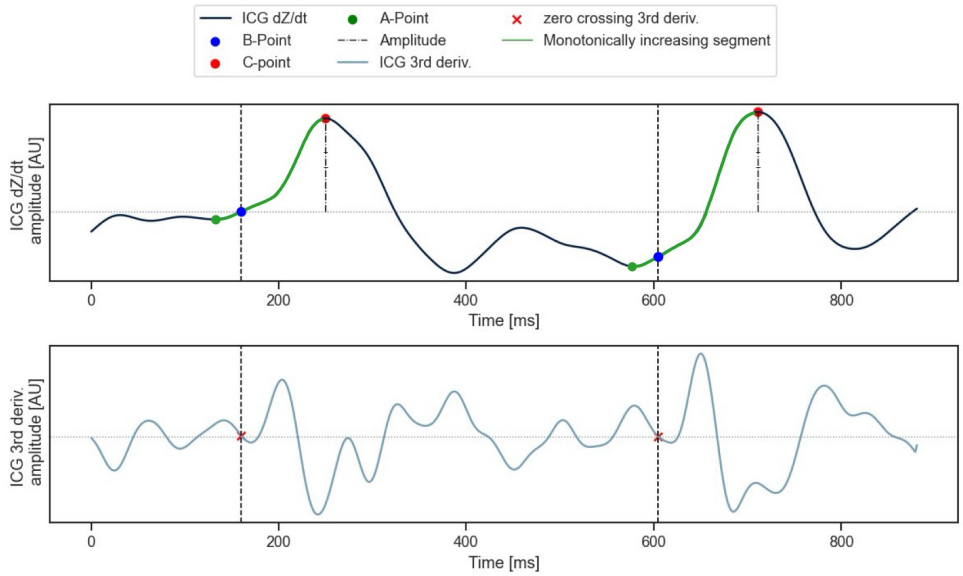

Figure 3.15: Labeled B-Points using the MultCon algorithm

To identify the B-Point within the specified interval, all zero crossings and local maxima of the dZ^3/dt^3 were determined, excluding those that didn't meet specific limits [For19]. These thresholds were set to $10 \cdot H/f_s$ for the zero crossings and $4 \cdot H/f_s$ for the local maxima, with H being the difference in amplitude between the A- and C-point and f_s representing the sampling rate [For19]. All zero crossings that corresponded to a greater value in the dZ^2/dt^2 signal than the defined threshold were excluded. Furthermore, all maxima with values below the threshold of $4 \cdot H/f_s$ were also discarded. In the end, the B-Point was determined as either the zero point or the local maximum, whichever is closest to the C-Point [For19]. If no point was detected with

this method, the onset of the monotonically increasing interval was chosen [For19]. Additionally, in cases requiring Outlier Correction, the R-Peak is designated as the B-Point. Using the R-Peak ensures to achieve the highest possible number of B-Points, since its detection is very reliable. An illustration of the described algorithm with the relevant features can be found in Figure 3.15. For ease of reference, this approach will be called MultCon in the following.

Table 3.5 provides an overview of all the introduced abbreviations of the algorithms that will be used in the following sections.

Table 3.5: Overview of abbreviations of the different Q-Wave onset algorithms, B-Point algorithms, and Outlier Correction methods

Algorithm Type	Description	Abbreviation
Q-Wave Onset	Algorithm that uses the Q-Peak as start point	QP
	Algorithm that uses the R-Peak as start point	RP
	Algorithm that subtracts 32 ms from the R-Peak to identify the start point	RP-32
	Algorithm that subtracts 34 ms from the R-Peak to identify the start point	RP-34
	Algorithm that subtracts 36 ms from the R-Peak to identify the start point	RP-36
	Algorithm that subtracts 40 ms from the R-Peak to identify the start point	RP-40
B-Point	Algorithm that uses reversal points of the second derivative to identify the end point	SecDer
	Algorithm that uses local maxima of the third derivative to identify the end point	ThirDer
	Algorithm that uses the maximal distance to a straight line between the C-Point and 150 ms before the C-Point to identify the end point	StrLin
	Algorithm that uses multiple conditions to identify the end point	MultCon
Outlier Correction	Interpolation	Intpol
	Autoregression	AutReg

3.3.5 Outlier Detection

Before applying Outlier Correction, the outliers in the detected B-Points were detected. Following the approach by Forouzanfar et al., the B-Point time data baseline, which was calculated by applying a Butterworth low-pass filter with a cut-off frequency of 0.1 Hz in forward and backward direction to the B-Point time values, were subtracted from the original B-Points first, which resulted in stationarized B-Points [For19]. Afterwards, the distance between the C- and B-Points was computed, as well as the median and median absolute deviation from the previously stationarized points [For19]. Outliers were ultimately considered to be points that deviated from the median by at least three median absolute deviations [For19].

3.3.6 Outlier-Correction

In this thesis, two different Outlier Correction algorithms were used, which will be described in the following. For both approaches, the process of outlier detection followed by Outlier Correction was repeated until no more outliers were found.

Interpolation (Intpol)

Within this work linear interpolation was used for correcting the B-Points marked as outliers. The *interpolate* method from the *pandas* library was chosen for implementation [McK10]. This method will be abbreviated with Intpol from now on.

Autoregression (AutReg)

As an alternative approach to Outlier Correction, an autoregressive model was used, as this method has previously been shown to successfully reduce the number of B-Point outliers [For19]. Autoregressive models are usually used to estimate the future trend of data [Lüt05]. For the B-Point Outlier Correction in this work, the principle of autoregression was applied in forward and backward direction. Following Forouzanfar et al., the average of both calculations was taken as the corrected point [For19]. To implement the autoregressive model the Autoregressive Integrated Moving Average (ARIMA) model provided by the *statsmodel* library was used, with the parameters *integration* and *moving average* set to zero, in order to solely obtain the autoregressive model [Sea10; For19]. Burg's method and the minimization of the Akaike information criterion (AIC) were employed to determine the parameters and order of the forward and backward models. [Aka69; For19]. Afterwards the baseline previously subtracted during the Outlier Detection (Section 3.3.5) was re-added to the data [For19]. The abbreviation used for the autoregression method is AutReg.

3.4 Manual Event Labeling

In order to obtain a gold standard against which the points found by the algorithms can be compared, the relevant points for PEP calculation were manually labeled. Due to the large amount of data available in both datasets, only parts of the data were labeled. Regarding TSST Dataset the ICG and ECG data of the 15 participants were labeled within the phases *Pause_1*, *Prep*, *Talk*, *Math* and *Pause_ 5* for each of the two conditions TSST and f-TSST. The sections for labeling were selected at random within the respective phases. For the phases *Pause_1* and *Pause_5*, a 10-second section was utilized, while for the remaining phases, the duration was 30 seconds. This results in a total of $5{,}094$ cardiac cycles that were manually labeled in this dataset. As for GUARDIAN Dataset the data of 24 participants were labeled. For this purpose, a time interval of 60 seconds was also randomly selected from each of the phases Resting, Valsalva, Apnea, TiltUp and TiltDown. The dataset includes $6{,}927$ cardiac cycles that were labeled by hand. The Python package *MaD GUI* allows data to be visualized in a user interface, where pre-annotated labels can be displayed and new labels can be set and saved [Oll22]. This package was used to produce the manually labeled data.

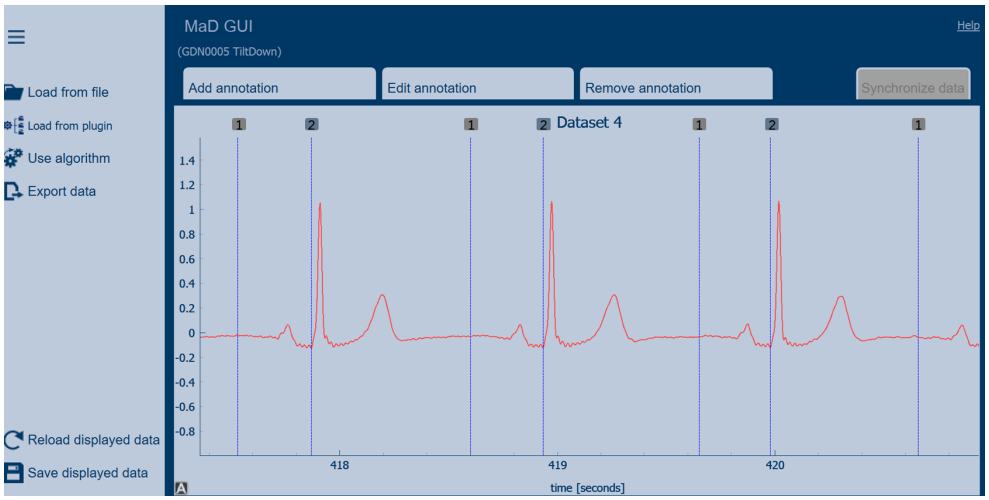

Figure 3.16: Example of labeled ECG signal in the *MaD GUI* package

To faciliate the labeling process, the pre-processed signal (as described in Section 3.3.1) was plotted together with the heartbeat borders which were also obtained from the preprocessing. Afterwards, the Q-Wave onset was labeled in the ECG signal and the B-Point in the ICG signal, respectively, according to the procedure explained below. Figure 3.16 exemplarily shows the described labeling using the *MaD GUI* with an ECG signal. The loaded signal is displayed in the *MaD GUI*, with the boundaries of heartbeats indicated by marker 1 and the onsets of Q-waves

identified by marker 2. If no reliable point for the Q-Wave Onset or the B-Point could be determined, this heartbeat was labeled as artefact. When excluding those heartbeats a total of $5{,}004$ cardiac cycles for the TSST Dataset and $6{,}764$ for the GUARDIAN Dataset remained. This results in a total of $11{,}768$ cardiac cycles that can be used as a reference and on which algorithms can be tested. A detailed breakdown of the numbers of labeled heartbeats excluding the artefacts is provided in Table 3.6.

Table 3.6: Amount of manually labeled cardiac cycles per phase

Dataset	Number Participants	Condition	Phase	Cardiac Cycles
TSST Dataset	**15**	TSST	Prep	683
			Pause_1	221
			Talk	776
			Math	764
			Pause_5	207
		f-TSST	Prep	612
			Pause_1	219
			Talk	664
			Math	668
			Pause_5	190
GUARDIAN Dataset	**24**	-	Resting	1,284
			Valsalva	1,318
			Apnea	1,377
			Tilt Up	1,541
			Tilt Down	1,244
Total				**11,768**

3.4.1 Q-Wave onset

Since the Q-Wave is not visible in the ECG of every person, the choice of the Q-Wave onset as the starting point of the PEP leads to problems in terms of the comparability of results between participants [She90]. For this reason, the R-onset was chosen as the starting point in this work, based on related work based on previous work, as it was found to be more reliably to detect [Ber04; See16]. The R-onset is defined as the abrupt change in gradient before the R-peak, which in the case of an existing Q-Wave corresponds with its peak [See16]. To simplify the identification of the points the preprocessed signal as described in Section 3.3.1 was used.

3.4.2 B-Point

The detection of the B-Point is challenging due to the variation in its morphology [She90; For19]. In order to achieve the highest possible accuracy in determining the B-Points, the guidelines for visual B-Point detection by Arbol et al. were followed, and the examples of difficult-to-identify B-Points by Forouzanfar et al. were consulted [Árb17; For19]. Additionally, the R-Peak was used as a reference point for the position of the B-Point as a lower limit, which is why the ECG was additionally plotted together with the ICG when determining the B-Points (Figure 3.17. The ECG signal is shown in red in the figure, and the ICG signal is in blue. Additionally, the heartbeat borders (marked with 1) and the labeled B-Points (marked with 2) are visible.

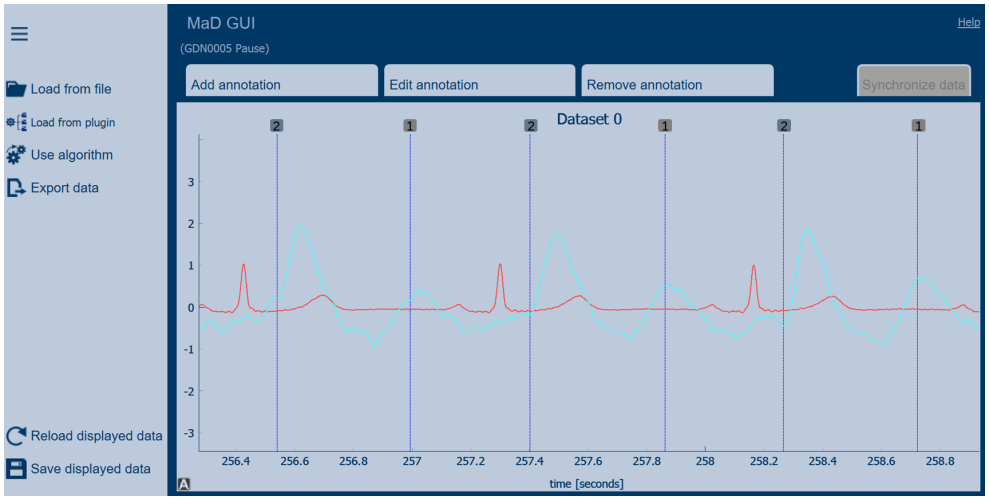

Figure 3.17: Example of labeled ICG signal in the *MaD GUI* package

3.5 Evaluation

Utilizing the *tpcp* Python package, pipelines were created by combining all Q-Wave onset algorithms and all B-Point algorithms, with the optional inclusion of one of the two Outlier Correction methods [Küd23]. These pipelines were evaluated regarding their performance in determining the PEP duration. Including the adapted variants of the RP-40 algorithm, a total of 72 algorithm combinations were analyzed in this regard.

In addition to assessing the pipelines' ability to accurately determine the PEP duration, the Q-Wave onset and B-Point algorithms were individually evaluated for their precision in identifying the respective points. The performance evaluation was conducted using the scoring implementation of the *tpcp* library [Küd23]. To assess the performance, the following metrics were calculated

based on the difference between the reference PEP values and the PEP values determined by the algorithms: the Mean Absolute Error (MAE) (3.1) and Mean Error (ME) (3.2), each accompanied by their standard deviation, and the Relative Error (RE) (3.3). Additionally, the number of cardiac cycles for which a PEP duration along with the rate of missing cardiac cycles was recorded. All these metrics were determined for each algorithm combination across all datapoints, where a datapoint refers to a specific phase of a particular participant. In addition to the overall results across all datapoints, the scoring functionality also provided the calculated values for each individual datapoint.

Moreover, for each heartbeat within every datapoint, the manually labeled start and end points, as well as the algorithmically detected points, were saved alongside the resulting PEP duration. This detailed recording faciliated a direct comparison between the manual and algorithmic detections for every single heartbeat.

All calculated error metrics were assessed for both datasets to determine the pipeline that provides the most accurate estimation of PEP duration for each dataset. Among these metrics, the MAE was considered the most critical. The MAE quantifies the absolute difference between reference values and computed values, providing a clear and unbiased measure of error. Unlike other metrics, the MAE does not allow positive and negative discrepancies to offset each other, thereby ensuring a true representation of estimation accuracy concerning the difference between the reference PEP duration and the calculated PEP duration. Additional insights were derived from the other error metrics, allowing conclusions to be drawn about whether the PEP duration was underestimated or overestimated and how robust and reliable the pipeline is in determining a PEP duration for each heartbeat.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\text{PEP\_manual,i} - \text{PEP\_calculated,i}| \tag{3.1}$$

$$\text{ME} = \frac{1}{n}\sum_{i=1}^{n} (\text{PEP\_manual,i} - \text{PEP\_calculated,i}) \tag{3.2}$$

$$\text{RE} = \frac{1}{n}\sum_{i=1}^{n} \frac{|\text{PEP\_manual,i} - \text{PEP\_calculated,i}|}{\text{PEP\_manual,i}} \tag{3.3}$$

Chapter 4

Results & Discussion

To assess the performance of various algorithm pipelines on the datasets, the relevant points for PEP calculation were manually labeled in a total of $11{,}768$ heartbeat cycles ($5{,}004$ for the TSST Dataset and $6{,}764$ for the GUARDIAN Dataset) in this study. This chapter presents an overview of the results obtained and discusses relations between findings of the two datasets.

4.1 Reference PEP data

Concerning the TSST Dataset, the mean and standard deviation of the reference PEP were 88.42 ms and 24.99 ms, with values ranging from 26 ms to 266 ms regarding the PEP duration. Distinguishing between TSST and f-TSST a mean value of 84.73 $\pm$ 24.08 ms was obtained for the TSST, while it was slightly higher for the f-TSST (92.57 $\pm$ 25.35 ms). For both conditions the minimum PEP mean values occurred for *Pause_1* and the maximum for *Pause_5*. A detailed list of all values for the individual phases per condition is given in Table 4.1, along with a visual representation in Figure 4.1.

In the GUARDIAN Dataset, the average PEP duration across all $6{,}764$ labeled cardiac cycles was 138.80 ms $\pm$ 27.03 ms, with a range between 36 ms and 292 ms. When breaking down into the individual phases the highest mean value was obtained for the *TiltUp* phase with 154.18 ms $\pm$ 22.70 ms. There were only minor differences in the mean values among the remaining phases, with the smallest mean value belonging to the *Resting* phase at 132.71 ms. When split into the phases, it was also observed that the smallest PEP value of 36 ms was recorded in 2 phases: *Resting* and *Apnea*. The *Apnea* phase also exhibited the maximum PEP value of 292 ms, showing a wide range of PEP duration during this phase.

Table 4.1: Reference PEP of the TSST Dataset divided into condition and phases

		Mean [ms]	**Min** [ms]	**Max** [ms]
TSST	**Prep**	81.69 ± 21.45	27	187
	Pause_1	76.61 ± 21.79	30	150
	Talk	85.70 ± 28.37	27	266
	Math	84.87 ± 20.62	31	144
	Pause_5	99.35 ± 22.48	52	161
	Total	84.73± 24.08	27	266
f-TSST	**Prep**	87.99 ± 24.35	33	167
	Pause_1	78.85 ± 20.67	27	165
	Talk	93.88 ± 24.67	26	180
	Math	94.59 ± 25.70	27	163
	Pause_5	111.49 ± 21.28	32	161
	Total	92.57 ± 25.35	26	180

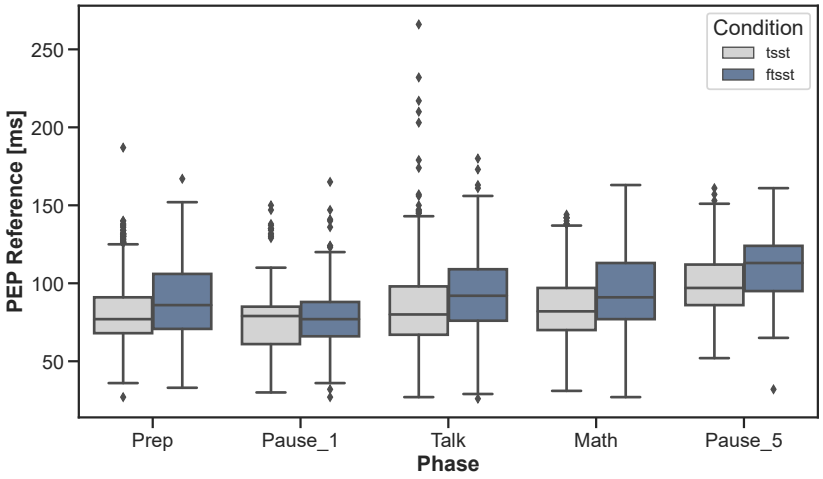

Figure 4.1: Box plot of the reference PEP data of the TSST Dataset divided into conditions and phases

Table 4.2 shows a precise breakdown of all parameters for the individual phases and a visualization is provided in Figure 4.2.

Discussion

At 138.80 ms, the mean PEP duration of the GUARDIAN Dataset was 56.8% longer than the one of the TSST Dataset, for which it was 88.49 ms. A possible explanation for this circumstance is

Table 4.2: Reference PEP data of the GUARDIAN Dataset divided into phases

	Mean [ms]	**Min** [ms]	**Max** [ms]
Resting	132.71 ± 25.41	36	208
Vasalva	134.99 ± 25.28	42	200
Apnea	133.76 ± 30.32	36	292
TiltUp	154.18 ± 22.70	44	250
TiltDown	135.63 ± 24.27	48	216
Total	138.80 ± 27.03	36	292

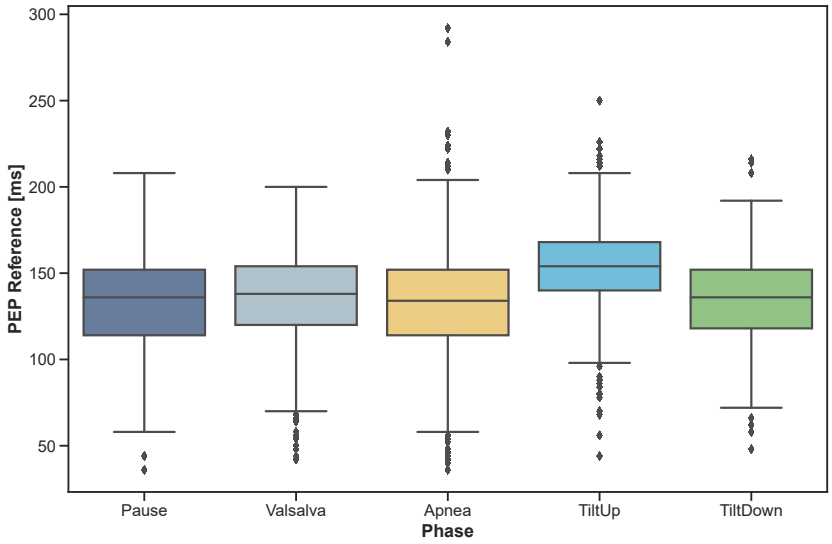

Figure 4.2: Box plot of the reference PEP data of the GUARDIAN Dataset divided into phases

the higher average age of the participants in the GUARDIAN Dataset, as a higher age is correlated with a longer PEP [New79].

The shorter PEP duration in every phase for the TSST in relation to the f-TSST in the TSST Dataset reflect the induced stress and points out that the PEP is capable of providing information about the stress level. For the GUARDIAN Dataset the observation of the highest mean value for the *TiltUp* phase is indicative of the physical challenge for the body experienced during the TTT.

Since the normal range for PEP duration is between 70 ms and 175 ms, both the overall mean and the mean values of the individual phases of both datasets are within this range [Árb17].

4.2 Fiducial Point Detection Algorithms

In total 72 algorithm combinations were tested regarding their ability to extract the PEP duration from ECG and ICG signals. These 72 pipelines were created by pairing each of the six Q-Wave onset algorithms (including the algorithms with adjusted time-intervals for subtraction from the R-Peak) with one of the four B-Point algorithm as well as with one of three Outlier Correction algorithms (none, Intpol, AutReg). To be able to compare the algorithms performances, the Mean Absolute Error (MAE), ME and the RE was calculated for each pipeline. Additionally, the number of cardiac cycles in which the algorithms were able to calculate a PEP duration was also examined (represented by #CC in the following Tables). Since the calculated values were subtracted from the reference data to obtain the ME, a positive value represents a shorter PEP duration calculated by the algorithms in comparison to the reference, while a negative value indicated a longer PEP duration. An overview of the results for the 36 pipelines, constructed using the QP, RP, and RP-40 as Q-Wave onset algorithms for the TSST dataset, is provided in Table 4.3.

Table 4.3: Overview of the performance of the 36 PEP computation pipelines using the QP, RP, and RP-40 as Q-Wave onset algorithms for the TSST Dataset. The best-performing pipeline with respect to the lowest MAE is highlighted in **bold**

Q-Wave onset	B-Point	Correction	**MAE** [ms]	**ME** [ms]	**RE**	**#CC**
QP	MultCon	None	21.74 ± 23.65	15.88 ± 27.93	24.98% ± 25.98%	4,545
		AutReg	20.80 ± 22.38	13.37 ± 27.47	23.99% ± 25.30%	4,676
		Intpol	20.60 ± 22.25	12.64 ± 27.56	23.86% ± 25.37%	4,674
	SecDer	None	17.83 ± 18.57	1.18 ± 25.71	22.79% ± 30.23%	4,480
		AutReg	18.34 ± 19.02	-0.76 ± 26.41	24.41% ± 31.54%	4,852
		Intpol	18.53 ± 19.31	-1.41 ± 26.72	24.74% ± 32.23%	4852
	StrLin	None	19.00 ± 17.80	-14.38 ± 21.70	24.45% ± 26.77%	4,836
		AutReg	19.51 ± 18.18	-16.82 ± 20.69	25.55% ± 28.95%	4,847
		Intpol	19.69 ± 18.32	-17.15 ± 20.71	25.80% ± 29.23%	4,847
	ThirDer	None	27.56 ± 21.17	-23.99 ± 25.14	35.97% ± 35.65%	4,817
		AutReg	29.14 ± 21.85	-27.85 ± 23.47	38.60% ± 38.06%	4,840
		Intpol	29.45 ± 21.94	-28.27 ± 23.44	39.02% ± 38.27%	4,840

Q-Wave onset	B-Point	Correction	MAE [ms]	ME [ms]	RE	#CC
RP	MultCon	None	47.39 ± 21.93	47.15 ± 22.45	54.18% ± 22.40%	4,152
		AutReg	46.73 ± 22.11	46.10 ± 23.40	53.15% ± 22.62%	4,395
		Intpol	46.23 ± 21.99	45.50 ± 23.46	52.68% ± 22.76%	4,412
	SecDer	None	41.27 ± 15.24	36.94 ± 23.89	47.52% ± 19.77%	4,617
		AutReg	39.83 ± 16.35	35.37 ± 24.55	47.90% ± 23.85%	5,000
		Intpol	39.46 ± 16.54	34.73 ± 24.99	47.48% ± 24.16%	5,000
	StrLin	None	22.92 ± 15.85	21.19 ± 18.10	25.91% ± 16.06%	4,953
		AutReg	21.91 ± 14.57	19.21 ± 17.98	24.93% ± 15.91%	4,991
		Intpol	21.71 ± 14.48	18.89 ± 18.01	24.73% ± 15.98%	4,991
	ThirDer	None	18.33 ± 14.73	11.10 ± 20.73	21.68% ± 18.86%	4,901
		AutReg	17.21 ± 13.34	8.15 ± 20.19	20.63% ± 18.83%	4,976
		Intpol	17.09 ± 13.27	7.75 ± 20.20	20.52% ± 18.87%	4,977
RP-40	MultCon	None	20.64 ± 21.86	12.67 ± 27.26	23.86% ± 23.53%	4,743
		AutReg	19.70 ± 20.44	10.08 ± 26.54	22.90% ± 22.96%	4,861
		Intpol	19.54 ± 20.36	9.40 ± 26.61	22.81% ± 23.10%	4,860
	SecDer	**None**	**14.55 ± 19.20**	**-3.06 ± 23.89**	**19.10% ± 31.53%**	**4,617**
		AutReg	15.64 ± 19.48	-4.63 ± 24.55	21.37% ± 32.96%	5,000
		Intpol	16.05 ± 19.86	-5.27 ± 24.99	21.98% ± 33.80%	5,000
	StrLin	None	22.32 ± 14.26	-18.28 ± 19.16	29.10% ± 23.71%	4,989
		AutReg	22.91 ± 15.23	-20.74 ± 18.06	30.30% ± 26.75%	4,995
		Intpol	23.16 ± 15.38	-21.05 ± 18.16	30.62% ± 26.98%	4,996
	ThirDer	None	31.18 ± 18.02	-27.55 ± 23.18	40.94% ± 33.12%	4,984
		AutReg	32.77 ± 18.83	-31.57 ± 20.77	43.56% ± 35.82%	4,994
		Intpol	33.10 ± 18.91	-31.97 ± 20.77	44.02% ± 36.02%	4,995

In addition to the RP-40 algorithm proposed by van Lien et al., which subtracts 40 ms from the R-Peak to determine the PEP start point, several adjusted algorithms that subtract 32, 34, and 36 ms were also tested [Lie13]. The results for the TSST dataset, obtained from the 36 pipelines constructed using these adjusted Q-Wave onset algorithms, are presented in TableA.1.

The lowest MAE of 14.55 ms ± 19.20 ms, was exhibited by the combination of the RP-40 and SecDer method without Outlier Correction. Furthermore, this pipeline achieved the lowest RE with 19.10% ± 31.53%. Using this combination, the PEP duration was determined for 4,617

cardiac cycles. With a MAE of 47.39 ms ± 21.93 ms, the worst performance was achieved by the pipeline consisting of the RP and MultCon algorithms without Outlier Correction. Additionally, this combination had the lowest number of cardiac cycles (4,152) for which a PEP duration was determined, corresponding to a missing heartbeat rate of 17.03%. The lowest rate of missing heartbeats, with 0.08% (5,000 out of 5,004 cardiac cycles), was achieved by the RP and all of the methods that subtract a certain time interval from the R-Peak in combination with the SecDer method and Outlier Correction.

The performances of the individual pipelines for the GUARDIAN Dataset were evaluated accordingly (Table 4.10). The results of the pipelines using the adjusted versions of the RP-40 algorithm to determine the Q-Wave onset can be found in Table A.2. The lowest MAE for this dataset was 18.41 ms ± 16.32 ms, achieved by the pipeline consisting of the RP-32 and StrLin algorithm without Outlier Correction. With a total of 6,754 cardiac cycles for which the PEP could be determined using these algorithms, the missing heartbeat rate was 0.15%. The RP algorithm in combination with MultCon algorithm and the use of the AutReg method for Outlier Correction exhibited the highest MAE of 53.60 ms ± 38.68 ms. Furthermore, this combination exhibited the lowest number of cardiac cycles (5,727) in which a PEP duration could be determined, which is equivalent to no PEP being extracted for 15.33% of the heartbeats.

Table 4.4: Overview of the performance of the 36 PEP computation pipelines using the QP, RP, and RP-40 as Q-Wave onset algorithms for the GUARDIAN Dataset

Q-Wave onset	B-Point	Correction	MAE [ms]	ME [ms]	RE	#cc
QP	MultCon	None	27.85 ± 35.89	19.78 ± 40.90	20.29% ± 25.35%	5,900
		AutReg	31.05 ± 38.34	22.04 ± 44.14	22.58% ± 27.09%	6,442
		Intpol	30.22 ± 37.82	20.45 ± 43.88	22.10% ± 26.95%	6,441
	SecDer	None	25.82 ± 26.60	-13.25 ± 34.62	20.87% ± 26.34%	6,643
		AutReg	26.31 ± 26.84	-15.10 ± 34.42	21.33% ± 26.72%	6,647
		Intpol	27.07 ± 27.13	-16.38 ± 34.65	22.00% ± 27.31%	6,647
	StrLin	None	20.44 ± 18.53	-16.84 ± 21.85	16.87% ± 20.58%	6,646
		AutReg	21.27 ± 19.04	-18.45 ± 21.79	17.68% ± 21.45%	6,647
		Intpol	21.60 ± 19.33	-18.91 ± 21.97	17.97% ± 21.73%	6,647
	ThirDer	None	26.30 ± 23.37	-16.89 ± 30.86	21.39% ± 24.86%	6,645
		AutReg	26.41 ± 23.12	-21.36 ± 27.85	21.82% ± 25.51%	6,645
		Intpol	27.19 ± 23.39	-22.88 ± 27.62	22.53% ± 26.01%	6,645

Q-Wave onset	B-Point	Correction	**MAE** [ms]	**ME** [ms]	**RE**	**#cc**
RP	MultCon	None	50.27 ± 34.05	49.88 ± 34.62	36.11% ± 22.79%	5,727
		AutReg	53.60 ± 38.68	52.98 ± 39.52	38.31% ± 26.09%	6,305
		Intpol	52.31 ± 38.13	51.49 ± 39.22	37.40% ± 25.82%	6,312
	SecDer	None	35.36 ± 17.04	21.43 ± 32.88	26.51% ± 16.71%	6,752
		AutReg	33.92 ± 17.27	19.61 ± 32.62	25.34% ± 16.66%	6,756
		Intpol	33.44 ± 17.46	18.35 ± 32.96	25.00% ± 17.04%	6,756
	StrLin	None	21.56 ± 15.61	17.72 ± 19.86	15.59% ± 12.78%	6,752
		AutReg	20.78 ± 14.74	16.13 ± 19.72	15.13% ± 12.69%	6,752
		Intpol	20.69 ± 14.65	15.68 ± 19.92	15.11% ± 12.81%	6,753
	ThirDer	None	26.53 ± 22.36	17.72 ± 29.83	20.05% ± 18.96%	6,752
		AutReg	23.19 ± 18.58	13.32 ± 26.56	17.64% ± 17.10%	6,752
		Intpol	22.55 ± 18.04	11.83 ± 26.35	17.26% ± 17.08%	6,752
RP-40	MultCon	None	28.69 ± 34.33	15.68 ± 41.90	21.19% ± 24.22%	6,057
		AutReg	31.41 ± 36.18	17.64 ± 44.55	23.16% ± 25.69%	6,598
		Intpol	30.73 ± 35.82	16.12 ± 44.36	22.79% ± 25.79%	6,599
	SecDer	None	24.66 ± 28.60	-18.57 ± 32.88	20.44% ± 28.91%	6,752
		AutReg	25.61 ± 28.70	-20.39 ± 32.62	21.30% ± 29.24%	6,756
		Intpol	26.63 ± 29.08	-21.65 ± 32.96	22.22% ± 29.96%	6,756
	StrLin	None	25.56 ± 15.69	-22.21 ± 20.15	21.20% ± 21.16%	6,755
		AutReg	26.39 ± 16.45	-23.79 ± 20.02	22.01% ± 22.09%	6,756
		Intpol	26.74 ± 16.79	-24.25 ± 20.21	22.31% ± 22.37%	6,756
	ThirDer	None	30.75 ± 21.13	-22.24 ± 29.96	25.17% ± 25.44%	6,754
		AutReg	31.18 ± 21.23	-26.64 ± 26.72	25.88% ± 26.37%	6,754
		Intpol	32.06 ± 21.58	-28.13 ± 26.50	26.66% ± 26.95%	6,754

The highest number of cardiac cycles that could be achieved in this dataset was $6{,}756$. This number occurred for several pipelines.

Discussion

In the TSST dataset, the best-performing combination failed to extract a PEP duration for a relatively high number of heartbeats (7.65%). By using AutReg, the number of detected PEP duration increased from $4{,}617$ to $5{,}000$, the maximum achieved by any of the pipelines. While this led to a slightly higher MAE of 15.64 ms, it may be more beneficial to use this combination for achieving the most complete detection of PEP values on the entire signal. The MAE of this combination was the third smallest among all pipelines.

The combination of the RP-36 and the SecDer method without Outlier Correction performed second best in terms of MAE (15.44 ms), but it detected the same number of heartbeats as the best combination, and is therefore offering no additional advantage.

The algorithm combination that identified the smallest and the largest number of cardiac cycles with a determined PEP duration was consistent across both datasets. This indicates that the algorithms performance in this regard is stable across different datasets.

To perform Outlier Correction on the detected B-Points, both an autoregressive model and an Intpol algorithm were implemented, following approaches by previous work [For19]. When looking at the number of heart cycles in which the algorithms were able to determine a PEP, it is evident across both datasets and all pipelines that both methods were able to increase the number of detected cycles. Notably, the combinations that used the MultCon algorithm for B-Point detection stand out. Using Outlier Correction with these combinations resulted in a considerably higher number of cardiac cycles where a valid PEP was extracted. Maximum improvement was achieved with the combination of the RP and MultCon algorithms in the GUARDIAN dataset by applying Intpol to the detected B-Points. The $5{,}727$ cardiac cycles could be increased to $6{,}305$, which corresponds to an increase of 10.09%.

However, both Outlier Correction methods improved the results concerning MAE only in a few cases. One possible explanation is that the points were determined using only the parts of the signal within the randomly selected intervals for manual labeling, which were either 10s, 30s or 60s, depending on the phase. In particular, the autoregressive model could benefit from a longer signal sequence to make more accurate forward and backward estimations [Bro04]. This limited data availability likely restricted the model's ability to effectively correct outliers. Consequently, the improvements in results were not as prominent as expected when using these correction methods. However, it can be argued that the Outlier Correction allowed for the inclusion of more points in the error value calculations, as it enabled the determination of the PEP for a higher number of cardiac cycles. The relatively small increase in the values of MAE, ME, and RE suggests that the newly identified values through Outlier Correction were determined with good accuracy with

respect to the reference data. Therefore the higher MAE cannot generally be associated with poorer performance. It is also notable that the values of the RE align with the results of the MAE, in the form that a smaller MAE corresponded to a smaller RE, and vice versa.

4.2.1 Q-Wave onset

In addition to evaluating the ability of the different algorithm combinations to determine the PEP duration, the accuracy of the Q-Wave onset algorithms in pinpointing this specific point was also assessed. To isolate the accuracy of the Q-Wave onset algorithms, the detected start points were combined with the manually labeled B-Points to calculate the PEP. This approach isolates the error arising solely from the start points in determining the PEP duration. Information about the MAE, the ME, and the RE for every Q-Wave onset detection algorithm, including the modified versions of the algorithm by Van Lien et al., using time intervals of 32, 34, and 36 ms instead of 40 ms, are listed in Table 4.5 [Lie13].

Table 4.5: Performance of the Q-Wave onset algorithms for the TSST Dataset. The best-performing algorithm with respect to the lowest MAE is highlighted in **bold**

Q-Wave onset	**MAE** [ms]	**ME** [ms]	**RE**	**#CC**
QP	**5.29 ± 9.55**	**-4.42 ± 9.98**	**6.76% ± 13.31%**	**4,856**
RP	31.59 ± 6.70	31.59 ± 6.70	37.88% ± 11.56%	5,004
RP-32	5.70 ± 3.56	-0.41 ± 6.70	7.26% ± 5.65%	5,004
RP-34	5.61 ± 4.39	-2.41 ± 6.70	7.36% ± 7.12%	5,004
RP-36	6.04 ± 5.28	-4.41 ± 6.70	8.09% ± 8.56%	5,004
RP-40	8.76 ± 6.24	-8.41 ± 6.70	11.63% ± 10.52%	5,004

The smallest MAE within this Dataset was found for the QP algorithm with 5.29 ms ± 9.55 ms. The ME and its standard deviation appeared at -4.42 ms and 9.98 ms. This algorithm showed with $4{,}856$ the smallest amount of cardiac cycles in which it was able to detect a point. This represents a 3.00% rate of heartbeats for which no start point was detected.

The second-best algorithm for TSST Dataset was the RP-34 algorithm, with a slightly larger MAE of 5.61 ms. However, the ME was smaller, at -2.41 ms, compared to the QP algorithm. Notably, the standard deviation for the RP-34 algorithm, as with all R-Peak based algorithms, was also slightly smaller than that of the QP algorithm.

The highest MAE along with the highest ME and RE was calculated for the RP algorithm. All methods based on the R-Peak were able to determine a starting point for all labeled cardiac cycles ($5{,}004$). Figure 4.3 contains the residual plots of all Q-Wave onset algorithms. Figures B.1 and B.2 display the same plots with color differentiation between the individual participants and phases.

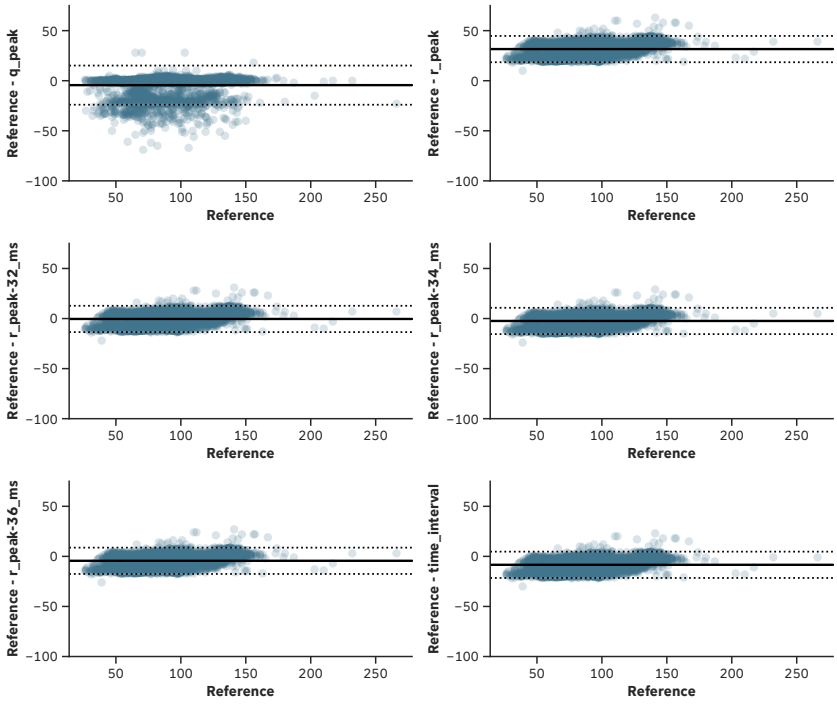

Figure 4.3: Residual plots of the Q-Wave onset algorithms for the TSST Dataset

Similarly to the TSST Dataset, Table 4.6 contains the MAE, ME, and RE along with the amount of cardiac cycles for all Q-Wave onset algorithms along with the modified versions of the RP-40 algorithm for the GUARDIAN Dataset. For this dataset, the QP algorithm demonstrated the lowest MAE of 4.15 ms $\pm$ 12.31 ms. With 2.99% $\pm$ 9.18% the RE was smallest for this algorithm as well. The number of heartbeats for which a starting point could be determined was $6{,}655$. Thus, the QP algorithm exhibits the highest rate of missing cardiac cycles at 1.6% for this dataset. All other Algorithms were able to detect a start point for all $6{,}764$ cardiac cycles.

For the GUARDIAN Dataset, the RP-32 method achieved the second-lowest MAE of 4.81 ms. The ME was -0.44 ms $\pm$ 7.11 ms, which was smaller than the one of the QP Algorithm. It can also be noted here that the standard deviation of the algorithms based on the R-Peak are smaller compared to the QP algorithm.

In contrast, the RP algorithm showed the highest error values for the GUARDIAN Dataset, with an MAE of 31.67 ms and an ME of 31.56 ms, along with the highest RE of 23.61% among all the Q-Wave onset algorithms.

Table 4.6: Performance of the Q-Wave onset algorithms for the GUARDIAN Dataset. The best-performing algorithm with respect to the lowest MAE is highlighted in **bold**

Q-Wave onset	**MAE** [ms]	**ME** [ms]	**RE**	**#CC**
QP	**4.15 ± 12.31**	**-3.09 ± 12.62**	**2.99% ± 9.18%**	**6,655**
RP	31.67 ± 6.62	31.56 ± 7.11	23.61% ± 7.15%	6,764
RP-32	4.81 ± 5.26	-0.44 ± 7.11	3.78% ± 5.07%	6,764
RP-34	4.97 ± 5.65	-2.44 ± 7.11	3.98% ± 5.46%	6,764
RP-36	5.65 ± 6.20	-4.44 ± 7.11	4.55% ± 5.95%	6,764
RP-40	8.56 ± 6.97	-8.44 ± 7.11	6.77% ± 6.77%	6,764

For this dataset, Figure 4.4 provides the residual plots of the Q-Wave onset algorithms, along with a breakdown between the individual participants and phases in Figures B.3 and B.4, respectively.

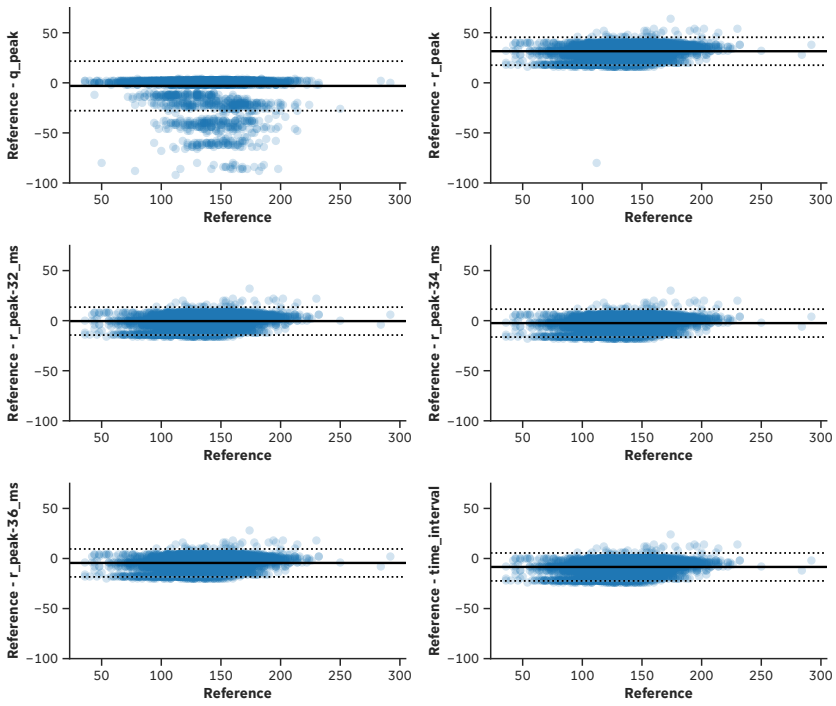

Figure 4.4: Residual plots of the Q-Wave onset algorithms for GUARDIAN Dataset

Discussion

The best-performing algorithm for pinpointing the PEP start points in both dataset was the QP algorithm. Given that the onset of the R-Peak was used for the manually labeled data, as well as within the QP algorithm, it seems logical that this algorithm showed the smallest average deviation from the reference. However, despite the QP algorithm having the MAE, it had the lowest amount of cardiac cycles in which it was able to identify a point.

Additionally, for both datasets, the RP algorithm consistently exhibited the poorest performance. Since the proposal to use the R-Peak as start point of the PEP is based on its simple and reliable detection and not on its physiological accuracy, this is not surprising. The poor performance is also evidenced by the fact that the RP algorithm was the only one that failed to accurately pinpoint the PEP start point in any cardiac cycle of both datasets without any deviation. In contrast, all other algorithms were able to detect the correct point in several heartbeats of both datasets.

It was also found that algorithms that exhibited low MAE also demonstrated low RE, and vice versa. Consequently, the relative error RE mirrors the findings of the MAE.

For the start algorithms, a positive ME value indicates that the Q-Wave onset was set later than in the reference data, resulting in a shorter PEP. Since the calculated PEP is subtracted from the reference, this produces a positive ME value. Conversely, a longer PEP, indicating an earlier Q-Wave onset, is represented by negative ME values. The negative values for ME for each algorithms except the RP one shows that these set the Q-Wave onset earlier than it is in the reference data, resulting in a longer PEP duration. The positive ME of the RP algorithm is expected, as the R-Peak occurs after the Q-Wave onset.

Examining the ME revealed that the algorithms with the second-lowest MAE for both datasets, had a smaller ME than the QP algorithm. This indicates that the QP algorithm produces more values that deviate noticeably in the negative direction from the ground truth. However, these negative deviations are balanced out by positive deviations in the MAE calculation. This observation is also evident in the Figures 4.3 and 4.4, where the plots for the QP algorithm show more outliers, particularly in the negative y-axis, than the algorithms based on subtracting a time period from the R-Peak. Upon examination of Figure B.3, it is evident from the residual plot of the QP algorithm that numerous outliers can be attributed to participants identified as “GDN0009” and “GDN0022” in the GUARDIAN Dataset. This observation suggests that the inaccurate detections were likely caused by artifacts or signal interference in the data from these participants. Since the R-Peak is easier to identify, the accuracies of the remaining algorithms were not affected by this. When divided into phases (Figure B.4), the outliers can primarily be attributed to the *TiltUp* and *TiltDown*

phases. Since these phases involve movement of the participant, this further supports the hypothesis that increased disturbances, such as noise in the signal, occurred in the cases of the outliers.

This assumption can be confirmed investigating the ECG signal more closely. Figure 4.5 shows a part of the ECG signal of participant "GDN0009" (top), where in comparison to the ECG signal of participant "GDN0030" (bottom), more noise can be observed especially in the area of the Q-Wave.

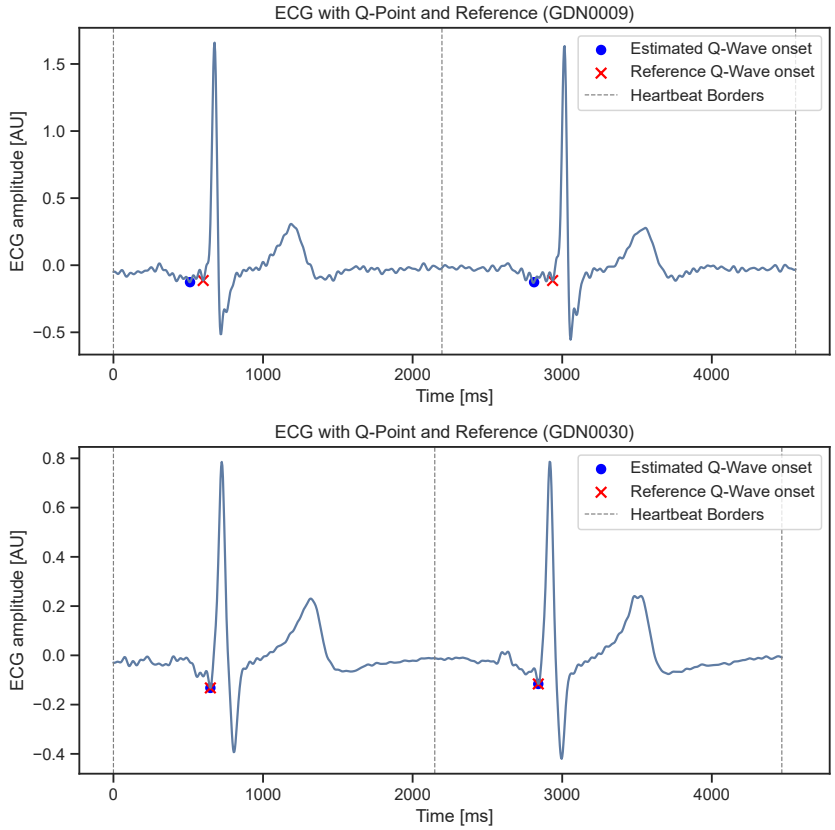

Figure 4.5: Selected ECG signal period of participant "GDN0009" (top) and "GDN0030" (bottom) from the GUARDIAN Dataset

In the TSST dataset, the outliers could not be attributed to few individual participants or phases. Generally, Figure B.1 and B.2 reveal fewer outliers with high negative deviation for the QP algorithm within this dataset. Nevertheless, the observed outliers can similarly be attributed to signal interference. However, in this case, the disturbances affect smaller segments of the signal rather than the entire signal of a participant.

Upon examination of the different adaptations of the algorithm that subtracts a specific time interval from the R-Peak, it was observed that the version with a 40ms interval exhibits the largest

MAE, ME, and RE among all algorithms of this type across both datasets. Therefore, for both datasets, the interval of 40ms proposed by Van Lien et al. could not be supported, since this time interval did not prove to be the best for subtraction in either dataset [Lie13].

4.2.2 B-Point

Using the same principle as for the Q-Wave Onset algorithms, the error caused only by the end points was determined for the B-Point algorithms. For this purpose, the B-Points detected with the different algorithms and possibly Outlier Correction were combined with the manually labeled start points to calculate the PEP. An overview of the results obtained for the TSST Dataset can be found in Table 4.7. The StrLin algorithm showed the lowest MAE of 14.92 ms $\pm$ 14.94 ms. The ME and its standard deviation of this algorithm were -9.99 ms and 18.60 ms, respectively. Furthermore, the StrLin algorithm had the lowest RE of 18.94% $\pm$ 20.80%. With a total of $4{,}983$ cardiac cycles for which a B-Point was identified, this algorithm has a rate of 0.42% missing heartbeats. The highest MAE with 25.03 ms $\pm$ 18.42 ms occurred for the ThirDer method in combination with Intpol for Outlier Correction. This pairing also showed the highest RE of 33.12% $\pm$ 32.12%.

The lowest ME of 3.14 ms $\pm$ 24.52 ms was achieved by the SecDer method while using Intpol to correct the found B-Points.

The lowest rate of heartbeats without a detected B-Point was achieved by using the SecDer algorithm with Outlier Correction. With $5{,}000$ detected B-Points, this rate was 0.08%. Figure 4.6 provides residual plots of every algorithm with either none or one of the two Outlier Correction methods. Additionally, two further plots were generated as shown in Figure B.5 and B.6, highlighting the different participants and phases.

Table 4.7: Performance of the B-Point algorithms for the TSST Dataset. The best-performing algorithm with respect to the lowest MAE is highlighted in **bold**

B-Point	**Correction**	**MAE** [ms]	**ME** [ms]	**RE**	**#CC**
MultCon	None	21.98 $\pm$ 24.91	20.26 $\pm$ 26.33	25.08% $\pm$ 26.89%	4,676
	AutReg	20.90 $\pm$ 23.37	17.87 $\pm$ 25.76	23.88% $\pm$ 25.74%	4,812
	Intpol	20.64 $\pm$ 23.18	17.20 $\pm$ 25.84	23.65% $\pm$ 25.69%	4,812
SecDer	None	17.08 $\pm$ 16.87	5.29 $\pm$ 23.41	21.82% $\pm$ 26.99%	4,617
	AutReg	17.40 $\pm$ 17.15	3.78 $\pm$ 24.14	23.27% $\pm$ 28.23%	5,000
	Intpol	17.57 $\pm$ 17.38	3.14 $\pm$ 24.52	23.55% $\pm$ 28.84%	5,000
StrLin	**None**	**14.92 $\pm$ 14.94**	**-9.99 $\pm$ 18.60**	**18.94% $\pm$ 20.80%**	**4,983**
	AutReg	15.31 $\pm$ 14.92	-12.36 $\pm$ 17.43	19.92% $\pm$ 22.97%	4,994
	Intpol	15.46 $\pm$ 15.04	-12.69 $\pm$ 17.44	20.13% $\pm$ 23.25%	4,994
ThirDer	None	23.38 $\pm$ 18.14	-19.68 $\pm$ 22.10	30.38% $\pm$ 29.85%	4,957
	AutReg	24.73 $\pm$ 18.44	-23.33 $\pm$ 20.18	32.72% $\pm$ 31.93%	4,986
	Intpol	25.03 $\pm$ 18.51	-23.74 $\pm$ 20.13	33.12% $\pm$ 32.12%	4,986

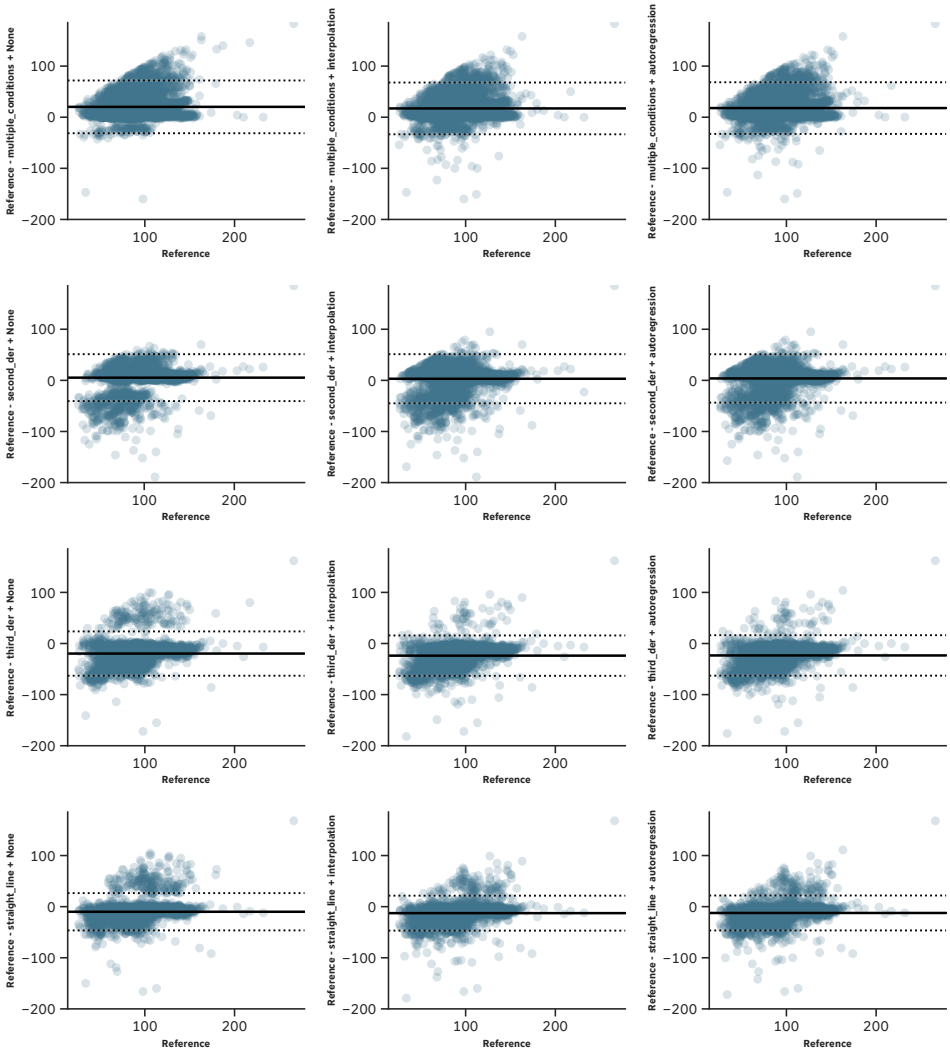

Figure 4.6: Residual plots of the B-Point algorithms for the TSST Dataset

Analogous to the TSST dataset, Table 4.8 presents the results of all B-Point algorithms for the GUARDIAN Dataset. The lowest MAE (17.79 pm 15.08) for this dataset was achieved by using the StrLin algorithm without Outlier Correction. The lowest RE of 14.93% $\pm$ 18.54% was also achieved using this algorithm. The associated ME was -13.82 ms $\pm$ 18.79 ms. Additionally, with $6{,}754$ detected B-Points, the StrLin algorithm had a rate of 0.15% cardiac cycles for which not B-Point could be determined. The highest MAE of 30.58 ms $\pm$ 40.19 ms was found for the MultCon algorithm with the usage of AutReg for Outlier Correction. This pairing showed with a value of 22.18% $\pm$ 28.12% also the highest RE within this dataset. The MultCon algorithm without Outlier Correction also exhibited the highest rate of heartbeats without a determined

B-Point (11.44%). The lowest ME was achieved by the SecDer algorithm with -10.16 ms ± 32.07 ms.

A visualization of the results in the form of a residual plot is available in Figure 4.7. Additional versions of this residual plot (B.7 and B.8) are provided, which highlight individual participants and phases.

Table 4.8: Performance of the B-Point algorithms for the GUARDIAN Dataset. The best-performing algorithm with respect to the lowest MAE is highlighted in **bold**

B-Point	**Correction**	**MAE** [ms]	**ME** [ms]	**RE**	**#CC**
MultCon	None	27.08 ± 37.69	22.62 ± 40.52	19.64% ± 26.26%	5,990
	AutReg	30.58 ± 40.19	24.98 ± 43.89	22.18% ± 28.12%	6,539
	Intpol	29.73 ± 39.63	23.43 ± 43.66	21.68% ± 27.95%	6,539
SecDer	None	23.47 ± 24.11	-10.16 ± 32.07	19.19% ± 24.65%	6,752
	AutReg	23.86 ± 24.29	-11.99 ± 31.87	19.58% ± 24.99%	6,755
	Intpol	24.55 ± 24.61	-13.26± 32.14	20.19% ± 25.61%	6,755
StrLin	**None**	**17.79 ± 15.08**	**-13.82 ± 18.79**	**14.93% ± 18.54%**	**6,754**
	AutReg	18.52 ± 15.62	-15.40 ± 18.70	15.67% ± 19.45%	6,755
	Intpol	18.82 ± 15.99	-15.86 ± 18.93	15.94% ± 19.77%	6,755
ThirDer	None	23.99 ± 21.35	-13.84 ± 28.98	19.68% ± 23.38%	6,753
	AutReg	23.75 ± 20.62	-18.24 ± 25.62	19.87% ± 23.86%	6,753
	Intpol	24.50 ± 20.92	-19.74 ± 25.47	20.56% ± 24.42%	6,753

Discussion

For both datasets, the StrLin algorithm did perform best in regard to the MAE. The second-best algorithm also matched for both datasets: the SecDer algorithm. The remaining two algorithms differed depending on the dataset: for the TSST Dataset, the MultCon approach by Forouzanfar et al. achieved better results, while for the GUARDIAN Dataset, the ThirDer algorithm performed better in calculating the correct B-Point. As with the Q-Wave onset algorithms, a correlation between MAE and RE was also observed for the B-Point algorithms. Higher MAE values were consistently associated with higher RE values, and vice versa.

In the case of the B-Point algorithms, a positive ME indicates that the point was found earlier compared to the reference. An earlier B-Point leads to a shorter PEP when using the same start point. Since the calculated time is subtracted from the reference, positive values result. Logically following, a negative value indicates that the points were found too late. When looking at the ME,

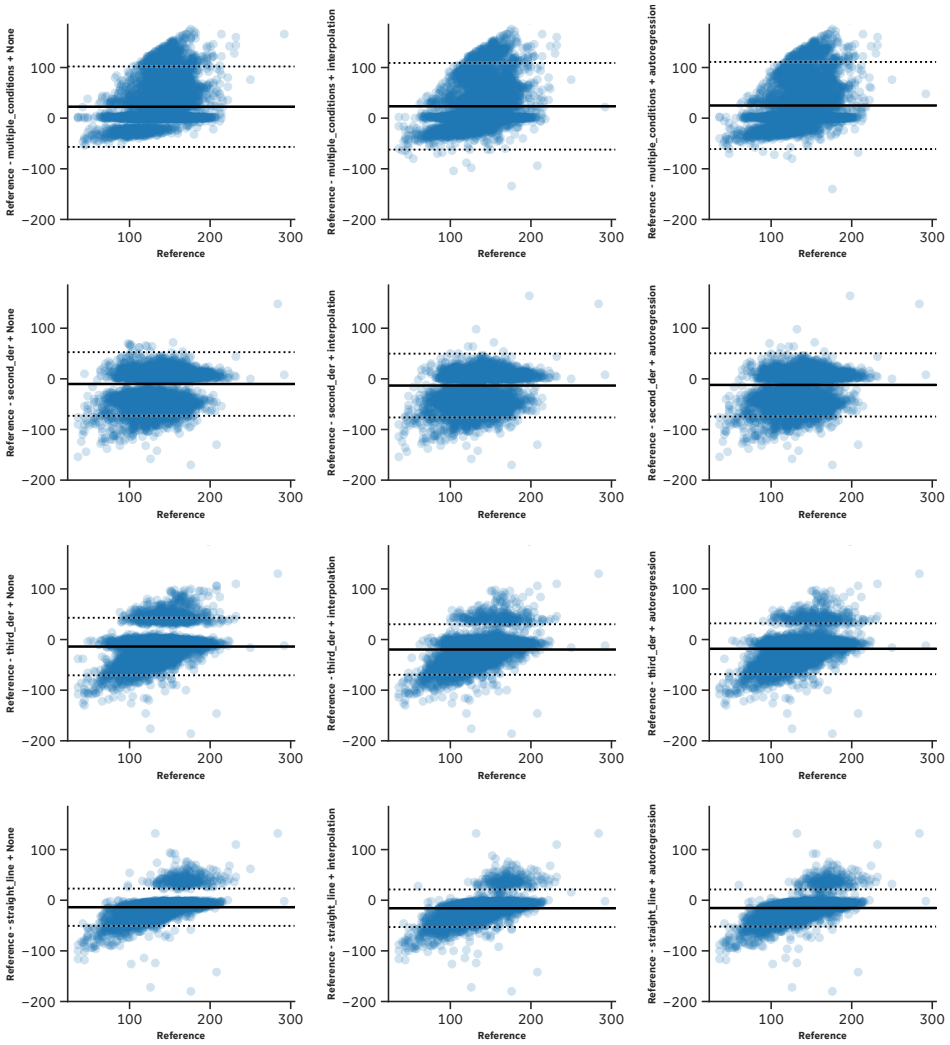

Figure 4.7: Residual plots of the B-Point algorithms for the GUARDIAN Dataset

it can be observed that for both datasets, the SecDer algorithm had the smallest absolute values, with 5.29 ms for the TSST Dataset and -10.16 ms for the GUARDIAN Dataset. The positive value for the TSST Dataset indicates that the B-Point was set earlier than in the reference, while for the GUARDIAN Dataset, the negative value indicates that it was set later than the manually labeled reference value. For the remaining algorithms, the sign of the ME is consistent for both datasets. Both the StrLin and ThirDer algorithms tend to detect the B-Point later than the reference value, while the MultCon algorithm tends to detect it earlier. These observations are also visually evident in the Figures 4.6 and 4.7 by the position of the offset.

Particularly in the residual plots containing the MultCon method, it can be seen that the deviation upwards is limited by the reference PEP length. Since the same start point was assumed

for both the reference and the determined PEP, and the B-Point cannot be located before the Q-Wave onset, the reference value itself represents the maximum possible difference in the positive direction when subtracting the calculated PEP duration from the reference. The fact, that this boundary is particularly noticeable in the plots of the MultCon algorithm, coincides with the ME of this algorithm over both datasets. The MultCon algorithm had the highest positive ME with 20.26 ms and 22.62 ms respectively. From this it can be deduced that this method has the strongest tendency to set the B-Point too early, closer to the Q-Wave onset, resulting in a shorter PEP. From the breakdown by participants and phases, no specific trends stand out for either dataset. Overall, the error values of the end-point algorithms were higher compared to the start-point algorithms, except for the RP algorithm. This suggests that the error in determining the PEP duration using the pipelines is more influenced by the end-point, which is understandable since the B-Point is considered difficult to identify correctly [She90; For19]. The higher errors are also apparent in the residual plots, where the outliers tend to deviate more substantially from the mean compared to those observed in the plots generated by the start-point algorithms.

The rate of identified points from all heartbeats is also lower for the B-Point algorithms compared to the Q-Wave onset algorithms. Consequently, it can be inferred that in this aspect as well, the B-Point algorithms have a greater impact on the results of the pipelines.

It is also noteworthy that using certain pipeline combinations resulted in a higher number of cardiac cycles where a PEP duration could be determined, compared to evaluating only the B-Point algorithms. For instance, in the TSST dataset, the MultCon approach combined with manually labeled start points was able to determine a PEP duration for $4{,}676$ heartbeats. In contrast, the combination of the RP-40 and MultCon approaches identified $4{,}743$ PEP values.

At first glance, this might seem counterintuitive. However, the explanation lies in the fact that the B-Point algorithm, when used with reference values, resulted in negative PEP duration for several heartbeats because the B-Point was detected before the Q-Wave onset. Since such negative duration are physiologically implausible, these heartbeats were classified as artefacts. By using the RP-40 algorithm, different start points were identified, which did not necessarily lead to negative values, thereby allowing the determination of a PEP duration for more cardiac cycles by this algorithm combination.

Furthermore, as already previously observed with the pipelines for determining the PEP duration, the Outlier Correction only improved the results in few cases with respect to the MAE. Concerning the TSST Dataset only the MAE of the MultCon algorithm was reduced from 21.93 ms to 20.60 ms through Intpol and to 20.85 ms by using the autoregressive approach. In the GUARDIAN Dataset, Outlier Correction improved the MAE only for the ThirDer method, with

a slight reduction from 23.83 ms to 23.65 ms through Intpol. For all other algorithms, Outlier Correction did not lead to a smaller MAE in either dataset. On the other hand, using Outlier Correction led to a higher number of cardiac cycles across all algorithms and both datasets. The most prominent increase was observed for the MultCon algorithm concerning the GUARDIAN Dataset: the number of cardiac cycles for which a B-Point was found increased from $5{,}861$ to $6{,}407$ through the use of the AutReg approach. This corresponds to an increase of 9.3%. As previously discussed, these slightly higher MAE values should not necessarily be interpreted as strictly negative. Since the values are only marginally higher when using Outlier Correction compared to without, and the number of cardiac cycles is also higher, it suggests that the newly identified points are determined with minimal deviation from the reference data. This indicates good performance of the Outlier Correction algorithms in this aspect.

4.2.3 Best-performing Algorithms

For the TSST Dataset, the combination of using the RP-40 algorithm for the start point and the SecDer method for determining the end point without Outlier Correction emerged as the most accurate pipeline. This combination showed a MAE of 14.55 ms ± 19.20 ms. Table 4.9 lists the PEP values calculated by this algorithm combination, categorized by condition and phase.

Table 4.9: PEP duration calculated by best-performing pipeline for the TSST Dataset divided into conditions and phases

		Mean [ms]	**Min** [ms]	**Max** [ms]
TSST	**Prep**	88.32 ± 28.10	41	327
	Pause_1	81.65 ± 21.31	41	159
	Talk	92.86± 32.48	41	242
	Math	88.04 ± 22.88	41	179
	Pause_5	97.15 ± 22.67	41	148
	Total	89.74± 27.39	41	327
f-TSST	**Prep**	95.07 ± 27.04	42	168
	Pause_1	83.90 ± 23.11	43	154
	Talk	99.95 ± 30.12	42	306
	Math	100.00 ± 28.50	41	247
	Pause_5	111.31 ± 22.31	50	155
	Total	98.34 ± 28.35	41	306

The mean duration of the by the best-performing pipeline computed PEP for the TSST was with 89.74 ms shorter than the calculated mean duration for the f-TSST with 98.34 ms. The highest estimated PEP duration was found within the *Prep* phase of the TSST where it amounted to 327 ms. The mean of the entire dataset, without subdivision into the two conditions, was 93.03 ms, which is 5.2% higher than the mean of the reference PEP data.

In Figure 4.8, the box plots of both the PEP duration calculated by the pipeline and the reference values are presented.

As previously discussed in Chapter (4.2), the number of cardiac cycles for which a PEP duration could be determined using this combination is relatively small. Therefore, it was suggested to additionally apply AutReg to increase this number. For the resulting combination of RP-40, SecDer, and AutReg, the PEP durations were also determined, segmented by condition and phase. The results are presented in Table A.3. Analogous to the version without AutReg, the box plots can be found in Figure B.9.

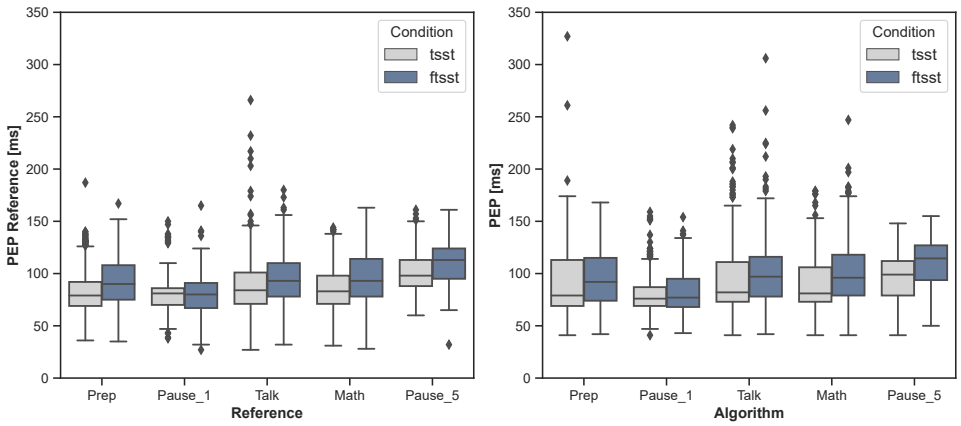

Figure 4.8: PEP duration per phase compared between best-performing pipeline of the TSST Dataset and the reference data

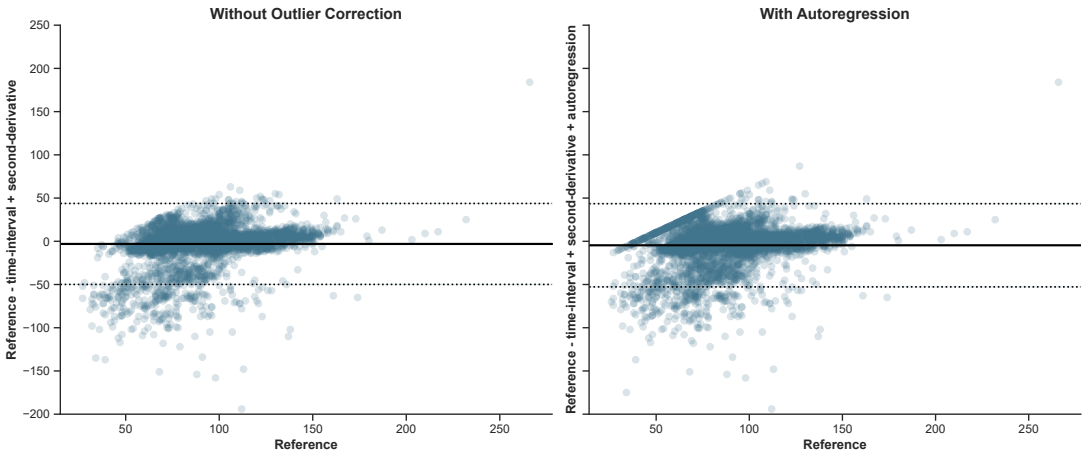

Figure 4.9: Residual plots of best-performing pipeline of the TSST Dataset without Outlier Correction (left) and with AutReg (right)

The residual plots in Figure 4.9 display the combination of the RP-40 and SecDer algorithms without outlier correction (left) and with AutReg (right).

For the GUARDIAN Dataset, the combination of the RP-32 and StrLin performed best with a MAE of 18.41 ms ± 16.32 ms. The PEP values determined by these algorithms are listed in Table 4.10, differentiated by phase. The mean of the calculated PEP values was 153.00 ms ± 20.73 ms, with a maximal duration of 356 ms. In comparison to the reference PEP data, the mean duration of the PEP calculated by the algorithm combination was 10.23% higher.

Table 4.10: PEP duration calculated by best-performing pipeline for the GUARDIAN Dataset divided into phases

	Mean [ms]	**Min** [ms]	**Max** [ms]
Resting	149.67 ± 16.70	84	200
Vasalva	149.27 ± 17.91	86	230
Apnea	147.82 ± 23.80	22	292
TiltUp	166.42 ± 19.09	70	356
TiltDown	149.59 ± 18.48	80	196
Total	153.00 ± 20.73	22	356

Figure 4.10 shows the box plot of both the calculated PEP and the reference PEP next to one another. In Figure 4.11 residual plots of the best-performing pipeline without Outlier Correction (left) and with AutReg (rigth) are displayed.

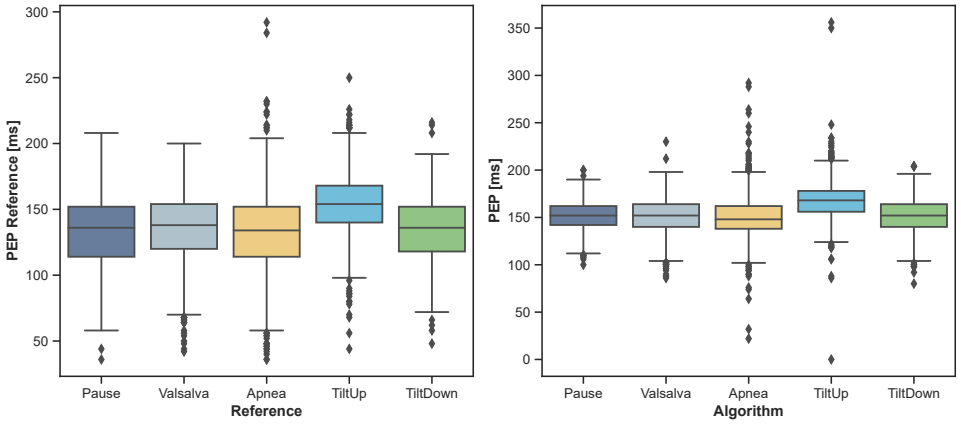

Figure 4.10: PEP duration per phase compared between best-performing pipeline of the GUARDIAN Dataset and the reference data

Discussion

For both datasets, calculating the PEP duration with the best-performing pipeline combination resulted in a higher mean PEP duration compared to the manually labeled reference (5.2% and

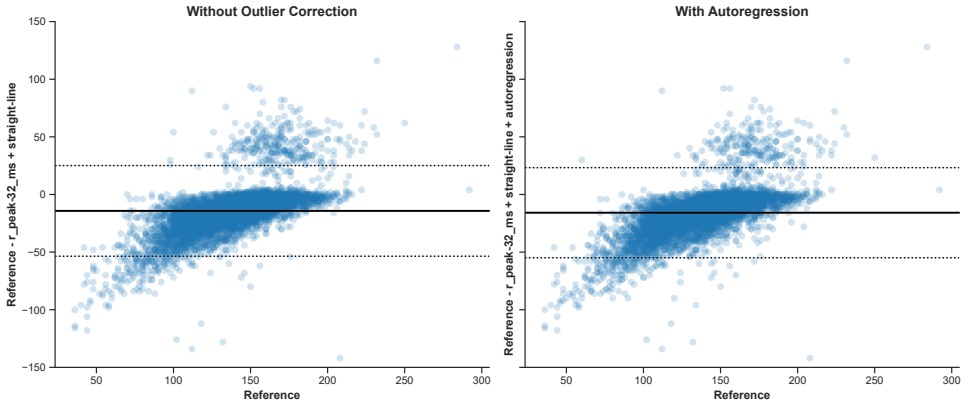

Figure 4.11: Residual plots of best-performing pipeline of the GUARDIAN Dataset without Outlier Correction (left) and with AutReg (right)

10.23%, respectively). This observation is also evident from the negative values of the ME of both algorithm combinations. Since the ME is produced by subtracting the value calculated by the algorithms from the actual value, the negative mean indicates a consistent trend of overestimating the PEP duration in the computed data.

This overestimation can also be observed in the box plots (Figures4.8 and 4.10, respectively), where the medians of the individual phases are noticeably higher compared to the reference plots, indicating longer PEP duration calculated by the algorithms. Furthermore, the box plots show that the relative differences across the individual phases are preserved. Regarding the TSST Dataset, it is notable that the difference between the two conditions, TSST and f-TSST, is also evident in the PEP times determined by the algorithm combinations.

Figures 4.9 and 4.11 display residual plots for the best-performing algorithm combinations of the two datasets. Notably, both residual plots exhibit a negative offset, again representing the overestimation in PEP duration.

Overall, it can be concluded that the overestimation by the algorithm combinations is consistently observed across the different phases of both datasets (Except for the *Pause_5* phases of both conditions in the TSST dataset). Consequently, the variation between the individual phases of the datasets remains accurately represented. For many applications, the relative change in PEP duration between different scenarios is more crucial, as the total duration can vary strongly between individuals. For instance, the change in PEP duration for an individual across various situations can provide insights into the level of stress experienced by that person [New79; Ber04]. Increased stress typically results in a reduced PEP duration. The preservation of the difference in PEP

duration between the two phases of the TSST dataset demonstrates that the algorithm-determined PEP is capable of reflecting the stress level.

These observations are also visible for the combination of RP-40, SecDer, and AutReg for the TSST Dataset. Visually, it is noticeable in the residual plot (Figure 4.9) that some points form a horizontal line at the upper left edge, which is not present without AutReg. This phenomenon can be attributed to the Outlier Correction process, where R-peaks were designated as B-points if no other points could be identified. Upon examining the residual plots for the best-performing pipeline of the GUARDIAN Dataset, it is evident that this line is not visible in the plot with AutReg. This can be explained by the fact, that when combining Algorithm that subtracts 32 ms from the R-Peak to identify the start point (RP-32) and StrLin algorithms with AutReg, PEP calculation was possible for only one additional cardiac cycle compared to the pipeline without Outlier Correction. Consequently, there aren't any points could build this line.

Given that the normal PEP duration ranges from 70 ms to 175 ms, the mean values from all discussed combinations for both datasets fall within the normal PEP range [Árb17]. Based on the previously discussed results, it is surprising that for both datasets, a combination using the RP algorithm for detecting the PEP start point ranks high in terms of performance, even though the RP is showing the highest MAE among all Q-Wave onset algorithms. For the TSST Dataset, this algorithm, in combination with the ThirDer method and using Intpol for Outlier Correction, results in an MAE of 18.33 ms. For the GUARDIAN Dataset, the combination of RP with StrLin and Intpol has a slightly higher MAE of 21.56 ms compared to the best-performing algorithm. The ME of this combination is 15.61 ms. An explanation for this can be found in the fact that both the RP algorithm and the ThirDer, as well as the StrLin algorithm tend to place the respective points on average later than the reference values. Because both the start and end points are consistently set too late, this offsets each other in terms of the distance between them, thereby ensuring that the PEP duration remains accurate. Figure 4.12 shows an example cardiac cycle from GUARDIAN Dataset, where the combination of RP and StrLin algorithm with Intpol for Outlier Correction achieved a deviation of 0 ms between the calculated PEP time and the reference. This clearly indicates that both algorithms did not detect the point labeled as the reference. Both the start and end points are located later than their respective references. However, due to the shift of both points, the final PEP duration remains the same at 118 ms.

For both datasets, the combination of the best-performing pipelines does not correspond to the algorithms that achieved the best results in the individual determination of the start and end points. This can also be attributed to the fact that the two algorithms used in the pipeline do not accurately determine the points, but their offsets compensate well when calculating the PEP duration.

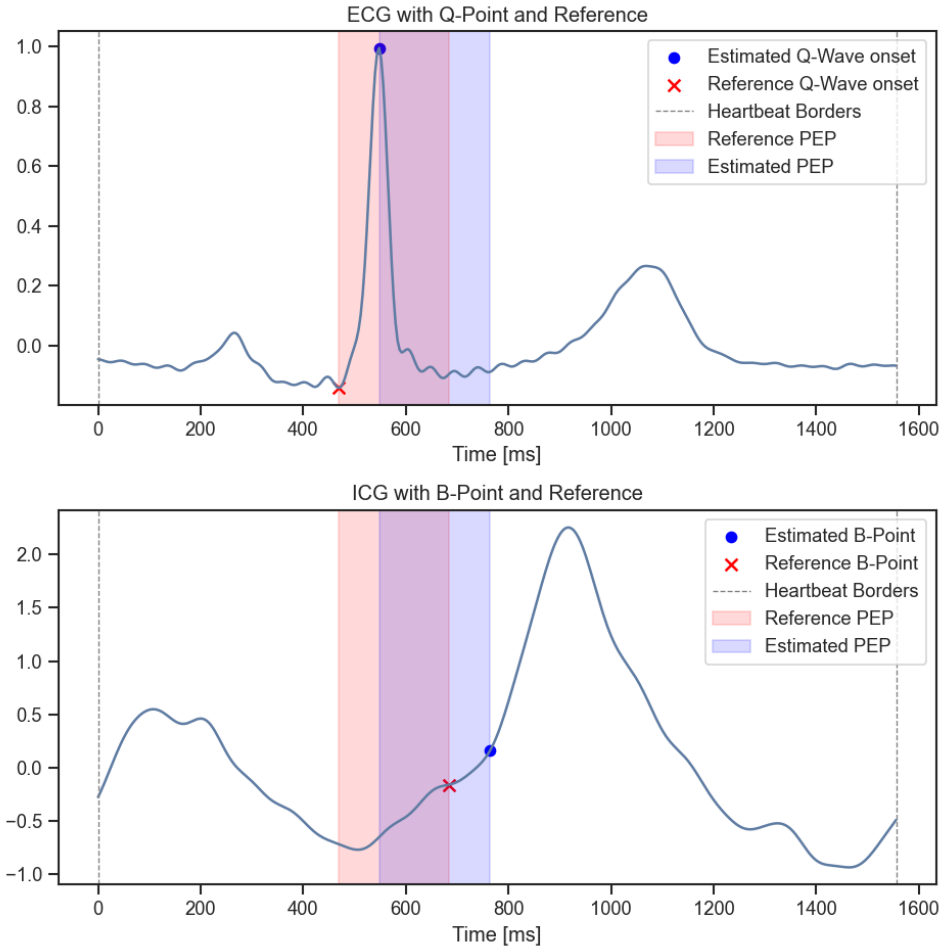

Figure 4.12: Detected Points with RP and StrLin algorithm for the GUARDIAN Dataset

4.3 General Discussions & Limitations

During the manual labeling process, heartbeats for which either one of the points could not be clearly identified were marked as artefacts and excluded from the evaluation. The approach aimed to prevent errors in automatic detection caused by strong signal disturbances from affecting the results, Since no reliable conclusions could have been drawn from this data.

The investigation of the pipeline performances showed that for locating the Q-Wave onset, the QP algorithm achieved the best results for both datasets. For determining the B-Point, the StrLin method presented by Drost et al. proved to be the best [Dro22]. When determining the best pipeline for PEP calculation, the results varied between the two datasets. For the TSST Dataset, the combination of the RP-40 and the SecDer algorithm emerged as the best. For the GUARDIAN Dataset, it was the RP-32 with the StrLin method. The difference in algorithm performances

between the two datasets highlights the importance of benchmarking to achieve the best possible results for specific data and therefore underscores the motivation of this thesis.

The two used datasets were labeled by two different individuals. Although both individuals adhered to the same guidelines and standards for labeling, it is important to acknowledge that achieving a 100% match between the labels of the two datasets cannot be guaranteed. Differences in interpretation, subtle variations in judgment, and human error are inherent in manual labeling processes. In the future, it may be beneficial to jointly label data from two individuals rather than labeling them individually. This approach would allow for the assessment of interrater reliability and enable consensus decisions in cases where points of uncertainty arise, to get two opinions which point is most accurate.

Due to the absence of Q-Waves fore some participants, the onset of the R-Wave was chosen as the starting point instead of the Q-Wave onset, following the suggestion of Berntson et al. [Ber04]. While this choice facilitates the comparability of PEP values across all participants it causes deviation from the precise physiological definition of PEP, leading to inaccuracies in its duration as defined. As a result, the PEP values obtained in this study are constrained in their ability to draw conclusions based on the absolute PEP values. However, for the evaluation of algorithm performance conducted in this study, these deviations are unlikely to significantly impact results. This is because the algorithms were evaluated based on their relative performance in detecting and measuring PEP durations, which remains consistent regardless of the chosen starting point.

Chapter 5

Conclusion & Outlook

The goal of this bachelor thesis was to present the first systematic benchmark for PEP extraction algorithms. For this purpose a total of $11{,}768$ cardiac cycles composed of two different datasets, that were conducted using individual study protocols and measurement systems, were manually labeled to build a basis of PEP reference data on which algorithms for automatic point detection can be tested. In one study, participants underwent the TSST and the f-TSST to induce acute psychosocial stress and a concurrent control condition, while the other part of the data was obtained from the execution of a TTT

Within this thesis, a total of 72 comprehensive data science pipelines consisting of algorithms to detect the start and end point of the PEP. All pipelines were tested on the two datasets and compared based on their accuracy in determining the PEP duration. In addition to the results for the determined PEP duration, the Q-Wave onset and B-Point algorithms were also individually compared regarding their ability to pinpoint the respective points.

Although no single combination of algorithms could be identified as the best-performing, as different results were observed for each dataset, some similarities between the datasets were determined and valuable insights were gained.

Given that the same algorithms were identified for both datasets as those providing results with the least deviation from the reference for determining the Q-Wave onset and B-Points individually, it would be beneficial to investigate whether this consistency holds true across additional datasets. This validation should be an integral part of future research to ensure the robustness and reliability of these algorithms. Similarly, the results regarding the worst-performing pipeline were consistent across both datasets. Further validation of this observation through evaluation with additional datasets could also be beneficial.

Moreover, it would be interesting to investigate the limited success achieved by using Outlier Correction concerning the MAE in more detail. Even though the Outlier Correction algorithms universally increased the number of heartbeats from which a PEP duration could be determined, it was noted that this also correlated with a higher MAE. As previously mentioned, the slightly higher MAE could alternatively indicate that the points newly identified through Outlier Correction are determined with high accuracy, thereby suggesting that the higher MAE values do not necessarily correlate with poorer performance. However, it would be interesting to examine if using longer signals, especially with the autoregression method, could potentially lead to better results.

To achieve meaningful and reliable results when benchmarking data, it is crucial to provide a comprehensive spectrum of data on which the algorithms can be tested. This data should include a variety of situations and measurement systems to ensure the algorithm's performance is robust across different conditions. Therefore, expanding the foundation established in this work with additional datasets is useful for creating a solid benchmark for PEP computing algorithms. Conclusively, this thesis provides a valuable first step towards the objective benchmarking of PEP extraction algorithms. In the spirit of open science, the implementation framework has been designed to allow the easy inclusion of additional datasets, thereby fostering continuous enhancement and extension to create a robust and reliable benchmark in this field, enabling it to exploit the potential of the PEP in clinical and research application.

List of Figures

3.1 Setup of the TSST for standing (a) and sitting (b) group 11
3.2 Temporal sequence of the (f-) TSST 11
3.3 Schematic representation of the setup of a TTT 14
3.4 Overview of significant ECG waveform 15
3.5 Overview electrode placement for ICG measurement 16
3.6 Overview of significant ICG waveform 17
3.7 Overview of Approach including algorithms used for PEP computation 18
3.8 ECG and ICG signal from the GUARDIAN Dataset before and after preprocessing 19
3.9 Labeled Q-Wave onsets using the QP algorithm 20
3.10 [Labeled Q-Wave onsets using the RP algorithm 20
3.11 Labeled Q-Wave onsets using the RP-40 algorithm 21
3.12 Labeled B-Points using the SecDer algorithm 22
3.13 Labeled B-Points using the ThirDer algorithm 23
3.14 Labeled B-Points using the StrLin algorithm 23
3.15 Labeled B-Points using the MultCon algorithm 24
3.16 Example of labeled ECG signal in the *MaD GUI* package 27
3.17 Example of labeled ICG signal in the *MaD GUI* package 29

4.1 Box plot of the reference PEP data of the TSST Dataset divided into conditions and phases 32
4.2 Box plot of the reference PEP data of the GUARDIAN Dataset divided into phases 33
4.3 Residual plots of the Q-Wave onset algorithms for the TSST Dataset 41
4.4 Residual plots of the Q-Wave onset algorithms for GUARDIAN Dataset 42
4.5 Selected ECG signal period of participant GDN0009 (top) and GDN0030 (bottom) from the GUARDIAN Dataset 44
4.6 Residual plots of the B-Point algorithms for the TSST Dataset 47

4.7 Residual plots of the B-Point algorithms for the GUARDIAN Dataset 49
4.8 PEP duration per phase compared between best-performing pipeline of TSST Dataset and the reference data 53
4.9 Residual plots of best-performing pipeline of the TSST Dataset without Outlier Correction (left) and with AutReg (right) 53
4.10 PEP duration per phase compared between best-performing pipeline of the GUARDIAN Dataset and the reference data 54
4.11 Residual plots of best-performing pipeline of the GUARDIAN Dataset without Outlier Correction (left) and with AutReg (right) 55
4.12 Detected Points with RP and StrLin algorithm for the GUARDIAN Dataset . . . 57

B.1 Residual plots of the Q-Wave onset algorithms for the TSST Dataset divided into participants 79
B.2 Residual plots of the Q-Wave onset algorithms for the TSST Dataset divided into phases 80
B.3 Residual plots of the Q-Wave onset algorithms for the GUARDIAN Dataset divided into participant 81
B.4 Residual plots of the Q-Wave onset algorithms for the GUARDIAN Dataset divided into phases 82
B.5 Residual plots of the B-Point algorithms for the TSST Dataset divided into participants 83
B.6 Residual plots of the B-Point algorithms for the TSST Dataset divided into phases 84
B.7 Residual plots of the B-Point algorithms for the GUARDIAN Dataset divided into participant 85
B.8 Residual plots of the B-Point algorithms for the GUARDIAN Dataset divided into phases 86
B.9 PEP duration per phase compared between best-performing pipeline with AutReg for the TSST Dataset and the reference data 87

List of Tables

3.1 Demographic and anthropometric data of the study participants for the TSST Dataset 10
3.2 Mean duration recordings per phase divided into conditions for the TSST Dataset 12
3.3 Demographic and anthropometric data of the study participants for the GUARDIAN Dataset . . . 13
3.4 Mean duration of recordings and quantity of participants per phase for the GUARDIAN Dataset . . . 14
3.5 Overview of abbreviations of the different Q-Wave onset algorithms, B-Point algorithms, and Outlier Correction methods . . . 25
3.6 Amount of manually labeled cardiac cycles per phase . . . 28

4.1 Reference PEP of the TSST Dataset divided into condition and phases . . . 32
4.2 Reference PEP data of the GUARDIAN Dataset divided into phases . . . 33
4.3 Overview of the performance of the 36 PEP computation pipelines using the QP, RP, and RP-40 as Q-Wave onset algorithms for the TSST Dataset. The best-performing pipeline with respect to the lowest MAE is highlighted in **bold** . . . 34
4.4 Overview of the performance of the 36 PEP computation pipelines using the QP, RP, and RP-40 as Q-Wave onset algorithms for the GUARDIAN Dataset . . . 36
4.5 Performance of the Q-Wave onset algorithms for the TSST Dataset. The best-performing algorithm with respect to the lowest MAE is highlighted in **bold** . . . 40
4.6 Performance of the Q-Wave onset algorithms for the GUARDIAN Dataset. The best-performing algorithm with respect to the lowest MAE is highlighted in **bold** 42
4.7 Performance of the B-Point algorithms for the TSST Dataset. The best-performing algorithm with respect to the lowest MAE is highlighted in **bold** . . . 46
4.8 Performance of the B-Point algorithms for the GUARDIAN Dataset. The best-performing algorithm with respect to the lowest MAE is highlighted in **bold** . . . 48

4.9 PEP duration calculated by best-performing pipeline for the TSST Dataset divided into conditions and phases 52
4.10 PEP duration calculated by best-performing pipeline for the GUARDIAN Dataset divided into phases 54

A.1 Overview of the performance of the 36 PEP computation pipelines using the RP-32, Algorithm that subtracts 34 ms from the R-Peak to identify the start point (RP-34), and Algorithm that subtracts 36 ms from the R-Peak to identify the start point (RP-36) as Q-Wave onset algorithms for the TSST Dataset 75
A.2 Overview of the performance of the 36 PEP computation pipelines using the RP-32, RP-34, and RP-36 as Q-Wave onset algorithms for the GUARDIAN Dataset. The best-performing pipeline with respect to the lowest MAE is highlighted in **bold** . 77
A.3 PEP of RP-40 and SecDer with AutReg for the TSST Dataset divided into condition and phases 78

Bibliography

[Aka69] Hirotugu Akaike. “Fitting autoregressive models for prediction”. en. In: *Annals of the Institute of Statistical Mathematics* 21.1 (Dec. 1969), pp. 243–247. ISSN: 1572-9052. DOI: 10.1007/BF02532251. URL: https://doi.org/10.1007/BF02532251 (visited on 05/15/2024).

[AL-15] Hussain AL-Ziarjawey. “Heart Rate Monitoring and PQRST Detection Based on Graphical User Interface with Matlab”. In: *International Journal of Information and Electronics Engineering* (2015). ISSN: 20103719. DOI: 10.7763/IJIEE.2015.V5.550. URL: http://www.ijiee.org/index.php?m=content&c=index&a=show&catid=52&id=602 (visited on 05/09/2024).

[All17] Andrew P. Allen, Paul J. Kennedy, Samantha Dockray, John F. Cryan, Timothy G. Dinan, and Gerard Clarke. “The Trier Social Stress Test: Principles and practice”. en. In: *Neurobiology of Stress* 6 (Feb. 2017), pp. 113–126. ISSN: 23522895. DOI: 10.1016/j.ynstr.2016.11.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S2352289516300224 (visited on 05/09/2024).

[Árb17] Javier Rodríguez Árbol, Pandelis Perakakis, Alba Garrido, José Luis Mata, M. Carmen Fernández-Santaella, and Jaime Vila. “Mathematical detection of aortic valve opening (B point) in impedance cardiography: A comparison of three popular algorithms”. en. In: *Psychophysiology* 54.3 (2017). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/psyp.12799, pp. 350–357. ISSN: 1469-8986. DOI: 10.1111/psyp.12799. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/psyp.12799 (visited on 05/11/2024).

[Ban24] Bangkok Heart Hospital. *Tilt Table Test*. 2024. URL: https://www.bangkokhearthospital.com/en/medical-service/tilt-table-test (visited on 06/21/2024).

[Bec06] Daniel E Becker. “Fundamentals of Electrocardiography Interpretation”. In: *Anesthesia Progress* 53.2 (2006), pp. 53–64. ISSN: 0003-3006. DOI: 10.2344/0003-3006(2006)

53[53:FOEI]2.0.CO;2. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1614214/ (visited on 05/09/2024).

[Ber04] Gary G. Berntson, David L. Lozano, Yun-Ju Chen, and John T. Cacioppo. "Where to Q in PEP". eng. In: *Psychophysiology* 41.2 (Mar. 2004), pp. 333–337. ISSN: 0048-5772. DOI: 10.1111/j.1469-8986.2004.00156.x.

[BIO24] BIOPAC Systems Inc. *MP160 Starter Systems*. 2024. URL: https://www.biopac.com/product-category/research/systems/mp150-starter-systems/ (visited on 06/29/2024).

[Bor12] Christine L. Borgman. "The conundrum of sharing research data". en. In: *Journal of the American Society for Information Science and Technology* 63.6 (2012). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.22634, pp. 1059–1078. ISSN: 1532-2890. DOI: 10.1002/asi.22634. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.22634 (visited on 05/10/2024).

[Bro04] Piet M. T. Broersen, Stijn de Waele, and Robert Bos. "Autoregressive spectral analysis when observations are missing". In: *Automatica* 40.9 (Sept. 2004), pp. 1495–1504. ISSN: 0005-1098. DOI: 10.1016/j.automatica.2004.04.011. URL: https://www.sciencedirect.com/science/article/pii/S0005109804001372 (visited on 06/30/2024).

[Cac94] John T. Cacioppo, Gary G. Berntson, Philip F. Binkley, Karen S. Quigley, Bert N. Uchino, and Annette Fieldstone. "Autonomic cardiac control. II. Noninvasive indices and basal response as revealed by autonomic blockades". en. In: *Psychophysiology* 31.6 (Nov. 1994), pp. 586–598. ISSN: 0048-5772, 1469-8986. DOI: 10.1111/j.1469-8986.1994.tb02351.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.1994.tb02351.x (visited on 05/09/2024).

[Chr04] Ivaylo I. Christov. "Real time electrocardiogram QRS detection using combined adaptive threshold". eng. In: *Biomedical Engineering Online* 3.1 (Aug. 2004), p. 28. ISSN: 1475-925X. DOI: 10.1186/1475-925X-3-28.

[CNS24] CNSystems Medizintechnik AG. *Task Force Monitor*. 2024. URL: https://www.cnsystems.com/de/produkte/task-force-monitor/ (visited on 06/29/2024).

[Deb93] T. T. Debski, Y. Zhang, J. R. Jennings, and T. W. Kamarck. "Stability of cardiac impedance measures: aortic opening (B-point) detection and scoring". eng. In: *Biological Psychology* 36.1-2 (Aug. 1993), pp. 63–74. ISSN: 0301-0511. DOI: 10.1016/0301-0511(93)90081-i.

[Dou24] Steven Douedi and Hani Douedi. “P wave”. eng. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2024. URL: http://www.ncbi.nlm.nih.gov/books/NBK551635/ (visited on 05/09/2024).

[Dro22] L. Drost, J. B. Finke, J. Port, and H. Schächinger. “Comparison of TWA and PEP as indices of α2- and ß-adrenergic activation”. eng. In: *Psychopharmacology* 239.7 (July 2022), pp. 2277–2288. ISSN: 1432-2072. DOI: 10.1007/s00213-022-06114-8.

[Ein12] Wilhelm Einthoven. “THE DIFFERENT FORMS OF THE HUMAN ELECTROCARDIOGRAM AND THEIR SIGNIFICATION.” In: *The Lancet*. Originally published as Volume 1, Issue 4622 179.4622 (Mar. 1912), pp. 853–861. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(00)50560-1. URL: https://www.sciencedirect.com/science/article/pii/S0140673600505601 (visited on 06/22/2024).

[Emp23] EmpkinS. *Website of the CRC 1483 EmpkinS › EmpkinS. de*. 2023. URL: https://www.empkins.de/ (visited on 06/06/2024).

[Erm12] V. Ermishkin, V. Kolesnikov, Elena Lukoshkova, V. Mokh, R. Sonina, N. Dupik, and S. Boitsov. “Variable impedance cardiography waveforms: How to evaluate the preejection period more accurately”. In: *Journal of Physics Conference Series* 407 (Dec. 2012), p. 2016. DOI: 10.1088/1742-6596/407/1/012016.

[For19] Mohamad Forouzanfar, Fiona C. Baker, Ian M. Colrain, Aimée Goldstone, and Massimiliano de Zambotti. “Automatic analysis of preejection period during sleep using impedance cardiogram”. In: *Psychophysiology* 56.7 (July 2019), e13355. ISSN: 0048-5772. DOI: 10.1111/psyp.13355. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6824194/ (visited on 05/10/2024).

[Fu18] Polly Fu, Carolyn J. Gibson, Wendy Berry Mendes, Michael Schembri, and Alison J. Huang. “Anxiety, Depressive Symptoms, and Cardiac Autonomic Function in Perimenopausal and Postmenopausal Women with Hot Flashes: A Brief Report”. In: *Menopause (New York, N.Y.)* 25.12 (Dec. 2018), pp. 1470–1475. ISSN: 1072-3714. DOI: 10.1097/GME.0000000000001153. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6265057/ (visited on 06/19/2024).

[Gen23] Gentemann, Chelle. “Why NASA and federal agencies are declaring this the Year of Open Science”. en. In: *Nature* 613.7943 (Jan. 2023). Bandiera_abtest: a Cg_type: World View Publisher: Nature Publishing Group Subject_term: Society, Research management, Lab life, pp. 217–217. DOI: 10.1038/d41586-023-00019-y. URL: https://www.nature.com/articles/d41586-023-00019-y (visited on 05/31/2024).

[Gib19] Christopher H. Gibbons. “Basics of autonomic nervous system function”. en. In: *Handbook of Clinical Neurology*. Vol. 160. Elsevier, 2019, pp. 407–418. ISBN: 978-0-444-64032-1. DOI: 10.1016/B978-0-444-64032-1.00027-8. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780444640321000278 (visited on 05/09/2024).

[Gir05] M P Girish, Mohit Dayal Gupta, Saibal Mukhopadhyay, Jamal Yusuf, Sunil Roy T N, and Vijay Trehan. “U wave: an Important Noninvasive Electrocardiographic Diagnostic Marker”. In: *Indian Pacing and Electrophysiology Journal* 5.1 (Jan. 2005), pp. 63–65. ISSN: 0972-6292. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1502069/ (visited on 05/09/2024).

[Hot05] Torsten Hothorn, Friedrich Leisch, Achim Zeileis, and Kurt Hornik. “The Design and Analysis of Benchmark Experiments”. In: *Journal of Computational and Graphical Statistics* 14.3 (2005). Publisher: [American Statistical Association, Taylor & Francis, Ltd., Institute of Mathematical Statistics, Interface Foundation of America], pp. 675–699. ISSN: 1061-8600. URL: https://www.jstor.org/stable/27594139 (visited on 05/10/2024).

[Koe12] S. Koelstra, C. Muhl, M. Soleymani, Jong-Seok Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras. “DEAP: A Database for Emotion Analysis ;Using Physiological Signals”. en. In: *IEEE Transactions on Affective Computing* 3.1 (Jan. 2012), pp. 18–31. ISSN: 1949-3045. DOI: 10.1109/T-AFFC.2011.15. URL: http://ieeexplore.ieee.org/document/5871728/ (visited on 06/18/2024).

[Küd23] Arne Küderle, Robert Richer, Raul Sîmpetru, and Bjoern Eskofier. “tpcp: Tiny Pipelines for Complex Problems - A set of framework independent helpers for algorithms development and evaluation”. In: *Journal of Open Source Software* 8 (Feb. 2023), p. 4953. DOI: 10.21105/joss.04953.

[Küd24] Arne Küderle, Martin Ullrich, Nils Roth, Malte Ollenschläger, Alzhraa A. Ibrahim, Hamid Moradi, Robert Richer, Ann-Kristin Seifer, Matthias Zürl, Raul C. Sîmpetru, Liv Herzer, Dominik Prossel, Felix Kluge, and Bjoern M. Eskofier. “Gaitmap—An Open Ecosystem for IMU-Based Human Gait Analysis and Algorithm Benchmarking”. In: *IEEE Open Journal of Engineering in Medicine and Biology* 5 (2024), pp. 163–172. ISSN: 2644-1276. DOI: 10.1109/OJEMB.2024.3356791. URL: https://ieeexplore.ieee.org/document/10411039/ (visited on 06/06/2024).

[Kur24] Miriam Kurz, Nicolas Rohleder, Luca Abel, Veronika Ringgold, Lena Gmelch (neé Schindler), Felicitas Hauck, and Robert Richer. “The effect of acute psychosocial

stress on micro- and macroscopic body movements using Empathokinaestetic Sensors (EmpkinS D03 AP3)". en-us. In: (Apr. 2024). Publisher: OSF. DOI: 10.17605/OSF.IO/YC5DJ. URL: https://osf.io/yc5dj (visited on 06/18/2024).

[Lab24] UbiComp Lab. *rPPG Toolbox*. 2024. URL: https://github.com/ubicomplab/rPPG-Toolbox (visited on 06/18/2024).

[Lab70] Zuhdi Lababidi, D. A. Ehmke, Robert E. Durnin, Paul E. Leaverton, and Ronald M. Lauer. "The First Derivative Thoracic Impedance Cardiogram". en. In: *Circulation* 41.4 (Apr. 1970), pp. 651–658. ISSN: 0009-7322, 1524-4539. DOI: 10.1161/01.CIR.41.4.651. URL: https://www.ahajournals.org/doi/10.1161/01.CIR.41.4.651 (visited on 05/09/2024).

[Lar86] Kevin T. Larkin and Alfred L. Kasprowicz. "Validation of a Simple Method of Assessing Cardiac Preejection Period: A Potential Index of Sympathetic Nervous System Activity". en. In: *Perceptual and Motor Skills* 63.1 (Aug. 1986), pp. 295–302. ISSN: 0031-5125, 1558-688X. DOI: 10.2466/pms.1986.63.1.295. URL: http://journals.sagepub.com/doi/10.2466/pms.1986.63.1.295 (visited on 05/09/2024).

[Lie13] René van Lien, Nienke M. Schutte, Jan H. Meijer, and Eco J. C. de Geus. "Estimated preejection period (PEP) based on the detection of the R-wave and dZ/dt-min peaks does not adequately reflect the actual PEP across a wide range of laboratory and ambulatory conditions". eng. In: *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology* 87.1 (Jan. 2013), pp. 60–69. ISSN: 1872-7697. DOI: 10.1016/j.ijpsycho.2012.11.001.

[Lüt05] Helmut Lütkepohl. *New Introduction to Multiple Time Series Analysis*. en. Berlin, Heidelberg: Springer, 2005. ISBN: 978-3-540-40172-8 978-3-540-27752-1. DOI: 10.1007/978-3-540-27752-1. URL: http://link.springer.com/10.1007/978-3-540-27752-1 (visited on 05/15/2024).

[Mak21] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. "NeuroKit2: A Python toolbox for neurophysiological signal processing". eng. In: *Behavior Research Methods* 53.4 (Aug. 2021), pp. 1689–1696. ISSN: 1554-3528. DOI: 10.3758/s13428-020-01516-y.

[Man18] Sofienne Mansouri, Tareq Alhadidi, Souhir Chabchoub, and Ridha Ben Salah. "Impedance cardiography: recent applications and developments". In: *Biomedical Research* 29.19 (2018). ISSN: 09761683. DOI: 10.4066/biomedicalresearch.29-17-3479. URL: http://

www.alliedacademies.org/articles/impedance-cardiography-recent-applications-and-developments-10909.html (visited on 05/09/2024).

[McC07] Laurie Kelly McCorry. "Physiology of the Autonomic Nervous System". In: *American Journal of Pharmaceutical Education* 71.4 (Aug. 2007), p. 78. ISSN: 0002-9459. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1959222/ (visited on 05/09/2024).

[McK10] Wes McKinney. "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56–61. DOI: 10.25080/Majora-92bf1922-00a.

[New79] David B. Newlin and Robert W. Levenson. "Pre-ejection Period: Measuring Beta-adrenergic Influences Upon the Heart". en. In: *Psychophysiology* 16.6 (Nov. 1979), pp. 546–552. ISSN: 0048-5772, 1469-8986. DOI: 10.1111/j.1469-8986.1979.tb01519.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.1979.tb01519.x (visited on 05/09/2024).

[Oll22] Malte Ollenschläger, Arne Küderle, Wolfgang Mehringer, Ann-Kristin Seifer, Jürgen Winkler, Heiko Gaßner, Felix Kluge, and Bjoern M. Eskofier. "MaD GUI: An Open-Source Python Package for Annotation and Analysis of Time-Series Data". In: *Sensors (Basel, Switzerland)* 22.15 (Aug. 2022), p. 5849. ISSN: 1424-8220. DOI: 10.3390/s22155849. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9371110/ (visited on 05/17/2024).

[Phy24a] PhysioNet. *PhysioBank Databases: ECG*. 2024. URL: https://physionet.org/content/?topic=ecg (visited on 06/14/2024).

[Phy24b] PhysioNet. *PhysioBank Databases: EEG*. 2024. URL: https://physionet.org/content/?topic=eeg (visited on 06/14/2024).

[Phy24c] PhysioNet. *PhysioBank Databases: Respiration*. 2024. URL: https://physionet.org/content/?topic=respiration (visited on 06/14/2024).

[Phy24d] PhysioNet. *PhysioNet: The research resource for complex physiologic signals*. 2024. URL: https://physionet.org/ (visited on 06/14/2024).

[Pil23] Niklas Pilz, Andreas Patzak, and Tomas L. Bothe. "The pre-ejection period is a highly stress dependent parameter of paramount importance for pulse-wave-velocity based applications". en. In: *Frontiers in Cardiovascular Medicine* 10 (Feb. 2023), p. 1138356.

ISSN: 2297-055X. DOI: 10.3389/fcvm.2023.1138356. URL: https://www.frontiersin.org/articles/10.3389/fcvm.2023.1138356/full (visited on 05/09/2024).

[Pre95] Lutz Prechelt. "Some notes on neural learning algorithm benchmarking". In: *Neurocomputing*. Control and Robotics, Part III 9.3 (Dec. 1995), pp. 343–347. ISSN: 0925-2312. DOI: 10.1016/0925-2312(95)00084-1. URL: https://www.sciencedirect.com/science/article/pii/0925231295000841 (visited on 05/10/2024).

[Pst16] L. Pstras, K. Thomaseth, J. Waniewski, I. Balzani, and F. Bellavere. "The Valsalva manoeuvre: physiology and clinical examples". en. In: *Acta Physiologica* 217.2 (2016). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/apha.12639, pp. 103–119. ISSN: 1748-1716. DOI: 10.1111/apha.12639. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/apha.12639 (visited on 05/30/2024).

[San13] Kulwinder S Sandhu, Pervez Khan, John Panting, and Sunil Nadar. "Tilt-table test: its role in modern practice". In: *Clinical Medicine* 13.3 (June 2013), pp. 227–232. ISSN: 1470-2118. DOI: 10.7861/clinmedicine.13-3-227. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5922663/ (visited on 05/09/2024).

[Sat24] Yasar Sattar and Lovely Chhabra. "Electrocardiogram". eng. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2024. URL: http://www.ncbi.nlm.nih.gov/books/NBK549803/ (visited on 05/09/2024).

[Sea10] Skipper Seabold and Josef Perktold. "Statsmodels: Econometric and Statistical Modeling with Python". In: Austin, Texas, 2010, pp. 92–96. DOI: 10.25080/Majora-92bf1922-011. URL: https://conference.scipy.org/proceedings/scipy2010/seabold.html (visited on 05/15/2024).

[See16] Mark D. Seery, Cheryl L. Kondrak, Lindsey Streamer, Thomas Saltsman, and Veronica M. Lamarche. "Preejection period can be calculated using R peak instead of Q". en. In: *Psychophysiology* 53.8 (Aug. 2016), pp. 1232–1240. ISSN: 0048-5772, 1469-8986. DOI: 10.1111/psyp.12657. URL: https://onlinelibrary.wiley.com/doi/10.1111/psyp.12657 (visited on 05/09/2024).

[She90] Andrew Sherwood(Chair), Michael T. Allen, Jochen Fahrenberg, Robert M. Kelsey, William R. Lovallo, and Lorenz J.P. Van Doornen. "Methodological Guidelines for Impedance Cardiography". en. In: *Psychophysiology* 27.1 (Jan. 1990), pp. 1–23. ISSN: 0048-5772, 1469-8986. DOI: 10.1111/j.1469-8986.1990.tb02171.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.1990.tb02171.x (visited on 05/09/2024).

[Sri24] Shival Srivastav, Radia T. Jamil, and Roman Zeltser. "Valsalva Maneuver". eng. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2024. URL: http://www.ncbi.nlm.nih.gov/books/NBK537248/ (visited on 05/30/2024).

[Ste23] Ulla Sternemann. "Extraction of Pre-Ejection Period as Marker for Acute Psychosocial Stress from Wearable Sensors and Interferometry Radar". MA thesis. Machine Learning, Data Analytics Lab, Department Artificial Intelligence in Biomedical Engineering in Cooperation with Chair of Health Psychology (FAU), and Institute of High-Frequency Technology, Hamburg University of Technology (TUHH): Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Apr. 2023.

[Sto10] Victoria Stodden. *The Scientific Method in Practice: Reproducibility in the Computational Sciences*. en. SSRN Scholarly Paper. Rochester, NY, Feb. 2010. DOI: 10.2139/ssrn.1550193. URL: https://papers.ssrn.com/abstract=1550193 (visited on 05/10/2024).

[Str20] Nils Strodthoff, Patrick Wagner, Tobias Schaeffter, and Wojciech Samek. *Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL*. en. Apr. 2020. URL: https://arxiv.org/abs/2004.13701v1 (visited on 06/06/2024).

[Stü23] Sebastian Stühler. "Investigation of the Pre-Ejection Period as a Marker for Sympathetic Activity during Acute Psychosocial Stress". Bachelor's Thesis. Erlangen, Germany: Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), May 2023. URL: https://www.mad.tf.fau.de/files/2023/03/BA_MA_Thesis_master_final.pdf.

[Teo16] Nicholay Teodorovich and Moshe Swissa. "Tilt table test today - state of the art". en. In: *World Journal of Cardiology* 8.3 (2016), p. 277. ISSN: 1949-8462. DOI: 10.4330/wjc.v8.i3.277. URL: http://www.wjgnet.com/1949-8462/full/v8/i3/277.htm (visited on 05/09/2024).

[Ulb14] Mark Ulbrich, Jens Muehlsteff, Steffen Leonhardt, and Marian Walter. "Influence of physiological sources on the impedance cardiogram analyzed using 4D FEM simulations". In: *Physiological Measurement* 35 (June 2014), p. 1451. DOI: 10.1088/0967-3334/35/7/1451.

[Uni23] Universitätsklinikum Erlangen. *Guardian - Klinisch-experimentelle Forschung*. 2023. URL: https://www.palliativmedizin.uk-erlangen.de/forschung/klinisch-experimentelle-forschung/guardian/ (visited on 06/06/2024).

[Vir20] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. "SciPy 1.0: fundamental algorithms for scientific computing in Python". eng. In: *Nature Methods* 17.3 (Mar. 2020), pp. 261–272. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0686-2.

[Vol23] Vanessa Volz, Dani Irawan, Koen Van Der Blom, and Boris Naujoks. "Benchmarking". en. In: *Many-Criteria Optimization and Decision Analysis*. Ed. by Dimo Brockhoff, Michael Emmerich, Boris Naujoks, and Robin Purshouse. Series Title: Natural Computing Series. Cham: Springer International Publishing, 2023, pp. 149–179. ISBN: 978-3-031-25262-4 978-3-031-25263-1. DOI: 10.1007/978-3-031-25263-1_6. URL: https://link.springer.com/10.1007/978-3-031-25263-1_6 (visited on 05/10/2024).

[Wan06] David J. Wang and Stephen S. Gottlieb. "Impedance cardiography: More questions than answers". en. In: *Current Cardiology Reports* 8.3 (May 2006), pp. 180–186. ISSN: 1523-3782, 1534-3170. DOI: 10.1007/s11886-006-0031-0. URL: http://link.springer.com/10.1007/s11886-006-0031-0 (visited on 05/09/2024).

[Wan23] Kathryn Wantlin, Chenwei Wu, Shih-Cheng Huang, Oishi Banerjee, Farah Dadabhoy, Veeral Vipin Mehta, Ryan Wonhee Han, Fang Cao, Raja R. Narayan, Errol Colak, Adewole Adamson, Laura Heacock, Geoffrey H. Tison, Alex Tamkin, and Pranav Rajpurkar. *BenchMD: A Benchmark for Unified Learning on Medical Images and Sensors*. arXiv:2304.08486 [cs]. June 2023. DOI: 10.48550/arXiv.2304.08486. URL: http://arxiv.org/abs/2304.08486 (visited on 05/10/2024).

[Wil18] Christoph Will, Kilin Shi, Sven Schellenberger, Tobias Steigleder, Fabian Michler, Jonas Fuchs, Robert Weigel, Christoph Ostgathe, and Alexander Koelpin. "Radar-Based Heart Sound Detection". en. In: *Scientific Reports* 8.1 (July 2018), p. 11551. ISSN: 2045-2322. DOI: 10.1038/s41598-018-29984-5. URL: https://www.nature.com/articles/s41598-018-29984-5 (visited on 06/29/2024).

[Win16] Raimond L. Winslow, Stephen Granite, and Christian Jurado. "WaveformECG: A Platform for Visualizing, Annotating, and Analyzing ECG Data". In: *Computing in*

Science & Engineering 18.5 (Sept. 2016), pp. 36–46. ISSN: 1521-9615. DOI: 10.1109/MCSE.2016.91. URL: http://ieeexplore.ieee.org/document/7548995/ (visited on 05/09/2024).

[Xia24] Hanguang Xiao, Tianqi Liu, Yisha Sun, Yulin Li, Shiyi Zhao, and Alberto Avolio. "Remote photoplethysmography for heart rate measurement: A review". In: *Biomedical Signal Processing and Control* 88 (Feb. 2024), p. 105608. ISSN: 1746-8094. DOI: 10.1016/j.bspc.2023.105608. URL: https://www.sciencedirect.com/science/article/pii/S1746809423010418 (visited on 05/10/2024).

[Zys24] Dorota Zysko, Radia T. Jamil, and Arayamparambil C. Anilkumar. "Tilt Table". eng. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2024. URL: http://www.ncbi.nlm.nih.gov/books/NBK482320/ (visited on 05/09/2024).

Appendix A

Additional Tables

Table A.1: Overview of the performance of the 36 PEP computation pipelines using the RP-32, RP-34, and RP-36 as Q-Wave onset algorithms for the TSST Dataset

Q-Wave onset	B-Point	Correction	**MAE** [ms]	**ME** [ms]	**RE**	**#CC**
RP-32	MultCon	None	23.60 ± 23.77	20.19 ± 26.73	26.87% ± 25.21%	4,707
		AutReg	22.50 ± 22.38	17.83 ± 26.25	25.64% ± 24.32%	4,841
		Intpol	22.24 ± 22.25	17.17 ± 26.36	25.40 % ± 24.30%	4,841
	SecDer	None	17.17 ± 17.34	4.94 ± 23.89	21.30% ± 27.83%	4,617
		AutReg	17.55 ± 17.49	3.37 ± 24.55	22.94% ± 29.24%	5,000
		Intpol	17.78 ± 17.77	2.73 ± 24.99	23.34% ± 29.96%	5,000
	StrLin	None	15.71 ± 14.99	-10.30 ± 19.11	20.59% ± 21.94%	4,988
		AutReg	16.11 ± 15.14	-12.74 ± 18.06	21.58% ± 24.34%	4,995
		Intpol	16.29 ± 15.26	-13.07 ± 18.10	21.82% ± 24.59%	4,995
	ThirDer	None	24.01 ± 18.23	-19.78 ± 22.75	32.04% ± 31.05%	4,971
		AutReg	25.28 ± 18.63	-23.60 ± 20.71	34.25% ± 33.32%	4,992
		Intpol	25.57 ± 18.72	-24.02 ± 20.67	34.65% ± 33.52%	4,992

Q-Wave onset	B-Point	Correction	**MAE** [ms]	**ME** [ms]	**RE**	**#CC**
RP-34	MultCon	None	22.63 ± 23.53	18.38 ± 26.98	25.82% ± 24.91%	4,720
		AutReg	21.55 ± 22.05	15.94 ± 26.39	24.65% ± 24.05%	4,849
		Intpol	21.30 ± 21.90	15.26 ± 26.47	24.43% ± 24.05%	4,848
	SecDer	None	16.19 ± 17.81	2.94 ± 23.89	20.36% ± 28.78%	4,617
		AutReg	16.76 ± 17.99	1.37 ± 24.55	22.16% ± 30.17%	5,000
		Intpol	17.05 ± 18.29	0.73 ± 24.99	22.62% ± 30.93%	5,000
	StrLin	None	17.25 ± 14.85	-12.28 ± 19.16	22.61% ± 22.39%	4,989
		AutReg	17.69 ± 15.19	-14.74 ± 18.06	23.64% ± 24.96%	4,995
		Intpol	17.90 ± 15.36	-15.05 ± 18.16	23.92% ± 25.23%	4,996
	ThirDer	None	25.74 ± 18.22	-21.74 ± 22.84	34.18 % ± 31.58%	4,973
		AutReg	27.09 ± 18.72	-25.60 ± 20.71	36.51% ± 33.97%	4,992
		Intpol	27.39 ± 18.82	-26.02 ± 20.67	36.92% ± 34.18%	4,992
RP-36	MultCon	None	21.76 ± 23.05	16.48 ± 27.08	24.93% ± 24.47%	4,728
		AutReg	20.71 ± 21.53	13.97 ± 26.41	23.81% ± 23.65%	4,852
		Intpol	20.49 ± 21.41	13.29 ± 26.49	23.63% ± 23.70%	4,851
	SecDer	None	15.44 ± 18.26	0.94 ± 23.89	19.69% ± 29.69%	4,617
		AutReg	16.19 ± 18.47	-0.63 ± 24.55	21.65% ± 31.09%	5,000
		Intpol	16.52 ± 18.79	-1.27 ± 24.99	22.17% ± 31.88%	5,000
	StrLin	None	18.88 ± 14.65	-14.28 ± 19.16	24.72% ± 22.81%	4,989
		AutReg	19.38 ± 15.20	-16.74 ± 18.06	25.81% ± 25.55%	4,995
		Intpol	19.60 ± 15.37	-17.05 ± 18.16	26.10% ± 25.81%	4,996
	ThirDer	None	27.54 ± 18.19	-23.63 ± 23.04	36.43% ± 32.13%	4,979
		AutReg	28.95 ± 18.78	-27.60 ± 20.71	38.82% ± 34.59%	4,992
		Intpol	29.27 ± 18.87	-28.00 ± 20.71	39.26% ± 34.81%	4,993

Table A.2: Overview of the performance of the 36 PEP computation pipelines using the RP-32, RP-34, and RP-36 as Q-Wave onset algorithms for the GUARDIAN Dataset. The best-performing pipeline with respect to the lowest MAE is highlighted in **bold**

Q-Wave onset	B-Point	Correction	**MAE** [ms]	**ME** [ms]	**RE**	**#CC**
RP-32	MultCon	None	29.26 ± 36.36	22.82 ± 40.71	21.29% ± 25.24%	6,016
		AutReg	32.26 ± 38.55	24.93 ± 43.65	23.44% ± 26.90%	6,561
		Intpol	31.39 ± 38.04	23.39 ± 43.42	22.92% ± 26.81%	6,561
	SecDer	None	24.42 ± 24.42	-10.57 ± 32.88	19.86% ± 25.49%	6,752
		AutReg	24.73 ± 24.62	-12.39 ± 32.62	20.19% ± 25.85%	6,756
		Intpol	25.50 ± 24.95	-13.65 ± 32.96	20.91% ± 26.53%	6,756
	StrLin	**None**	**18.41 ± 16.32**	**-14.23 ± 20.07**	**15.67% ± 19.95%**	**6,754**
		AutReg	19.17 ± 16.74	-15.82 ± 19.94	16.43% ± 20.75%	6,755
		Intpol	19.46 ± 17.07	-16.27 ± 20.13	16.69% ± 21.04%	6,755
	ThirDer	None	24.77 ± 22.06	-14.24 ± 29.96	20.55% ± 24.50%	6,754
		AutReg	24.54 ± 21.43	-18.64 ± 26.72	20.76% ± 25.06%	6,754
		Intpol	25.24 ± 21.69	-20.13 ± 26.50	21.41% ± 25.58%	6,754
RP-34	MultCon	None	28.66 ± 36.14	21.06 ± 41.03	20.93% ± 25.15%	6,027
		AutReg	31.66 ± 38.24	23.14 ± 43.92	23.09% ± 26.76%	6,572
		Intpol	30.83 ± 37.75	21.60 ± 43.69	22.61% ± 26.71%	6,572
	SecDer	None	24.19 ± 25.57	-12.57 ± 32.88	19.79% ± 26.41%	6,752
		AutReg	24.66 ± 25.75	-14.39 ± 32.62	20.26% ± 26.76%	6,756
		Intpol	25.50 ± 26.10	-15.65 ± 32.96	21.04% ± 27.45%	6,756
	StrLin	None	20.17 ± 16.19	-16.21 ± 20.15	17.04% ± 20.27%	6,755
		AutReg	20.94 ± 16.70	-17.79 ± 20.02	17.81% ± 21.10%	6,756
		Intpol	21.24 ± 17.04	-18.25 ± 20.21	18.08% ± 21.39%	6,756
	ThirDer	None	26.16 ± 21.83	-16.24 ± 29.96	21.63% ± 24.74%	6,754
		AutReg	26.11 ± 21.40	-20.64 ± 26.72	21.98% ± 25.40%	6,754
		Intpol	26.86 ± 21.69	-22.13 ± 26.54	22.66% ± 25.94%	6,754

Q-Wave onset	B-Point	Correction	MAE [ms]	ME [ms]	RE	#CC
RP-36	MultCon	None	28.33 ± 35.66	19.24 ± 41.28	20.78% ± 24.93%	6,036
		AutReg	31.26 ± 37.67	21.28 ± 44.09	22.89% ± 26.47%	6,579
		Intpol	30.47 ± 37.22	19.73 ± 43.87	22.44% ± 26.47%	6,579
	SecDer	None	24.13 ± 26.66	-14.57 ± 32.88	19.86% ± 27.29%	6,752
		AutReg	24.77 ± 26.82	-16.39 ± 32.62	20.46% ± 27.64%	6,756
		Intpol	25.68 ± 27.17	-17.65 ± 32.96	21.29% ± 28.34%	6,756
	StrLin	None	21.95 ± 15.99	-18.21 ± 20.15	18.42% ± 20.55%	6,755
		AutReg	22.74 ± 16.60	-19.79 ± 20.02	19.20% ± 21.42%	6,756
		Intpol	23.06 ± 16.94	-20.25 ± 20.21	19.48% ± 21.71%	6,756
	ThirDer	None	27.64 ± 21.59	-18.24 ± 29.96	22.77% ± 24.97%	6,754
		AutReg	27.76 ± 21.34	-22.64 ± 26.72	23.25% ± 25.73%	6,754
		Intpol	28.56 ± 21.66	-24.13 ± 26.50	23.97% ± 26.28%	6,754

Table A.3: PEP of RP-40 and SecDer with AutReg for the TSST Dataset divided into condition and phases

		Mean ± Std	Min	Max
TSST	**Prep**	86.64 ± 31.28	40	337
	Pause 1	77.35 ± 24.64	40	159
	Talk	91.68± 34.33	40	242
	Math	87.44 ± 24.03	40	179
	Pause 5	99.05 ± 23.25	40	150
	Total	88.54± 29.63	40	337
f-TSST	**Prep**	92.00 ± 30.11	40	168
	Pause 1	79.99 ± 26.12	40	154
	Talk	101.12 ± 30.85	40	306
	Math	102.34 ± 29.05	40	247
	Pause 5	112.98 ± 21.36	67	155
	Total	98.09 ± 30.15	40	306

Appendix B

Additional Figures

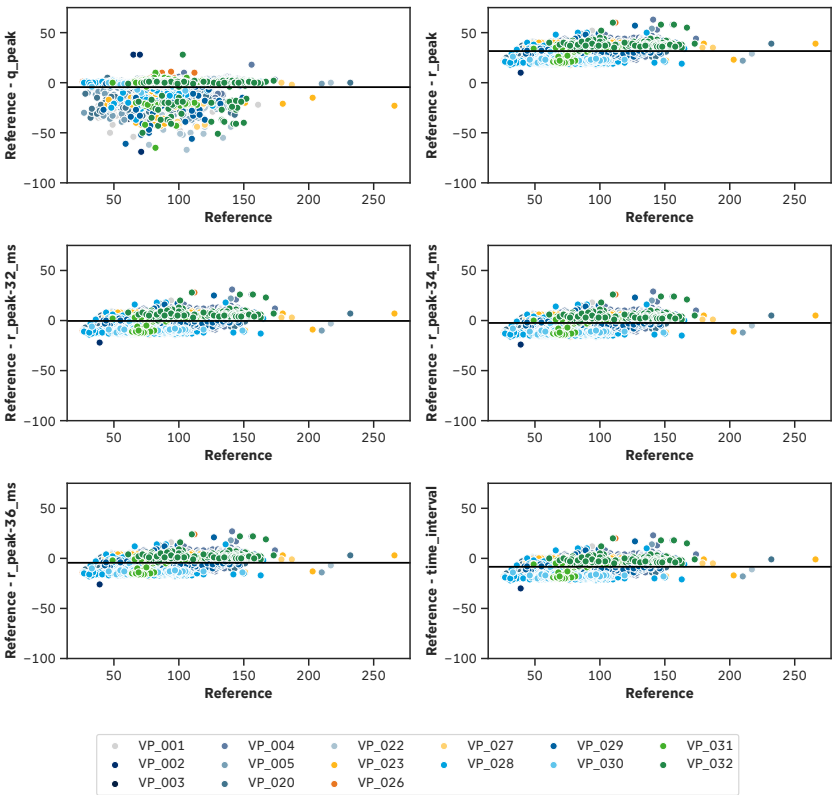

Figure B.1: Residual plots of the Q-Wave onset algorithms for the TSST Dataset divided into participants

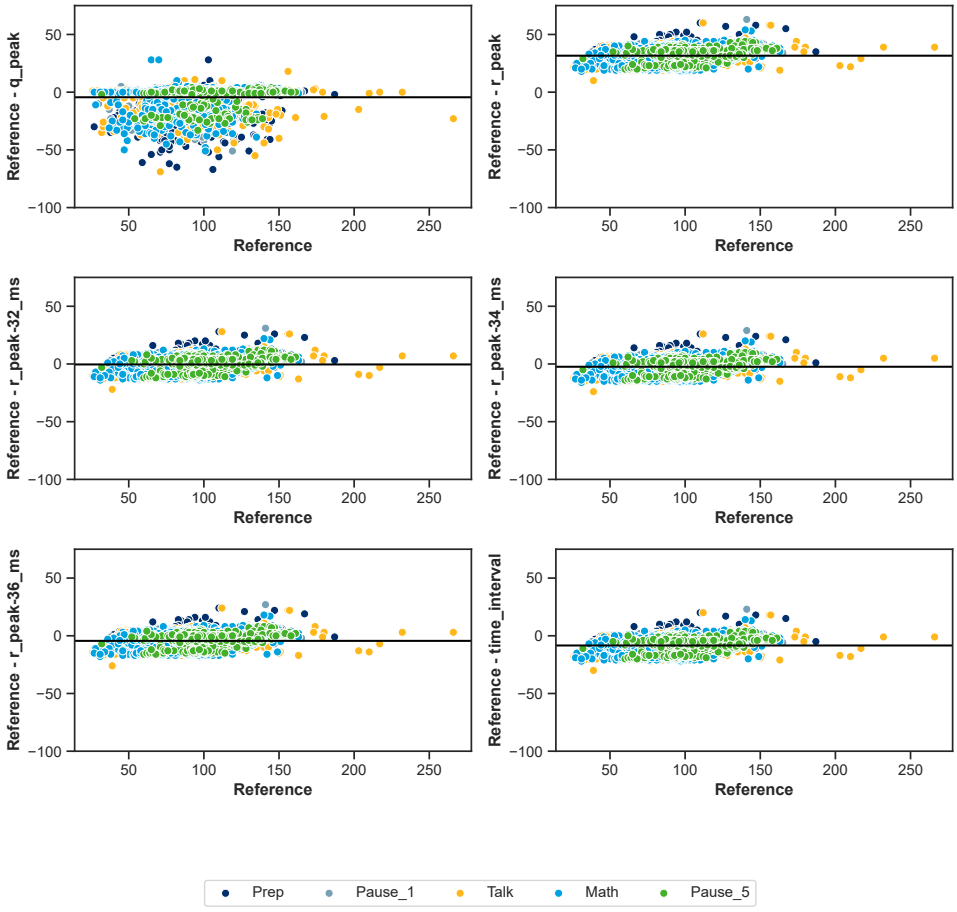

Figure B.2: Residual plots of the Q-Wave onset algorithms for the TSST Dataset divided into phases

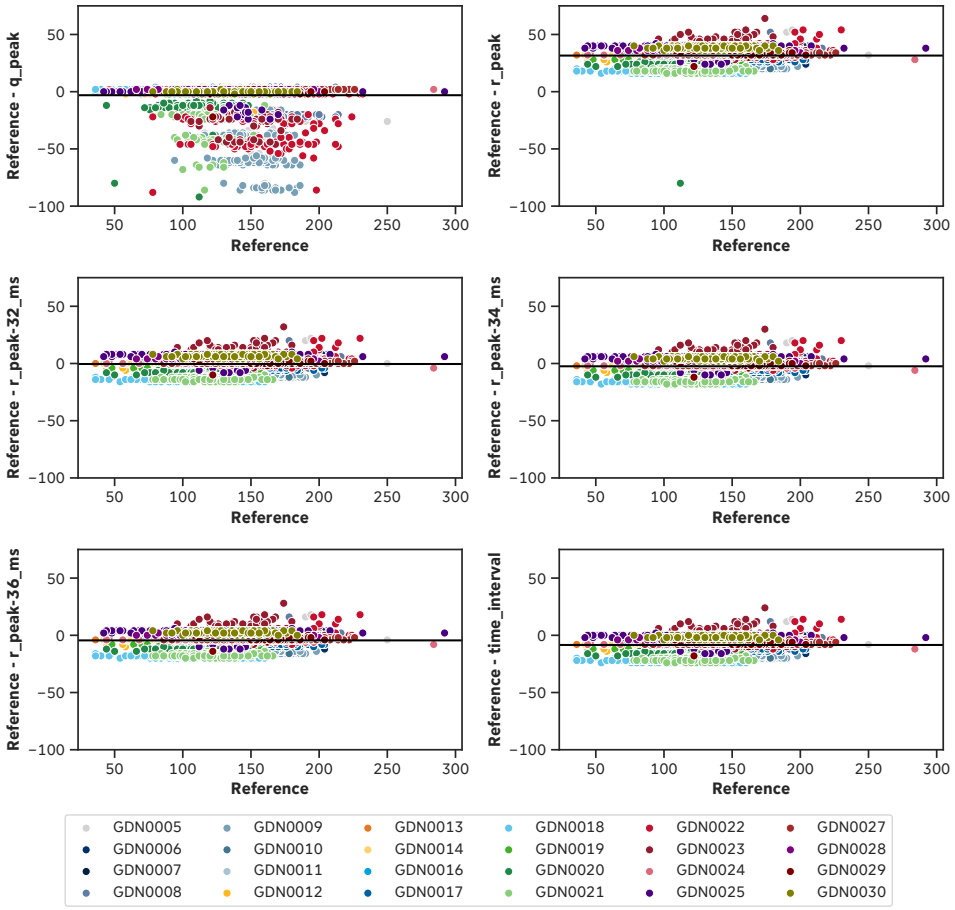

Figure B.3: Residual plots of the Q-Wave onset algorithms for the GUARDIAN Dataset divided into participants

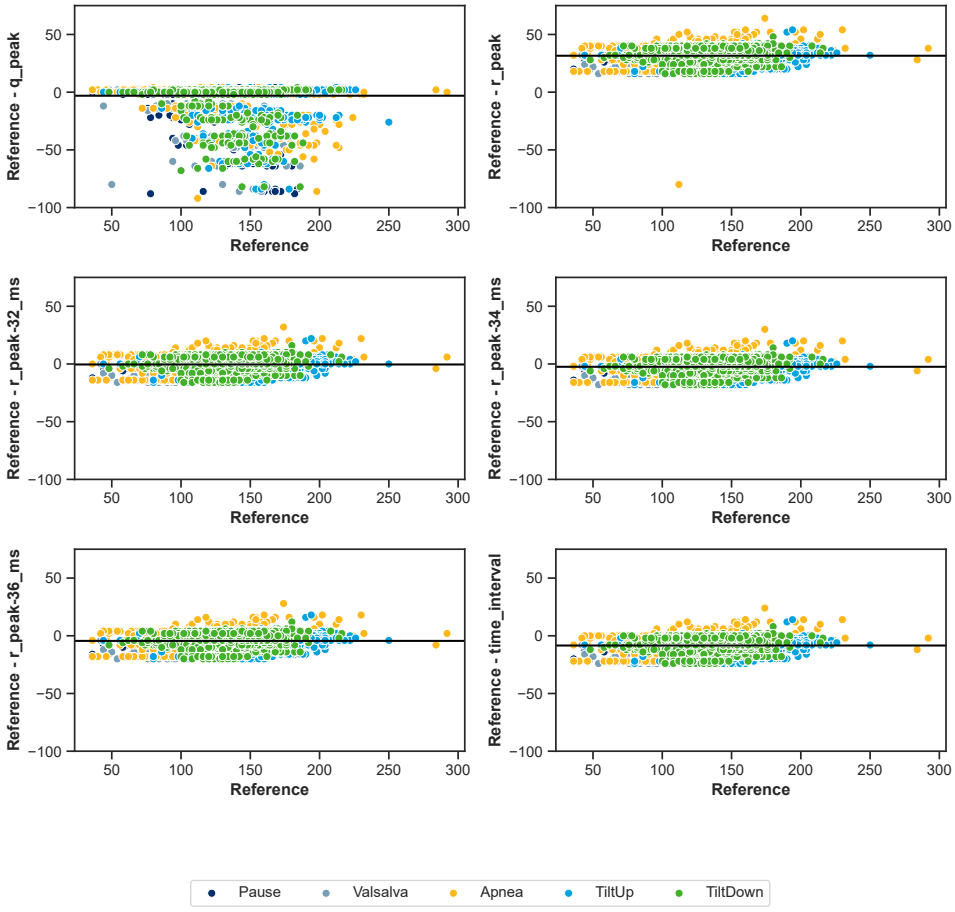

Figure B.4: Residual plots of the Q-Wave onset algorithms for the GUARDIAN Dataset divided into phases

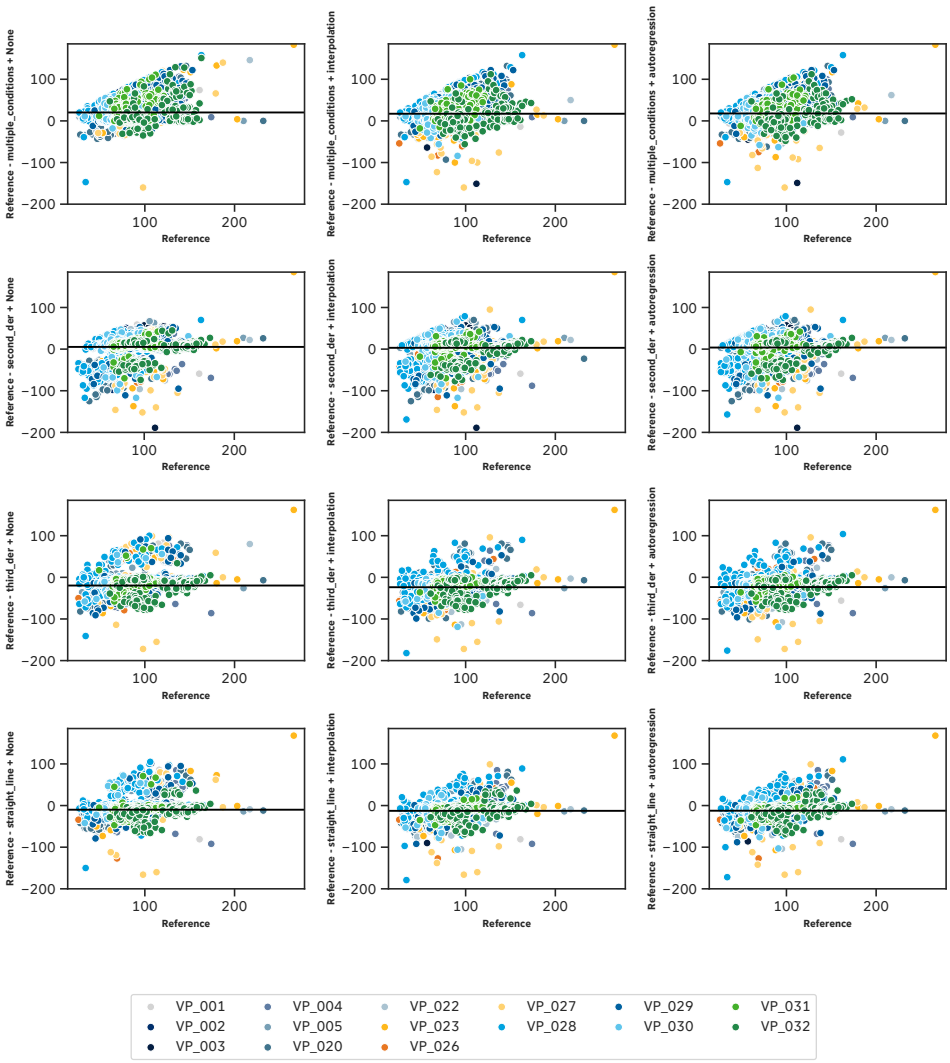

Figure B.5: Residual plots of the B-Point algorithms for the TSST Dataset divided into participants

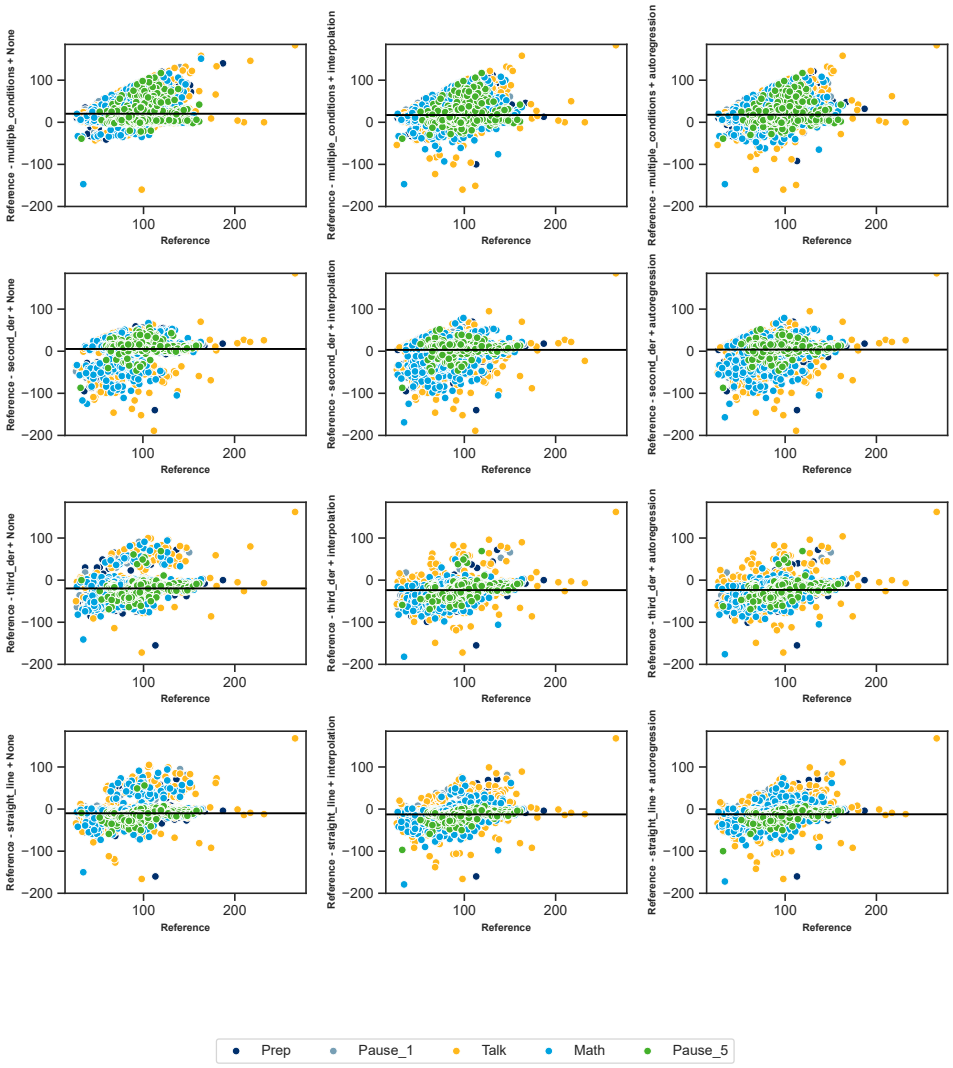

Figure B.6: Residual plots of the B-Point algorithms for the TSST Dataset divided into phases

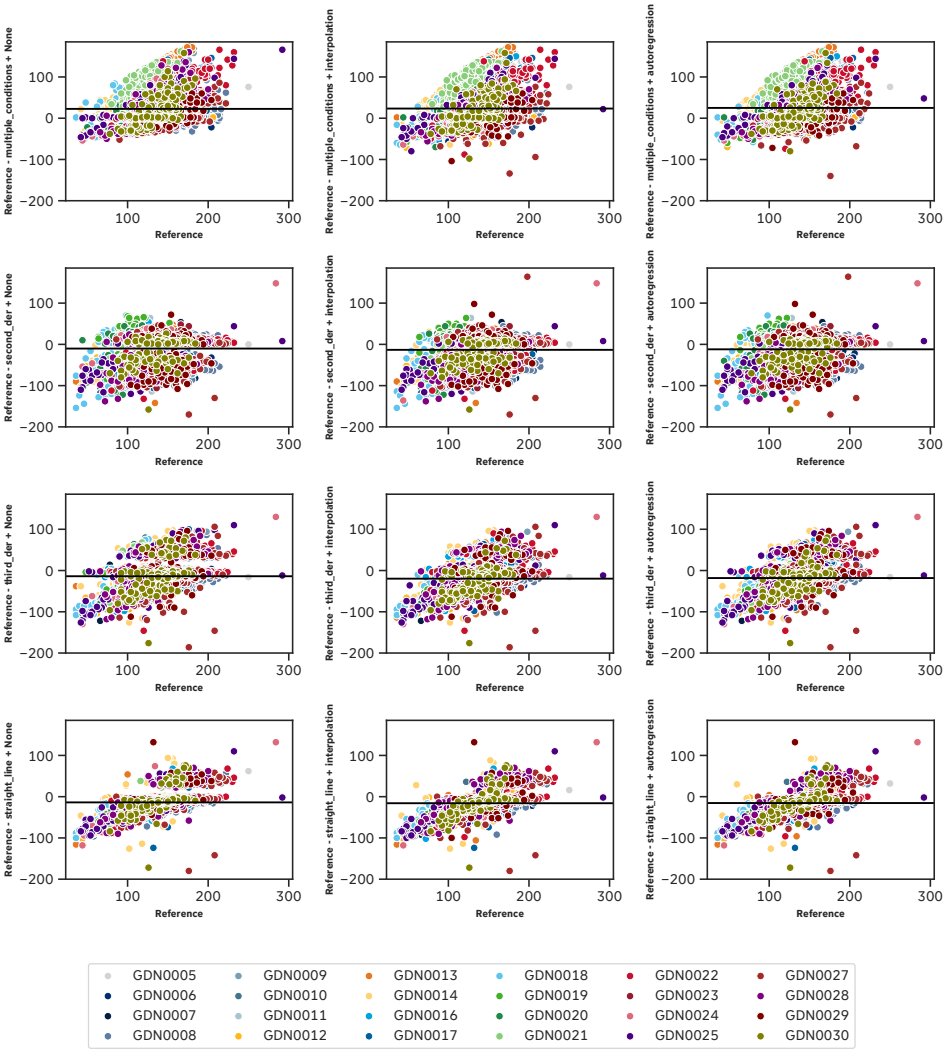

Figure B.7: Residual plots of the B-Point algorithms for the GUARDIAN Dataset divided into participants

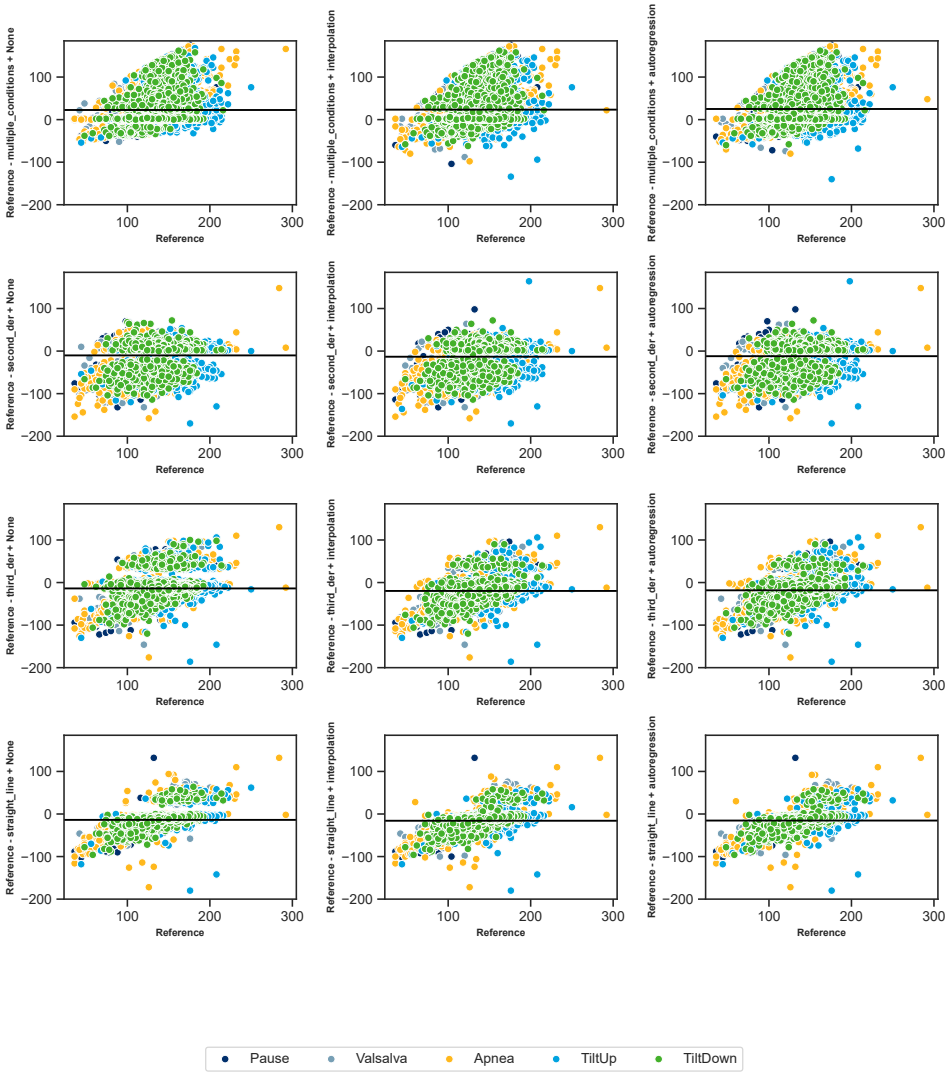

Figure B.8: Residual plots of the B-Point algorithms for the GUARDIAN Dataset divided into phases

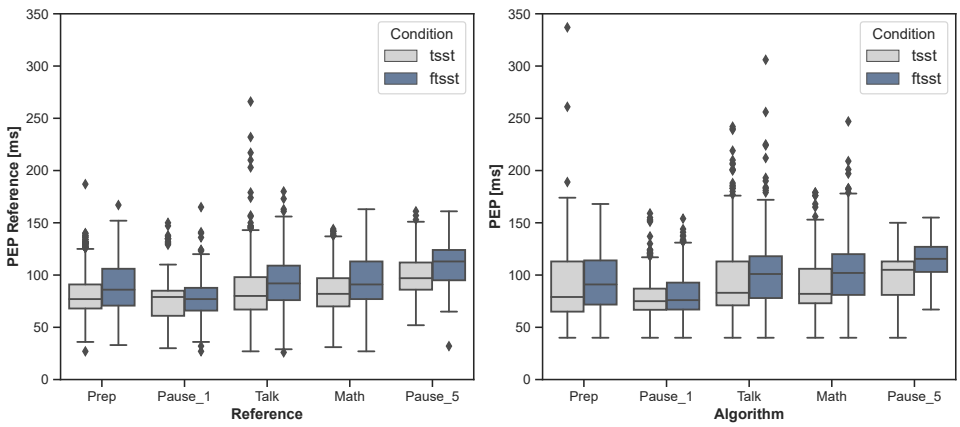

Figure B.9: PEP duration per phase compared between best-performing pipeline with AutReg for the TSST Dataset and the reference data

Appendix C

Acronyms

GUARDIAN **GU**arded by **A**dvanced **R**adar technology-based **DI**agnostics **A**pplied in palliative and intensive care **N**ursing

ECG Electrocardiogram

TSST Trier Social Stress Test

f-TSST Friendly Trier Social Stress Test

TTT Tilt Table Test

PEP Pre Ejection Period

ICG Impedance cardiogram

dZ/dt first derivative of the cardiac impedance

ANS Autonomic Nervous System

SNS Sympathetic Nervous System

PNS Parasympathetic Nervous System

HR Heart Rate

HRV Heart Rate Variability

dZ^2/dt^2 second derivative of the cardiac impedance

dZ^3/dt^3 third derivative of the cardiac impedance

ARIMA Autoregressive Integrated Moving Average

AIC Akaike information criterion

IMU Inertial Measurement Units

MAE Mean Absolute Error

ME Mean Error

RE Relative Error

PPG Photoplethysmogram

EEG Electroencephalography

DEAP Dataset for Emotion Analysis using Physiological Signals

BP blood pressure

QP Algorithm that uses the Q-Peak as start point

RP Algorithm that uses the R-Peak as start point

RP-40 Algorithm that subtracts 40 ms from the R-Peak to identify the start point

RP-32 Algorithm that subtracts 32 ms from the R-Peak to identify the start point

RP-34 Algorithm that subtracts 34 ms from the R-Peak to identify the start point

RP-36 Algorithm that subtracts 36 ms from the R-Peak to identify the start point

RP-36 Algorithm that subtracts 36 ms from the R-Peak to identify the start point

MultCon Algorithm that uses multiple conditions to identify the end point

SecDer Algorithm that uses reversal points of the second derivative to identify the end point

ThirDer Algorithm that uses local maxima of the third derivative to identify the end point

StrLin Algorithm that uses the maximal distance to a straight line between the C-Point and 150 ms before the C-Point to identify the end point

Intpol Interpolation

AutReg Autoregression