MASTER

Heart Rate Variability Analysis within a Data-Enabled Design research

Nazarchuk, Iryna

Award date:
2021

Department of Mathematics and Computer Science
Architecture of Information Systems Research Group

Heart Rate Variability Analysis within a Data-Enabled Design research

Master Thesis

Iryna Nazarchuk

Supervisors:
Dr. Natalia Sidorova
n.sidorova@tue.nl

Peter Lovei
peter.lovei@philips.com

Final version

Eindhoven, August 2021

Abstract

During the past few decades the field of medical healthcare services experienced massive developments with regard to the wide spread of affordable user-friendly health monitoring devices. These wearable, non-invasive sensors enabled collecting data about user behaviour and health condition, and automatic storage on the cloud for the further analysis. The integration of the wearable devices with wireless technology gave a potential for conducting insightful analysis and data processing based on immense generation of physiological data that should be supported by scalable and adaptive infrastructure. This thesis project was focused on taking the Heart Rate Variability (HRV) data as a the main input source for the analysis that was aimed at bringing insights into the personal experience. Existing solutions in the field of Heart Rate Variability focused on the cloud-architecture designing side, rather than the actual profound research into a reliability of the heart rate measuring devices. A more efficient *data-enabled* approach, which is presented in this thesis, is to *explore the constraints of the Heart Rate Variability data to understand the minimum requirements in terms of quality that affects the feasibility of using it for measuring the emotional state of the healthcare practitioners within the hospital setting.* Within a MEX Oncology project at Philips Design, the goal was to develop a toolkit that enabled analyzing the experience of the healthcare practitioners and identifying the key moments of interest that influenced their emotional state. Hence, the task to create a system that allowed collection and analysis of the experience-related data was targeted in this thesis.

Research questions were addressed: (1) how the combination of physiological sensing and self-reflecting annotations can be utilized for emotional state and physiology analysis? (2) how much missing data and specific patterns of it can affect the reliability and validity of HRV measurements? (3) how to estimate the influence of the device on the reliability of the derived HRV features?

These research objectives are achieved in three phases, namely designing an ecosystem of micro-intelligences, performing missing data pattern identification and analysis, and conducting device reliability evaluation. By incorporating the Data-Enabled Design approach, we were able to derive a comprehensive set of design decisions that allowed to develop a system for collecting behavioural, contextual and experiential data. In the experimental use cases, we were able to demonstrate that by combining physiological signals data, self-annotations and post-hoc reflections it is feasible to design a full coherent system for conducting affective computing in a situated manner. Afterwards, a more profound research was carried out to evaluate the accuracy and reliability of the Heart Rate Variability data collected by the particular PPG device. The thesis proposes a framework on identifying the specific patterns of missing data based on their continuous probability distribution type as well as presents the analytical steps for estimating the error in the Heart Rate Variability data. Furthermore, the thesis provides systematic methods that enable performing a benchmark comparison between the acquired data from multiple heterogeneous sources to assess the level of discrepancy between the gold-standard HRV values and those derived from the tested device.

This Master thesis work brings three core contributions to the field of the Heart Rate Variability research, specifically (1) introducing a micro-intelligence approach for designing and testing the ecosystem for situated affective computing, (2) making the missing data pattern analysis more specific with regard to the Heart Rate Variability context, (3) validating particular PPG sensor accuracy with regard to the two ECG devices.

Keywords: heart rate variability, affective computing, data-enabled design, situated design experiments

Preface

This thesis project represents the work carried out to fulfill the graduation requirements of the Big Data Management and Analytics Master Programme on the track of Business Process Analytics, a joint programme between Université Libre de Bruxelles (ULB) in Belgium, Universitat Politècnica de Catalunya (UPC) in Spain and Eindhoven University of Technology (TU/e) in the Netherlands. It has been conducted at Department of Mathematics and Computer Science at the Eindhoven University of Technology and Philips Design.

The research was performed under the supervision of Dr. Natalia Sidorova from Eindhoven University of Technology. I am very grateful to have been working with her for the last six months, as throughout this period I was able to deepen technical skills to the extensive level and significantly progress professionally. I would like to thank Dr. Natalia Sidorova for constantly giving me the direct feedback, defining the directional vector of my research and sharing her knowledge with me.

This thesis project summarizes the results of my internship process organized by Philips Design. I would like to express the gratitude to my supervisor Peter Lovei for always providing me with a practical guidance regarding the implementation, suggesting relevant ideas on the areas of improvement and supporting my research initiatives. I would also like to profoundly thank Eva Deckers for providing me with an opportunity of being part of Philips Design and making an invaluable contribution to my future career.

Lastly, I would like to thank my family that was always giving me an unconditional support during the challenging times and feeling proud of me as I was reaching my goals. Without them I would never be able to achieve some of my most unattainable dreams.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This master thesis is conducted as a part of Erasmus Mundus Joint Master Degree Programme in Big Data Management and Analytics (BDMA) in the specialization of Business Process Analytics, a joint programme between Université Libre de Bruxelles (ULB) in Belgium, Universitat Politècnica de Catalunya (UPC) in Spain and Eindhoven University of Technology (TU/e) in the Netherlands. It has been performed at the Process Analytics group of the Department of Mathematics and Computer Science at the Eindhoven University of Technology and Philips Design.

1.1 Thesis Context

During the past few decades the field of medical healthcare services experienced massive developments with regard to the wide spread of affordable user-friendly health monitoring devices. From the user perspective there was noticed a grown demand in reliable non-invasive physiological measurements for monitoring the biological indicators during workouts and routinely performed tasks (e.g. breathing rate, heart rate, energy expenditure). An increasing number of emerging *wearable activity trackers* (WAT) (e.g. Fitbit, Xiaomi, Garmin, Samsung Gear Fit) enabled self-monitoring and provided with the capability to enhance human behaviour toward a healthy lifestyle.

The wearable devices are usually attached to the body, mainly to a wrist, arm or chest to aid the facilitation of human health management opportunities. Moreover, WAT enabled collecting data about user behaviour and health condition, and automatically store it on the cloud for the further analysis. Hence, the integration of wearable sensors with wireless communication technologies made it possible to change the perspective of healthcare services from a clinic-centric level to the patient-centric level. This inter-connection gives a potential for conducting insightful analysis and data processing based on the immense generation of physiological data that should be supported by scalable and adaptive infrastructure. Furthermore, it brought a room for research to analyze human experiences and provide the biological feedback that through monitoring the quantifiable body functions of a person provides guidance on how to respond to specific emotional situations.

One of the quantifiable physiological data source is the heart rate (HR), which represents heartbeat frequency as beats per minute (bpm). Although HR is able to reflect the drastic changes in the human body, it holds a limitation for analysis when it comes to understand the trends and distinguish stress from workout in the high HR ranges, as it is very individualized to the subject of age, physical health state, fitness level, time of the day, etc [3]. Those shortcomings were resolved by introducing the Heart Rate Variability (HRV), which is a more precise and highly specific measurement, that indicates beat-to-beat variation in the time between consecutive heartbeats. HRV represents non-invasive, unobtrusive information about modulation of heart rate by the autonomic nervous system in a variety of dynamic circumstances, including evoked emotions, and exercise [4]. Health and self-regulatory capability, and resilience, are correlated with an optimal level of HRV. In general, variability reflects the capacity of the body to deal with stressors. Any significant change away from the baseline may be an indication of overtraining, sickness or just lack of sleep. Hence, within the scope of this research we are aiming at performing the HRV analysis that would enable linking and interpreting cause and effect states, i.e. physiological changes that trigger human emotional shifts.

When designing a framework for physiological and affective analysis, Data-Enabled Design (DED) methodology can be applied. The DED approach is commonly used in the decision-making process during the researches that aim at analyzing human experiences. Data-Enabled Design methodology lies at the intersection of an experimental prototype-centric perspective and a user-experience oriented perspective. It proposes a situated design approach that uses data as creative design material [5]. Thus, to enable the development of the framework for systematic analysis based on the physiological data, the DED process should be performed at every iteration, from prototyping to testing and releasing.

1.2 State of the Art

Heart Rate Variability measurements were previously integrated into the health monitoring systems within a wide range of researches. From existing literature we identified that the main approaches in developing cloud-based HRV monitoring systems were utilizing a micro-service architecture.

Hussein et al. (2018) study developed a remote and automated diagnostic system for monitoring the patient's heart condition. The idea was to help in preventing the potential heart diseases of the patients that were recovering from diseases. The system comprised a wearable ECG device, ECG system nodes, a main processing server, data files and a web Graphical User Interface (GUI) server. ECG system node consisted of 3 components, specifically pre-processing, heartbeat peak detection, and data transmission. The researchers presented a proof-of-concept system tested on an extensively collected data that enables remote monitoring as well as provision of the urgent aid for the elderly people [6].

In the study conducted by Haoyu et al. (2018) a cloud-based real-time apnea detection scheme using HRV signals was proposed. The approach adopted the Internet of Medical Things (IoMT) techniques for real-time detection that enables triggering the true alarms and predict an alarm as well as reduce the false alarm cases. The proposed scheme consisted of 3 components: the SpO2 sensor, the gateway and the cloud infrastructure. SpO2 is a wearable device based on PPG technology that collects signals for extracting the HRV parameters and oxygen level in the blood [7].

Another research performed by Zhou et al. (2016) presented an iHRV system that focused on mobile HRV data collection and subsequent analysis via the Canarie Cloud. The designed system is composed of both hardware and software is able to measure HRV in real-time, analyze the trends in HRV and send a medical alarm when the HRV data is abnormal. Once being uploaded to the cloud, the data was being pre-processed and analyzed to determine if it is normal. In case the strong deviations were found in the HRV pattern, a system generated and sent an alert email to the doctor [8].

Based on the outlined studies in the domain of HRV monitoring systems we can conclude that the majority of existing solutions developed coherent frameworks for accurate analysis of the human physiological conditions. Therefore, there is a need for developing an intelligent ecosystem for conducting a precise HRV analysis in a systematic way.

However, all the previously developed solutions were primarily focused on the cloud-architecture designing side, rather than the actual profound research into a reliability of the heart rate measuring devices. Hence, to address the shortcomings of the existing researches, this Master project is targeted at developing an intelligent ecosystem for affective computing and providing a framework that enables benchmarking the heart rate recording devices to understand the feasibility of their integration within specific context into the system of micro-services.

1.3 Research Problem

The field of Heart Rate Variability research has been actively under exploration since the 1960s, and in 1996 the majority of the measurements and calculations were made according to the standard of the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology [9]. Albeit, the appearance of modern, fast-performing devices brought new challenges into the approach of Heart Rate Variability analysis:

- Inaccuracy caused by immense data loss due to wireless connectivity issues and motion artifacts that produce noise
- Lack of comprehensive non-commercial research aimed at verifying the reliability of using PPG-based heart rate monitors as the data acquisition tool for objective physiological measurement
- Ambiguity of the insights derived solely from physiological data and disconnection from the ground-truth

Data-enabled design methodology has been already proven to be efficient for designing frameworks that facilitate collection of behavioural, contextual and experimental data [5]. Hence, this approach is an essential part for developing systems for coherent affective computing.

The conducted work in the thesis is performed within the context of the Philips MEX Oncology project, whose aim is to "design a systematic framework using data-enabled design approach that enables collecting, analyzing and deriving insights from reliable Heart Rate Variability data as well as allows integration into an ecosystem of micro-intelligences".

Consequently, the work presented in *this thesis* is pursuing to address the following goal:

Explore the constraints of the Heart Rate Variability data to understand the minimum requirements in terms of quality that affects the feasibility of using it for measuring the emotional state of the healthcare practitioners within the hospital setting.

To achieve this goal, one has to take into account that the data collection must be performed by making the healthcare practitioners (HCPs) wear unobtrusive sensors for recording the physiological data. In order to measure the emotional state of the HCPs, the Heart Rate Variability data has to be used as a proxy for representing the changes in human physiology. Finally, there is a need to investigate to which extent the collected data is reliable and accurate enough to derive the insights regarding the emotional condition of the HCPs.

1.4 Research Questions

The following research questions arise as a result of the research problem outlined in the previous section:

1. How the combination of physiological sensing and self-reflecting annotations can be utilized for emotional state and physiology analysis?
2. How much missing data and specific patterns of it can affect the reliability and validity of HRV measurements?
3. How to estimate the influence of the device on the reliability of the derived HRV features?

1.5 Approach

Within a MEX Oncology project at Philips Design, the goal was to develop a toolkit that enabled analyzing the experience of the healthcare practitioners and identifying the key moments of interest that influenced their emotional state. Hence, the first task was to create a system that allowed collection and analysis of the experience-related data. The Data-Enabled Design approach, that is commonly adopted within Philips Design working projects, was used as main methodology during prototyping and testing phases. Thus, the main idea was to check the feasibility of the chosen devices for data collection within a system for affective computing.

As a first step, the Data-Enabled Design was utilized for facilitating the decision-making process regarding on which devices to use for data collection, what is the feasible duration of the recording, how to synchronize all the sensors into one coherent system for data acquisition. Following this methodology, we aimed at designing a loosely-coupled and highly maintainable infrastructure architecture. Hence, the goal is to present an approach to design this intelligent ecosystem, test all its standalone components, understand how the combinations of physiological sensors and affective annotations can bring the meaningful insights, and integrate all layers of collected data into a coherent ecosystem for situated computing. In addition, by incorporating data from the open-source datasets there is a need to develop a set of steps through visual analytics for personalized analysis based on the derived Heart Rate Variability parameters.

Taking into account constraints introduced during the contextual step of DED study on unobtrusive devices, the aim is to test the feasibility of using a particular PPG device, introduced as a component in the designed ecosystem, for collecting data during a set of specific activities that involve huge amount of movements. Hence, we need to present a method that allows understanding whether the device can be incorporated into the physiological data layer and estimate to which extent the data loss limitation can be tolerated for providing the accurate Heart Rate Variability parameters. For this purpose, firstly the data was collected within a home-working setting due to Corona reasons and afterwards visually inspected during the data exploration step. The concept of missing data patterns was introduced and continuous probability distributions were identified for reproducing the missing patterns on the complete dataset. To evaluate the change in the accuracy of the Heart Rate Variability parameters, an open-source dataset was used for imputing with artificially generated missing data patterns and utilized later on to derive the HRV values.

Once the nature of missing patterns has been researched, further benchmark analysis of the PPG device is conducted to evaluate the reliability of incorporating it as a source of physiological signal in the intelligent ecosystem with regard to the gold-standard measurements collected by the ECG sensor. The idea is to simultaneously perform data collection using two different sensors and assess to which extent Heart Rate Variability parameters features computed from PPG-obtained data significantly deviate from the ones recorded by ECG sensor. We performed the pre-processing and transformation of the raw ECG signal data into datasets containing clean inter-beat interval values. Afterwards, the comparison between datasets is implemented using visual analytics and statistical tests. Finally, the error estimation between the derived Heart Rate Variability parameters is performed using a set of regression error metrics.

To summarize, the approach presented in this thesis consists of (1) designing the ecosystem of micro-intelligences for situated affective computing, (2) assessing the accuracy of the device used for the HRV data collection and (3) estimating the feasibility of integrating it into the intelligent ecosystem for performing reliable data acquisition that leads to deriving quality insights.

1.6 Structure of Thesis

By answering the research questions that were presented, this work aims to design a systematic framework using data-enabled design approach that enables collecting, analyzing and deriving insights from reliable Heart Rate Variability data. The remainder of this thesis is structured as follows:

- Chapter 2 introduces the preliminary information that is incorporated throughout the methodological parts of this thesis, such as Heart Rate Variability concept, measuring techniques and analysis tools. The chapter also provides the comprehensive discussion of all prior work in the area of this subject.

- Chapter 3 presents the designed, following the Data-Enabled Design approach, ecosystem of micro-intelligences, that are used to remotely collect behavioral, contextual and experiential data within the conducted experiment. The ecosystem architecture is being discussed and a method for personalized analysis of the physiological data is demonstrated.
- Chapter 4 illustrates the method for evaluating the amount and impact of missing data during cardio-active recorded activities, identifying missing data patterns and conducting the error estimation analysis on the datasets with artificially imputed missing data.
- In Chapter 5 the method for heart rate monitor reliability assessment with regard to the benchmark device is described. The Chapter illustrates the insights obtained through the HRV analysis of the data coming from multiple sensors and discusses the level of discrepancy between regular measurements and gold-standard ones.
- Chapter 6 presents the conclusion of the thesis and addresses the future work that can be performed as a continuation of this thesis.

Chapter 2

Background

2.1 Heart Rate Variability

Heart Rate Variability (HRV) represents a non-invasive, unobtrusive information about modulation of heart rate by the autonomic nervous system in a variety of dynamic circumstances, including evoked emotions, and exercise [4]. HRV is an evolving property of interdependent regulatory systems that work to help us respond to environmental and psychological challenges on various time scales. HRV represents autonomic balance, blood pressure (BP), gas exchange, gut, heart, and vascular tone regulation, which refers to the diameter of the BP-regulating blood vessels, as well as the muscles of the face. It is commonly defined as the fluctuation in the time intervals between adjacent heartbeats that are called R-waves [10]. Health and self-regulatory capability, and adaptability or resilience, are correlated with an optimal level of HRV. Higher levels of resting vagally mediated HRV are correlated with the prefrontal cortex output of executive functions such as attention and emotional processing [11].

In general, variability reflects the capacity of the body to deal with stressors. Any big change away from the baseline may be an indication that overtraining, sickness, or just lack of sleep is not right [12, 13]. HRV analysis has been widely used in numerous cohorts, and plays an important role in describing the patients' autonomic dysfunctions, tracking the natural fluctuations of autonomic function, evaluating the autonomic changes following various interventions, and predicting prognosis [14].

In recent years, the number of studies related to HRV has risen steeply. For example, studies have indicated that higher HRV is associated with decreased morbidity and mortality, psychological well-being and quality of life, as well as improved physical health and reduced individual age. Moreover, acute stress has been associated with decreased HRV during sleep and during daytime. In addition, decreased HRV has been associated with work stress in many studies [15]. Furthermore, heart rate variability is often utilized as a measure to assess the risk of a sudden death and developing pneumonia, provide prognosis for patients with heart transplants, monitor hypertension, track immune response, provide biofeedback and optimize strength training [12].

Whilst HRV analysis has previously been restricted to research applications, the increasing availability of significant amounts of computational power is making the widespread clinical use of HRV analysis feasible. It is therefore opportune to consider the possibility of improving methods for acquisition of the physiological signal from which HRV measures are derived [16].

2.1.1 HRV measuring techniques

Measurement domains of HRV

One of the approaches to measure the Heart Rate Variability is by using the spectral analysis of the data derived from *ECG* or *PPG* signal. While ECG signals enable to calculate the *R-R time* *interval* (RRI) of two successive R-peaks, similarly PPG signal can be used to derive *peak-to-peak interval* (PPI) as the time interval between two successive P-peaks [16].

Variations in heart rate are most commonly calculated in the *time domain* and *frequency domain* as well as with non-linear analysis. This work is focused primarily on the HRV analysis based on time domain and frequency domain variables. Time domain indices of HRV quantify the amount of variability in measurements of the *inter-beat interval* (IBI), which is the time period between successive heartbeats. These values may be expressed in original units or as the natural logarithm (Ln) of original units to achieve a more normal distribution [10]. The most established *time domain* parameters include:

- *IBI or R-R*: Inter-beat interval
- *BPM*: Beats per minute
- *SDNN*: Standard deviation of all normal inter-beat intervals
- *rMSSD*: Square root of the mean squared differences of successive inter-beat intervals
- *Mean RRI*: Average inter-beat interval
- *NN50*: Number of interval differences of successive inter-beat intervals larger than 50 ms
- *pNN50*: Percentage of the number of interval differences of successive inter-beat intervals larger than 50 ms divided by the total number of inter-beat intervals
- *HR Max and HR Min*: Average difference between the highest and lowest heart rates during each respiratory cycle

Frequency domain measures perform more complex calculations on IBI (Fourier transforms), expressing variability in terms of a power density spectrum (energy in specific frequency bands). Frequency domain measures can be calculated for any frequency band [17]. The most common frequency domain parameters are:

- *VLF*: Very low frequency component (0.0033–0.04 Hz).This metric indicates both PNS and SNS activities
- *LF*: Low frequency component, 0.04–0.15 Hz. This metric indicates both PNS and SNS activities
- *HF*: High frequency component, 0.15–0.4 Hz. This metric describes PNS activity
- *LF/HF ratio*: Ratio of the power in LF divided by the power in HF. This metric reflects the balance between PNS and SNS activities

Overview of ECG and PPG technologies

The monitoring of heart rate variability activity is commonly detected through *electrocardiogram* (ECG) or *photoplethysmography* (PPG) [18].

Electrocardiogram is usually considered as a gold-standard method to obtain the HRV. HRV analysis and extraction of a respiratory signal from the ECG are important ways to assess the heartbeat [19]. Each R wave in the electrocardiogram is caused by depolarization of the main mass of the ventricular myocardium. However in theory any discrete event in the cardiac cycle may be repeatedly measured to produce a record of successive heartbeats [16].

The ECG requires solid contact via adhesive pads attached on skin (electrodes). The QRS complex, defined by a combination of the Q wave, R wave and S wave, is the product of ventricular depolarization, accompanied by heart muscle contractions that pump blood to the pulmonary and aortic arteries.

The R-R interval is considered as the time between the initiations of succeeding heart beats. Every heart contraction results in pulses of blood volume which circulate in the blood circulation.

There is a clear need for a comfortable indicator that does not interfere with rest in order to compare long-term data with lifestyle adjustments over weeks, months and years, as the ECG is not comfortable enough to wear for continuous use [20].

Furthermore, ECG recordings are often imperfect. Popular noise sources are those produced by physiological processes, including contamination by electromyographs, signal interference and baseline drift caused by respiration, as well as those generated by non-physiological effects such as interference with the power line and electrode contact motion [16].

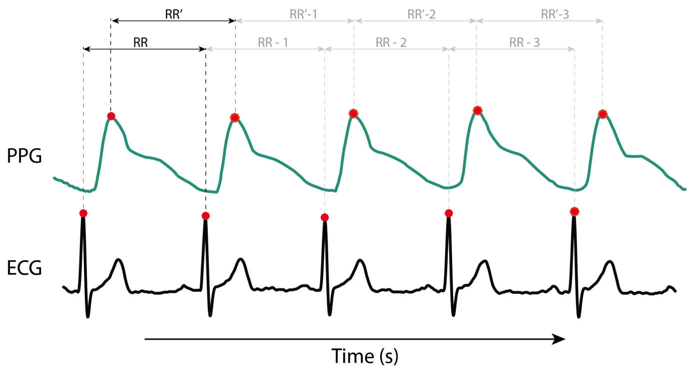

Figure 2.1: ECG and PPG signals [1]

Compared to ECG monitoring which needs to attach several electrodes on the body, PPG monitoring at peripheral positions is much more convenient and non-obtrusive. The sensors can be embedded in the phone camera, ring, glass, chair, or bed sheet for unobtrusive measurement. Therefore, PPG was proposed as a surrogate of ECG for the analysis of HRV [21].

PPG signal can be obtained with a more comfortable measurement that is also suitable for long-term nocturnal use. As the pulse propagates to the peripheral arteries and capillaries, the PPG signal originates from blood volume pulses. Two wavelength PPG is routinely used in hospitals to derive oxygen saturation, usually with fingertip sensors. More recently, HRV has also been estimated using pulse intervals from PPG. Its accuracy seems to vary depending on the application [20].

Several studies examined the difference between ECG and PPG recordings aided to analyse heart rate variability. Lu G et al. (2009) [16] investigated the feasibility of using PPG signals for analysing the HRV in comparison with ECG techniques. In the case of ECG recordings, the extraction method incorporated a peak detection algorithm that found the durations between successive peak locations were calculated to produce a time series of R–R intervals (RRIs). For PPG obtained signals, a neighbouring peak searching method was used to derive the peak events from the amplitude of the filtered PPG signals and then the intervals between the successive detected peaks (PPIs) were calculated. The conducted experiment of comparing 5-minute recordings from PPG and ECG demonstrated a high degree of correlation in the temporal and frequency domains and in nonlinear dynamic analyses between HRV measures. The study confirmed that PPG provides accurate interpulse intervals from which HRV measures can be accurately derived in healthy subjects under ideal conditions as an alternative to ECG for HRV analysis [16].

In another research [19], authors were aiming to identify the differences between heart rate variability from ECG and pulse rate variability from PPG based on the breathing frequency and volume. Heart rate (HR) and pulse rate (PR) were compared in the breathing tasks, including spontaneous breathing and controlled breathing. Obvious coherence was observed between PR and HR in both the resting and controlled breathing tasks. As a result, the recording of PPG-derived PR was consistent with that of ECG-derived HR recorded in the breathing interval. It was found that the PPG-derived PR was an acceptable surrogate of ECG-derived HR [19].

Plews et al. (2017) study was comparing the accuracy and validity of HRV recordings obtained via a PPG smartphone application (HRV4Training), and via the Polar H7 Chest Strap alongside electrocardiography (ECG). The root mean sum of the squared differences between R-R intervals (rMSSD) was determined from each device after conducting measurements. Research identified that rMSSD derived via PPG and Polar H7 during guided and normal breathing both shared acceptable agreement to HRV recorded via ECG. This method was therefore established as the most appropriate solution to choose when evaluating HRV, given the superior practicality and strong validity of HRV obtained through PPG with directed breathing [22].

Time-window length in the HRV analysis

The *time-window* used for spectral analysis of the HRV data from both ECG and PPG recordings is a key component for the assessment of cardiovascular autonomic function.

Kai Li et al. (2019) study reviewed the characteristics of spectral HRV studies using different lengths of time windows. The majority of existing works that use spectral analysis of HRV via fast Fourier transform or auto-regressive method work on ECG segments of 2–5 min, and previous applications of multiple TRS analysis are on ECG segments of 1–2 min. The main benefits of short-term HRV analysis consists of its ability to perform efficiently and explain the complex change in HRV within a short period of time, as well as allowing confounding variables such as body position, physical activity, breathing to be regulated. The key disadvantage is that the short-term HRV analysis might not be stable owing to the constant fluctuation of HRV parameters.

On the other hand, long-term HRV analysis can collect ECG information from 1h to an entire day. It is more stable than short-term analysis. Directly analyzing the entire long-term target time window can estimate longer fluctuations. Long-term recordings, however, are more costly and time consuming; more noise is collected by long-term recordings and it is difficult to deal with the noise.

It was reported that short-term HRV analysis is a convenient method for the estimation of autonomic status, and can track dynamic changes of cardiac autonomic function within minutes. Contrary, the long-term HRV analysis is a more stable tool for assessing autonomic function, it describes the autonomic function change over hours or even longer time spans, and can reliably predict prognosis [14].

Hagad et al. (2020) work focused on using the spectral HRV data representations based on PPG signal together with deep learning models to predict mental stress. Specifically, the study concentrated on utilizing a sliding window *Lomb-Scargle approach* to build a spectrogram of heart rate in lieu of the traditional *short-term Fourier transform* (STFT). In order to segment the data, a sliding window method was applied to the stress dataset. Afterwards, the analysis was performed based on different accuracy metrics obtained from model training on data segments of changing window lengths. It was observed that shorter segments tended to be more stable but not very expressive. Meanwhile, longer segments were more expressive but tended to overlap across samples with different labels. Through empirical testing, it was found that 505-beat windows (6 mins) showed the best balance between stability and expressiveness while staying above the 5-minute minimum length recommended for accurate short-term HRV analysis [23].

Sensors for physiological signals measurement

In order to obtain the physiological signal and derive heart rate variability data for conducting an analysis we focus on utilizing *PPG* sensors as the main HRV monitoring technology. Due to the ease of wearable integration and unobtrusive nature, it enables to continuously acquire heart function information, facilitates long duration monitoring and thus aids to provide long term insights into the health of a person. *Photoplethysmography* sensors are commonly used for measuring physiological signals in the vast majority of previous researches.

In the study towards hypertension prediction based on PPG-derived HRV signals [24], Kun-chan Lan et al. (2018) used the sensor board that comprised the ring probe wearable device and accelerometer. The ring probe was detecting the waveform of the user's PPG. The waveform was read by the sensor, which consists of a processing unit and a wireless radio transmission unit. The sensor was taking the waveform received from the ring probe and used it as the input to calculate the HR.

In another research conducted by Bhowmik et al. (2017) wavelength wristband PPG sensor was used to address the problem of RR series extraction. After applying the noise removal filtering and artefact correction algorithm it was proved that HRV extraction from long duration PPG signal acquired from the wrist can be utilized to monitor stress, myocardial infarction prognosis, fatigue levels and other health conditions [25].

HRV and accelerometer sensor streams were also used by Min Wu et al. (2015) to monitor perceived stress levels in daily life. Data collected by Mega Electronics sensors of two modalities (i.e. HRV and accelerometer) was used to generate the features for training the classification models aimed at stress recognition [26].

Jindal et al. (2016) study proposed a technique for heart rate calculation using smartphone accelerometer and PPG signal. The PPG signal was obtained using a wearable sensor system such as Fitbit, or by positioning the index finger on the camera lens using a mobile phone camera. The accelerometer sensor of the mobile was used to retrieve power spectrum density for all channels of signals. From each segment of PPG signal 17 different features were extracted and subsequently used for training the deep belief network. The resulting model was able to accurately determine and predict heart rate during intensive physical exercise [27].

For this research project PPG-based sensors are designed to be worn by health care practitioners (HCPs) during the whole working day to obtain the long-term and reliable physiological signals while not presenting any limitations on the movements and performing the occupation-specific tasks. Following the aforementioned requirements, Scosche Rhythm24 Heart Rate Monitor was used as a wearable wristband for monitoring and collecting the PPG signals. A study conducted by Plews et al. (2017) performed an analysis of the accuracy of the HRV measurement obtained through Rhythm24 Heart Rate Monitor built-in PPG sensors. It was concluded that measures of rMSSD derived from PPG sensors during guided and normal breathing both shared moderately acceptable agreement to HRV recorded via ECG [22].

2.1.2 HRV physiological insights

Applications of HRV

A growing number of recent researches are continuously advocating that organized variability in heart rate could serve as a reasonable index of general health, both physical and emotional. Lehrer et al. (2020) systematic review demonstrates that HRV can be utilized for measuring outcomes on activity reduction, anxiety, cognitive flexibility, fatigue, mindfulness, relaxation, positive feelings, sleep quality, etc [28].

In the review [17] authors discover that there is ample evidence that when physicians are under stress, quality of care is indeed reduced. Furthermore, a number of block factors such as fatigue, boredom, and stress can adversely affect cognitive performance. Reduced cognitive performance may continue to result in poor outcomes such as more error, low productivity which in turn cause financial loss [18].

Interest in the relationship between HRV and persistent emotional stress has increased after findings that perceived stress is predictive of transient myocardial ischemia TMI and that HRV is associated with perceived stress [4].

The autonomic nervous (ANS) system is stimulated when an individual is subjected to a stressor: the parasympathetic nervous system is suppressed and the sympathetic nervous system is activated. This results in the secretion into the bloodstream of the hormones epinephrine and norepinephrine, which leads to, for example, blood vessel vasoconstriction, elevated blood pressure, increased muscle tension, and a change in heart rate (HR) and variability in heart rate (HRV). When the stressor is no longer present, a negative feedback system stops cortisol production in the body, and a sympathovagal balance is established through homeostasis between the parasympathetic (vagal) and sympathetic system [28].

The et al. (2020) study on HRV-based measuring mental stress in surgery revealed that

HRV has proven to be a good objective assessment method of stress induced in the workplace environment and was able to pinpoint stressors during operations. In addition, HRV was able to determine which operating techniques provided most stress for surgeons and to determine differences in stress levels between performing and assisting in the surgical procedure [17].

In the research organized by Koldijk et al. (2018) challenges on detecting work stress using several unobtrusive sensors as well as taking into account individual differences were addressed. It was found that neutral and stressful working conditions can be distinguished with 90 percent accuracy by means of trained SVM model. Furthermore, by considering individual differences hence training models on particular subgroups of similar users the performance of stress prediction is highly increased [29].

Oskooei et al. (2019) performed a study of various unsupervised methods to identify mental stress in firefighter trainees based on unlabeled heart rate variability data. By means of clustering the data by convolutional autoencoders they were able to successfully stratify stressed versus normal samples thus predicting the stress patterns [30].

In addition to perceived stress, existing studies were also leveraging heart rate variability as a measure to identify the level of fatigue. Schmitt et al. (2013) longitudinal study was aimed at comparing heart rate variability (HRV) in elite athletes to identify either the 'fatigue' or 'no-fatigue' state in 'real life' conditions. It was discovered that values of HRV descriptors were significantly lower in 'fatigue' than 'no-fatigue' states in elite endurance athletes, both in supine and standing positions, over a multi-year period and without any 'artificial' change in their training regimen. The main trend by far was a 'fatigue'-linked lowering of total HRV power [31].

In another study by Butkevičiūtė et al. (2019) HRV parameters were analyzed during five stage training sessions for stages with less signal artefacts and the training impact on human fatigue was evaluated. The results of this study suggest that HRV is a potentially important measure of adaptive capacity in Chronic Fatigue Syndrome. While techniques such as HRV analysis show potential to detect both cognitive and physical fatigue, it seems important to distinguish various physical activities to improve the accuracy and robustness of this detection [32].

Interpretation of HRV trends

In order to derive meaningful insights on stress and fatigue level from the calculated HRV parameters there is a need to determine the ground truth from the HRV trend fluctuations.

Key physiological parameters that are derived from the RRIs and are commonly used in stress and fatigue evaluations are outlined in Section 2.1.1. In general, during stress arousal, the individual's heart rate(HR) is elevated, HRV is decreased from the basic resting level and respiration rate is low, relative to HR [15].

Speaking of domain parameters, the *frequency-domain* metric LF/HF that is usually derived to determine sympathovagal balance, is directly proportional to the stress levels. Low stress levels were identified to correspond to the low *LF/HF ratio* [33]. Moreover, lower HF power is also commonly correlated with stress, panic, anxiety and worry.

Regarding the *time-domain* metrics, since HRV is decreased during stress conditions hence all derived parameters in the time domain are also decreased. In the case of fatigue level assessment, an increase in *rMSSD* is usually associated with good physical performance and high energy level. Although, the correlation of this metric with fatigue condition should be always considered with regard to specific context, as decrease in *rMSSD* can be linked not with fatigue arousal, but associated to parasympathetic saturation or tapering.

HRV Biofeedback

To ensure the stable work performance in highly stressed hospital environments, there is a need to have an HCP-focused methodology that helps to deal with performance block factors. Psychophysiological self-regulation refers to a person's ability to regulate affective and cognitive states based on the activity of the autonomous and central nervous system. It is defined as a training approach that can be employed to achieve optimal performance through biofeedback "learning" method.

One of the commonly used biofeedbacks is Heart Rate Variability (HRV) biofeedback [18]. *Biofeedback* is an intervention that involves monitoring the quantifiable body functions of a person (e.g., blood pressure, heart rate, muscle tension) and transmitting the information to the individual in real time, is a helpful way to provide guidance and support for the effective control of the physiological response to stress [34].

In recent years a wide number of studies have been organized to investigate and prove the validity of HRV biofeedback methodology. Sutarto et al. (2010) study presented HRV-biofeedback as a potential tool for cognitive performance enhancement. HRV biofeedback was designed to control oscillatory variability in heart rate, hence directly targeting and exercising the body's own physiological control mechanisms. The research indicated that subjects who received biofeedback training improved their cognitive performance significantly [18].

Study conducted by Lemaire et al. (2011) was aimed to determine whether the use of a biofeedback-based stress management tool (consisting of rhythmic breathing, actively self-generated positive emotions and a portable biofeedback device) helps to reduce physician stress. It was concluded that biofeedback-based stress management tool is considered to be a simple and effective stress-reduction strategy for physicians [34].

In another meta review by Lehrer et al. (2020) all problems addressed by HRV biofeedback as well as all outcome measures were analyzed. Different biological and psychological conditions and issues with athletic, cognitive and creative performance were included in the objectives of the research.

It was found in the number of studies that HRV biofeedback does produce an improvement in the variety of physical and emotional conditions including anxiety, depression, hypertension, asthma, and pain, as well as improvement in various kinds of human performance including mental concentration and agility, athletics, dance, and music. Moreover, HRV biofeedback has been proposed as a psychotherapy component that specifically targets the neurovegetative components of emotional problems and may improve treatment effectiveness.

2.2 HRV Data Processing and Analysis

As was stated above, the Heart Rate Variability (HRV) signal conveys critical information about the systems that govern heart rate and blood pressure, which are primarily induced by *autonomic nervous system* (sympathetic and parasympathetic) regulation [35]. This section outlines the preliminary information regarding methods of processing the raw physiological data and existing tools that enable analytics applied on HRV.

2.2.1 HRV Data Processing

R-peaks detection algorithms

The majority of the robust ECG sensors provide raw data as a tachogram with millivolts (mV) units. A succession of beat-to-beat time intervals comprise the *R-R tachogram.* Hence, to construct a reliable R-R tachogram, a prominent spot on the ECG waveform that can be found consistently on each beat must be determined. Because the R-peak is the most easily identified feature, most beat detection algorithms are designed to locate this point. The *sampling rate* at which the ECG is digitised determines the precision of the location in time of each peak, and thus the accuracy of the value of each inter-beat interval that comprises the R-R tachogram. Moreover, as deviating from normal (sinus rhythm) beats are produced by processes other than normal conduction mechanisms, they are not recognized to be representative of autonomic control systems that manifest the observed variability in the R-R tachogram. The R-R tachogram must therefore exclude beat-to-beat periods that do not correlate to time gaps between two sinus beats [36].

Currently, the primary method of obtaining the HRV parametets is to record the *electrocardiogram* (ECG) signal, use appropriate *QRS complex* detection algorithms to locate the *R wave* and its peak, determine the *R-R intervals*, and conduct appropriate interpolation and re-sampling to produce a uniformly sampled tachogram [37].

The most scientifically acknowledged and widespread QRS detection algorithm within the scientific community is *Pan-Tompkins* real-time algorithm for ECG signals [38]. The algorithm was validated with regard to MIT/BIH arrhythmia database [39] to reliably detect QRS complexes using slope, amplitude, and width information. A band-pass filter pre-processes the signal to decrease interference, enabling low amplitude thresholds to be used to achieve high detection sensitivity. A dual-thresholds strategy was employed in the algorithm to explore backward for missed beats [38].

Hamilton-Tompkins algorithm was developed within the extended research to amplify the accuracy of original Pan-Tompkins algorithm. The key enhancement of the algorithm resides in the decision-making phase of the QRS complex detection. The algorithm is primarily focused on optimizing the decision rules, which is tested with regard to performance using three estimators (e.g. mean, median and an iterative peak level) to locate the adaptive threshold. The role of the decision rule section is to discriminate the QRS events from the noise events. Hamilton-Tompkins algorithm, as the previous one, incorporates a technique of discovering lost peaks that consists of a search back when a peak is not identified within a particular time interval, as well as refractory blanking to reject peaks that are closer to an already detected peak [40]. The optimized QRS complex detection algorithm has a sensitivity of 99.69 % and specificity of 99.77 % when evaluated with the MIT/BIH arrhythmia database [39, 41].

Recently, a novel real-time QRS complex detection algorithm was proposed by Kim et al. (2016), based on the physiological features of the ECG waveform. The introduced algorithm identifies the QRS complex using the dual criterion of amplitude and duration of the QRS complex. It is composed of basic processes such as a finite impulse response filter, differentiation, or thresholding, rather than complex and computationally expensive activities such as a wavelet transformation. Furthermore, the algorithm operates in a steady manner by employing extra parameters that correlate to different conditions, such as when there is surrounding noise or when the signal amplitude changes. Using the MIT-BIH arrhythmia and AHA ECG databases, the suggested technique was assessed. The obtained sensitivity and specificity after evaluation were 99.85 % and 99.86 %, respectively [42].

Ectopic beats and Artifacts Detection

In the most of the cases, even the most accurate heart rate monitors suffer from including noise when recording physiological signals. The majority of people have a normal cardiac (sinus) rhythm, which is comprised of relatively comparable beats that occur at regular (but slightly fluctuating) intervals. Ectopic beats are beats that emerge prematurely in the cycle, often with a deviating morphology [43].

According to Malik [44], there are two established methods that aim to detect and remove the ectopic beats.

In cases where the ectopic beats are infrequent, they are eliminated, and interpolation is employed to add a beat where a sinus beat would have been expected to occur. This is common in spectral frequency analysis approaches, where interpolation and resampling are frequently incorporated to make the typical methods for computing the power spectral density (PSD) easier.

Otherwise, if the occurrence of ectopic beats is highly frequent during a certain segment, it is recommended to discard the segments of the HRV signal that contain such a high occurrence from the analysis [36, 44]. In the research conducted by Mølgaard et al. (1991), it was elaborated how particular HRV features (e.g. rMSSD) are incredibly sensitive to missing beats during signal loss, particularly in patients with low HRV values, hence it is critical to assess the reliability of the data used in such scenarios [45].

When considering the threshold for rejecting specific R-R intervals, there is no basis of agreement within the HRV research community. However, a meta-study proposed by Karlsson [46] revealed that on average, the sequences with sinus outside the range 350 - 1350 milliseconds should be disregarded from subsequent HRV analysis to preserve its robustness. Most of the times, when experiments have been conducted to evaluate the performance of algorithms utilizing various R-R interval exclusion criteria, the types of noise and artifacts that are eliminated were not mentioned. Usually, the researches visually examined and removed data segments that appeared to be highly noisy.

In a study carried out by Wen et al. (2011), an efficient method to deal with ectopic beats and artifacts was presented. The strategy was based on the trend correlation of the cardiac timing signal. The weight computation and slope estimation of the previous normal RRI were used to generate a predictor of R-R interval (RRI) value at ectopic beat time. The predictor of RRI detected the type of ectopic beat and was replacing it. The standard value was used to estimate the performance of the simulated signal after ectopic correction, and the findings of clinical data with ectopic beats were compared to adjacent ectopic-free data. The examination of synthetic signal and clinical data revealed that frequency indexes corrected for ectopy had less error than other existing methods [47].

Research conducted by Chikh et al. (2004), the novel method was developed to identify normal beats and ectopic beats in the ECG signal using two artificial neural network classifiers. The approach consists of noise handling, feature extraction, and neural classification, all integrated in a three-stage procedure. The morphology of the detected QRS complexes are used to extract thirty features, and reduce them to seven coefficients using principal component analysis (PCA) and two coefficients using linear predictive coding (LPC) technique. The obtained data is passed as an input for two neural network classifiers. The neural network classifiers were evaluated on the MIT-BIH database and high scores were obtained for sensitivity and specificity (84.88 % and 91.92 % respectively using ACP technique) [48].

2.2.2 Overview of tools for HRV analysis

To enable the off-the-shelf Heart Rate Variability analysis, a number of analytical tools have been developed during the past few years.

HRVanalysis is a software developed by Pichot et al. [49] to allow the straightforward analysis over Heart Rate Variability data. The wide scope application of the HRVanalysis includes standard analysis of the short-term and long-term inter-beat timeseries, time-frequency analysis employing the wavelet transform, as well as autonomic nervous system status analysis on surrounding scored events and labeled areas. The supported input comprises RR or EKG files of various types, such as cardiac frequency meters, holters ECG, polygraphs, and data acquisition devices.

Bartels et al. [50] presented an alternative free software for advanced HRV analysis that enables the classical time- and frequency domain indices as well as techniques for non-stationary data analyses, *SinusCor*. SinusCor was created to satisfy the needs of physiologists who work with HRV in both clinical and research settings. This software allows the user to conduct the most common and clinically validated stationary and non-stationary HRV assessments. Moreover, the software provides the separated visualization of the findings of time-varying and time–frequency analysis, making user interpretation clearer and easier, which is a unique feature with regard to other existing solutions.

Kubios HRV software [51] is a free analytical toolkit for researches and clinicians with a non-commercial purpose of usage. It is considered to be most widely used tool in the research community and was acknowledged to provide gold-standard results for HRV analysis [52]. The software allows analysis of the HRV in time- and frequency-domain as well as supports the non-linear indices. The software is a cross-platform solution that is compatible with both Windows and Linux operating systems, supports both ECG and R-R interval data formats, and conducts the essential pre-processing activities such as QRS identification and artifact elimination.

ARTiiFACT [53] is a MatLab software tool for processing ECG and inter-beat intervals data. In a graphical user interface, the tool provides both automated and manual artifact detection and correction. Furthermore, ARTiiFACT provides time- and frequency-based HRV analyses, as well as descriptive statistics, providing the fundamental tools for HRV analysis. Due to its modular architecture and compatibility, it can be integrated with other software tools, replacing one or more of ARTiiFACT's subcomponents to optimize benefits by combining the benefits of other software solutions.

In addition to the graphical drag-and-drop software tool, an extensive number of Python packages have been developed to facilitate the HRV analysis. Among them reside those established libraries:

- *HeartPy*, Python Heart Rate Analysis Toolkit. The toolkit is designed to handle (noisy) PPG data collected with either PPG or camera sensors [2].
- *hrvanalysis*, a Python module for Heart Rate Variability analysis of RR-intervals built on top of SciPy, AstroPy, Nolds and NumPy and distributed under the GPLv3 license [54].
- *hrv*, a simple Python module that brings the most widely used techniques to extract information about cardiac autonomic functions through RRi series and Heart Rate Variability (HRV) analyses [55].
- *pyHRV*, a toolbox for Heart Rate Variability (HRV) that bundles a selection of functions to compute time-domain, frequency-domain, and nonlinear HRV parameters, along with other additional features designed to support the HRV research [56].

Chapter 3

Ecosystem of micro-intelligences for affective computing

The field of Artificial Intelligence (AI) has grown significantly in the previous decade and currently is on the verge of altering many parts of our daily lives and experiences that affect them. Services such as customer assistance and patient care, which were previously solely available through human–human interaction, can now be provided through robotic daily assistants and AI-enabled conversational chatbots [57]. Despite that, the emotional state of humans still continue to be hardly detectable from the standpoint of computing. Relying fully on the physiological measuring can lead to ambiguous and irrelevant insights, that are completely disconnected from the ground-truth. Hence, there is a need to incorporate the HRV physiological data into a coherent system that will allow measuring human experience with regard to emotional states.

The area of research is thus revolving around the interdisciplinary field of *Affective Computing* (also know as Emotional AI), that addresses the problem by designing machines that enable recognition, interpretation, modelling and responsive adaptation to human emotions [58]. The standardized Affective Computing processing pipeline entails (1) collecting objective, measurable indicators of human emotions, (2) collecting subjective annotations of internal emotions in a form of post-hoc reflections, and (3) modeling the relationship between these indicators and annotations to produce insights about the user's emotional state [59].

In order to design for the area of Affective Computing, it is possible to apply design methodologies, acknowledged by the scientific community. One of them is the Data-Enabled Design process. *Data-Enabled Design* (DED) has proven to be useful for gathering behavioral, contextual, and experimental data [59, 60]. As DED is an explorative and situated design approach, it enables researchers to investigate a remote context outside the design studio [61]. The DED process consists of the multiple steps, the first one being the *research-oriented contextual step*, where contextual data is gathered throughout the design process and the *design-led informed* step, when the acquired data is used to inform continuous exploration and updating the design [59].

Data-Enabled Design methodology was successfully incorporated by multiple researches for designing *intelligent ecosystems*. Bogers et al. (2016) conducted a design case study of a connected baby bottle, which is used as an exploration towards a data-enabled design framework for designing intelligent products, services and ecosystems targeting behavior change. The main focus of the research was to explore how contextual, behavioral and experiential data can be used as design material for remote insight collection and remote design intervention [62].

In another study, the Data-Enabled Design process was utilized for building Field Lab Sleep and Energy system for longitudinal remote sleep tracking and prototyping. Within the study, a communication platform was built and embedded aiming to prototype and test the RelaxBreathe program representing the human-IoT experience [63].

In a research by Jansen et al. (2020) the potential value of the intelligent ecosystem that is designed for the co-responsibility between patient, Health Care Professionals (HCPs) was investigated. The research presented a clinical trial following an explorative data-enabled approach and developed an ecosystem consisting of data trackers and personalized coaching interventions with six patients and their partners that allowed to identify six ways in which designing for co-responsibility can bring value [59].

In this Chapter we present an approach to design the ecosystem of micro-intelligences that enables conducting an analysis on the physiological data layer and combining it with the affect annotation layer for post-hoc reflections and event annotations. We introduce the concept of micro-intelligences and describe the essential components of the ecosystem for situated affective computing. By performing a pairwise analysis between multiple components we would like to understand how separate layers in one intelligent ecosystem can be used for conducting the contextual DED step. Hence, this chapter is addressing the problem of how the combination of physiological sensing and self-reflecting annotations can be utilized for emotional state and physiology analysis.

The content of this chapter outlines the conceptual and technical work conducted within the MEX Oncology research project at Philips Experience Design and was presented in the paper "Designing Micro-intelligences for Situated Affective Computing" for the CHIIot 2, a 2nd workshop on Computer Interaction in IoT applications [64].

3.1 Data-Enabled Design decisions

The *Data-Enabled Design* method is effective for situations in which researchers are unable to meet their study participants in person. Nonetheless, as the research-oriented contextual step aids at informing and inspiring the design process it is more beneficial to perform it in the on-site context. During this stage, there is importance in gathering more qualitative data through contextual inquiries or interviews. These activities entail more in-person engagement. Furthermore, many of the *"design plumbing"* [61] actions prior to remote data collection may necessitate the gathered contextual insights.

However, in the process of setting up the contextual phase of the research in the hospital, the Coronavirus pandemic started. Due to the circumstances, we were unable to carry out our initial plans for iterative development, testing, and improvement of our intelligent ecosystem with healthcare professionals (HCPs) as study participants. We concluded that our initial arrangement for the research-oriented contextual phase relied far too heavily on the option of being present in real-life at the hospital. As a result, we decided to take a step back and focus on improving the experience-related and functionality-related parts of our forthcoming contextual step by conducting small experiments in our study team's working from home environment.

The general intention of this part of the research is to systematically design separate input source layers that will allow to remotely collect behavioral, contextual and experimental data. As the main goal of the research is reflected in understanding the emotional states of the person by analyzing the objective and subjective collected data, we applied the best practices of DED approach. During the preparation phase of our project, we were still able to perform interviews, co-creation sessions, and a shadowing study with our clinical partners. Based on the gained insights, we made the following design decisions for the hospital's intended contextual DED phase.

3.1.1 Post-hoc reflections

As a first step, we explored the methods that would enable us to collect post-hoc reflections on the workdays of the participating healthcare professionals. Our decisions were constrained by the limitations of precautitiosness, as we were not intended to interfere into the workflow of the HCPs and increase their workload. As a result, there is a need to minimize the amount of interactions our employees have with the ecosystem during the course of their workday. Hence, we focused on allowing them to reflect on their workday once they had completed all of their responsibilities.

In practice, the whole process takes not more than 5 minutes of their time to answer the questions imposed by a Chatbot running on the Flow.ai platform [1]. We wanted HCPs to be able

[1] "Flow.ai—Chatbot design and management solution for professionals". Available at: https://flow.ai

to recollect important moments that occurred during their workday that significantly affected their emotional state with the help of short reflection questionnaire uploaded to their phone. Furthermore, from the conducted interviews we concluded that the majority of the HCPs have an access to the personal work calendar on their phone and can thus use this data source as a reference to recall all their performed activities at the exact time frames they would like to discuss with us about their workday. In this way we are able to provide one of the components related to the subjective input data layer.

3.1.2 Physiological signals

In addition to collecting the post-hoc reflections, there is also a need to include a method to capture the objective behavioural data from the context. Following this, we decided to embed the physiological signals data layer, that will allow us to measure the physiological state during the moments of interest for the HCPs. Taking into account the previous conducted research around reliability and accuracy assessment of the heart rate variability, we decided to proceed with the HRV as a behavioural input data source.

HRV represents non-invasive, unobtrusive information regarding the autonomic nervous system's modulation of heart rate under a variety of dynamic circumstances, including induced emotions and exercise [4]. It is typically defined as a change in the time intervals between consecutive heartbeats known as R-waves. An optimal level of HRV is associated with good health, self-regulatory ability, and resilience. In general, variability reflects the body's capability to deal with external stimuli. Any significant deviation from the baseline may indicate over-training, illness, or just a lack of sleep. During the past few years the research field of HRV was experiencing a rapid growth of interest, which yielded in the range of scientific studies. In a wide set of conducted researches, decreased HRV was associated with acute stress in the working environment [65, 66, 67]. As a result, we believe HRV to be a valuable data source for gaining physiological insights. According to our interviews, health-care professionals preferred to wear the heart rate monitor on their arm rather than on their wrists or chests, that influenced our choices in selecting the appropriate wearable sensor device.

3.1.3 Event annotations

Based on our pre-contextual study stages, we discovered that sharing the overall explanation of one's workday with our design team is a bottleneck within the real-time analysis process, as it forces us to wait for the participant's workday completion. In our post-hoc reflection section, we discussed how it is possible to recollect an overall impression of the day and use other additional data sources that one has access to. Nonetheless, there is still a number of relevant events that must be identified and recorded at the exact time of their occurrence. This approach would facilitate more accurate experience measuring as well as enable us to anticipate similar events in case of their re-occurrence.

Hence, our team decided to investigate the idea of keeping it simple in the contextual stage and mark the actual time of such events occurring during a participant's workday using a physical design probe (e.g. an event annotation button). Marking these episodes has two major advantages: (1) it can serve as an extra data stream for the participant when answering post-hoc reflection questionnaire, and (2) the researchers can use the annotated events in conjunction with the physiological signals. Both of these strategies lead to a stronger comprehension of the experiences happened throughout the workday.

3.2 Ecosystem architecture

Applying the Data-enabled Design (DED) approach enables designing prototypes for the intelligent ecosystems. Using these prototypes, a research team is ables to collect objective (i.e. physiological signals) and subjective data (i.e. post-hoc reflections and even annotations). By the term *intelligent ecosystems* we refer to dynamic clusters of interconnected products, services, and people. Data and artificial intelligence can be used by these products and services to develop a comprehensive and nuanced understanding of its context and user. Over time, intelligent ecosystems can build this understanding and adapt to personalized interactions [5].

Within this project, we attempted to translate our design decisions into micro-intelligences that were inspired by the micro-services architecture [68]. By *micro-intelligence* we consider an intelligent, self-contained layer that embeds all of the experience-related and functionality-related components required to collect, analyze, and visualize behavioral, contextual, and experimental data. The objectives of designing micro-intelligences rather than the overall intelligent ecosystem are as follows: (1) to enable the researchers reusing existing micro-intelligences that were developed within multiple DED studies, (2) to be able to separately assess experience-related and functionality-related aspects of intelligent ecosystem constituents, and (3) when included in the intelligent ecosystem, micro-intelligences will have varying degrees of automation.

In the scope of this research project, we designed two micro-intelligences for our future DED study: (1) the physiological layer and (2) the affect annotation layer. We were able to execute four asynchronous experiments with members of our research team, who were working from home.

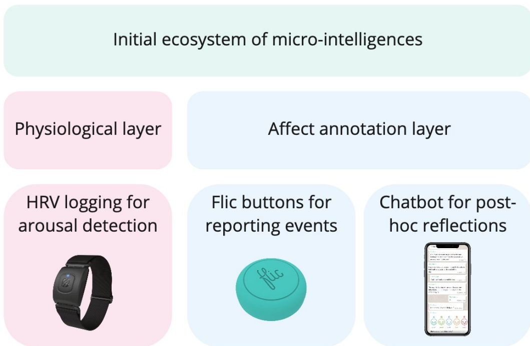

Figure 3.1: Ecosystem of micro-intelligences

3.2.1 Physiological layer

The physiological layer represents a micro-intelligence, that was developed by following our design decisions for the purpose of (1) facilitating the remote HRV data collection, and (2) enabling the post-hoc reflections on the participant's workday. The layer consists of the following components: (1) a Scosche Rhythm24 Heart Rate Monitor, (2) a Bluetooth Low Energy (BLE) enabled device (e.g. a computer or a smart phone), (3) software tools enabling HRV data extraction and analysis (Kubios HRV software [2], Elite HRV [69]), and (4) a messaging platform. The experience-related and the functionality-related capabilities of this micro-intelligence were tested by gathering data from two members of our research team while they were working from home.

Collecting physiological signals

In order to align with the requirements imposed by the HCPs, that were discovered during contextual research phase, we opted for a wearable heart rate monitor that can be worn on the arm of our participant. The Scosche Rhythm24 Heart Rate Monitor is a Bluetooth Low Energy optical HR band that can collect heart rate and inter-beat interval data using a Valencell PPG-based sensor that is supposed to provide reliable measurements. We performed this experiment setup for one person to determine its feasibility. Using the wearable HR monitor, data was collected throughout

[2]"Kubios HRV Analysis Software - Heart Rate Variability". Available at: https://www.kubios.com

a specified set of tasks across 6 unique intervals, each lasting 25 minutes and comprising of 5 repeated episodes, all of which included varying amounts of physical activity. The set of activities for recording within this pilot experiment was the same, as for the data collection described in Chapter 4. To be precise, the data collection was performed for the following recorded physical states: resting phase with deep breathing, training, cooking, cleaning, washing the dishes and watering the plants.

IBI time series data was extracted using an Elite HRV application, that was installed on the participant's mobile phone and enabled the real-time data acquisition from the sensor using the Bluetooth Low Energy connection. The data was sent in packages, containing information about timestamp, heart rate measurements and R-R intervals data. At the end of each recording session, data in a form of CSV files was automatically uploaded from the Elite HRV mobile application to the Elite HRV cloud storage.

Table 3.1: Data loss frequency evaluation during DED study

Dataset type	Frequency of data loss
Resting phase	4.5%
Watering the plants	27%
Washing the dishes	32%
Cooking	39%
Cleaning	45%
Training	49.8%

The results of the experimental data collection are shown in the Table 3.1. According to the illustrated statistics, the highest percentage of the missing data occurs during training activity (49.8%). This can be explained by the weak connectivity during data collection between the recording and receiving devices, which subsequently led to significant amount of data being lost. Nevertheless, findings outlined in the Chapter 4 suggest that even with high amount of missing data, we can still obtain quite accurate HRV results. Meanwhile, the lowest proportion of lost data appears for resting phase data acquisition (4.5%) when the individual is close to the signal receiving device and not engaged in any dynamic activity.

Overall, the presence of lost data as a matter of fact still highlights the limitation of relying solely on the HRV data. The HRV only indicates arousal, but not valence. Thus, in order to effectively interpret the experience and discern the context between the same HRV results, the emotional states deducted from the HRV must be connected to the ground truth in order to represent valence. This can be accomplished by incorporating extra input sources that allow for the collection of post-hoc reflection and event annotation data. Albeit, our micro-intelligence setup has the advantage of enabling us to continue this experiment with the goal of determining possible improvements (e.g., the location of the data receiving device) that would improve the quality of calculated HRV parameters while we continue our preparation for the contextual DED step [70, 71, 72].

3.2.2 Affect annotation layer

The affect annotation layer represents a micro-intelligence, that was developed by following our design decisions for the purpose of (1) facilitating event annotations, and (2) enabling the post-hoc reflections on the participant's workday. The layer comprises the following components: (1) two Flic Smart buttons, (2) a messaging platform that can be customized by Flow.ai, (3) a smartphone as a receiving data device with installed chatbot and Flic application, and (4) back-end supporting infrastructure for enabling event annotation analysis (e.g. deployed AWS Lambda function, custom Python script). The experience-related and the functionality-related capabilities of this micro-intelligence were tested by gathering data from three members of our research team while they were working from home.

Collecting post-hoc reflections

Home-context setting during the development of the DED contextual step as a micro-intelligence enabled our team to carry out the experiment focused on the hypothesis that is related to the experience-oriented aspect of our design decisions to post-hoc reflection's on one's workday. In the case of this experiment, the idea was to explore the feasibility of the chatbot as a tool for daily reflective journaling. Chatbot was developed and customized using the Flow.ai solution [3]. We created a WhatsApp group with our third participant and encouraged them to use it as a personal diary to record their reflections on working from home as a designer.

At the end of each workday, our design researcher delivered a list of generic working from home-related topics. The participant could select something from the list to reflect upon. This served as a prompt for them to provide us with their written post-hoc reflections. The participant was also asked to reflect on something not on the list. This experiment was carried out for seven days. Our key discovery was that this method of post-hoc reflection quickly became tedious and overly repetitive, resulting in less and fewer nuanced comments from our participants. As a result, we chose to direct our team's efforts toward further developing the affect annotation layer in order to address these deficiencies.

Collecting event annotations

Our discovered insight about the need to provide participants with an ability to report important events happening throughout the day at the exact time of their occurrence led to developing an event annotation functionality for them. Furthermore, by doing this we intended to avoid the repetitiveness and monotonicity of the post-hoc reflection process. Flic Smart buttons are Bluetooth Low Energy enabled IoT devices that can be used for this purpose [4]. The buttons are connected to the native Flic application running on a smartphone. Once the Flic Smart button is configured with an application, it enables sending a notification to the smart phone (or perform specific functionality) while being pressed. Flic native application allows to fully program the specific actions that would be triggered during physical pressing of the button.

For organizing this experiment, each of our second and fourth participants received two Flic Smart buttons. They were instructed to press the button during the course of their working day in the moments, when they experienced positive or negative events. Within the Flic app, we configured a POST request to send JSON data to our backend containing the timestamp and whether the positive or negative button was pushed. Our findings regarding integration of the Flic Smart buttons into the affect annotation layer alongside with chatbot combination will be discussed in Section 3.4.2.

3.3 Personalized analysis of the physiological data layer

In order to develop the analytical framework for identifying the change in the emotional state with regard to physiological sensing we utilized two open-source extensive datasets: CASE and WESAD. This decision was driven by the idea that the physiological data recorded during the experiment was not sufficient in volume to enable profound analysis. Hence, by utilizing external data sources we were able to realize a systematic methodology for physiology analysis that is integrated as a component into physiological layer of the ecosystem.

3.3.1 CASE dataset

The Continuously Annotated Signals of Emotion (CASE) dataset was developed as a part of the scientific project conducted by Sharma et al. to provide a data solution that focuses on real-time continuous annotation of emotions, as experienced by the participants, while watching various

[3] "Flow.ai—Chatbot design and management solution for professionals". Available at: https://flow.ai
[4] "Flic 2—The Smart Button for Lights, Music, Smart Home and More". Available at: https://flic.io

videos. The dataset includes data from multiple physiological sensors and continuous annotations of emotion. To address the issue that separate annotation of valence and arousal does not account for the inherent relationship between them, researchers developed a novel, intuitive joystick-based annotation interface that facilitated simultaneous recording of valence and arousal, Joystick-based Emotion Reporting Interface (JERI). The data was collected from 30 volunteers while they watched various video stimuli and reported their emotional responses using JERI. Electrocardiograph (ECG), Blood Volume Pulse (BVP), Galvanic Skin Response (GSR), Respiration (RSP), Skin Temperature (SKT), and Electromyography (EMG) sensors were used to collect physiological data [73].

Data collection protocol

Thirty volunteers from various cultural backgrounds participated in the CASE data collection experiment (15 males, age 28.6±4.8 years and 15 females, age 25.7±3.1 years; range of age 22–37 years). The "within subjects" design had been used to set up the experiment. As a result, repeated measurements were taken, and all participants watched and annotated different video-stimuli employed in the experiment. To eliminate carry-over effects, the order of the movies in a viewing session was changed in a pseudo-random manner, so that the resulting video sequence was unique for each participant. The different videos were interspersed with a two-minute long blue screen to isolate the emotional response evoked by them. This two-minute break also allowed participants to take a rest in between annotating the videos [73].

The participants were given an oral and written description of the experiment on the day of the experiment. Once all of the participants' questions about the experiment had been answered, they were asked to sign the informed consent form. Then, a quick introduction to the 2D circumplex model was given, and any questions concerning it were answered. Following that, physiological sensors were attached to the individual, who was then seated in front of a 42" flat-panel TV. The annotation process was then explained in detail. It was highlighted to the participants that they should annotate their emotional experience as a result of watching the videos, not the emotional content of the videos. After the preparation, the experiment was initiated and lasted for a duration of 40 minutes. The SUS questionnaire was used to collect feedback on the annotation system at the end of the experiment [73].

During the experiment, the participants were shown 20 short video-stimuli, that were aimed at eliciting *amusing*, *boring*, *relaxing* and *scary* emotional states. The first video was always a peaceful documentary extract intended to soothe participants before the emotional videos were shown. The end-video was created for the same reason, to serve as a "cool-down" interlude before the experiment's conclusion. The annotation was performed by the participants via the Joystick-based Emotion Reporting Interface, that allowed pointing and moving the red pointer in the appropriate region of the interface corresponding to specific level of arousal and valence [73].

As our research is focused on analyzing the heart rate variability, we focused on working only with the ECG-obtained data. Within the CASE experiment, ECG sensor was used to detect the electrical signal generated by the heart muscles during contraction. The performed procedure involved placing three electrodes in a triangle configuration on the participant's chest. First, the skin installation site was prepared by (1) removing any excess hair around the site if necessary, (2) abrading the skin using Nuprep abrasive cream, and (3) wiping the skin with an alcohol (70 percent isopropanol) pad. The cleaned locations were then covered with pre-gelled electrodes, with two resting on the right and left coracoid processes and the third on the xiphoid process. The detected electric signal is additionally pre-amplified and filtered by this sensor [73].

The full, transformed dataset contained processed data from 30 participants. Columns within every participant dataset contained information gathered through physiological sensors (i.e. raw ECG) as well as annotation labels and video-stimuli labels.

3.3.2 WESAD dataset

WESAD is a multimodal dataset for Wearable Stress and Affect Detection, that was developed within the scope of research by Schmidt et al. The dataset contains physiological and motion data, recorded of 15 subjects during a lab study from different multimodal sensors. The data was collected utilizing two separate devices (one chest-based and one wrist-based), each of them including high-resolution physiological (BVP, ECG, EDA, EMG, RESP, and TEMP) and motion (ACC) modalities. By having three separate emotional states, the dataset bridges the gap between prior lab studies on stress and emotions (neutral, stress, amusement). Furthermore, the dataset includes self-reported values on the respondents' perceived affective state, which were gathered using a variety of standard questionnaires. These self-reports can help train personalised classifiers. In addition, a benchmark was created using a large amount of well-known features (extracted from physiological and motion signals) [74].

Data collection protocol

In total, 17 people took part in WESAD research. The data of two individuals had to be discarded due to the sensor malfunctioning. The remaining 15 individuals were 27.5±2.4 years old on average. Twelve individuals were male, while the remaining three were female [74].

Data acquisition was carried out using both a chest-worn device and a wrist-worn device: a RespiBAN Professional and an Empatica E4, respectively. The RespiBAN is equipped with sensors that measure ACC and RESP, and it can also serve as a hub for up to four additional modalities. ECG, EDA, EMG, and TEMP were recorded using the four analog ports. All signals were sampled at 700 Hz. The subject's chest was used to place around the RespiBan device. A respiratory inductive plethysmograph sensor was used to record the RESP. A standard three-point ECG was used to get the ECG data. On both sides of the spine, EMG data was recorded from the upper trapezius muscle. To avoid wireless packet loss, the collected data was saved locally and uploaded to a computer after the experiment for further analysis. The Empatica E4 was worn by all individuals on their non-dominant hand. BVP (64 Hz), EDA (4 Hz), TEMP (4 Hz), and ACC were all recorded by the E4 (32 Hz) [74].

The study's purpose was to induce three different affective states in the subjects (neutral, stress, and amusement). In addition, after the stress and amusement conditions, the subjects were requested to follow a guided meditation. Once the participants arrived at the study site, they were provided with the sensors and subjected to a brief sensor test. The RespiBAN and E4 were then manually synchronized using a double tap gesture [74].

The *baseline condition* was recorded for 20 minutes straight from the beginning of the experiment. The goal of this part of the study protocol was to induce a neutral affective state. During the baseline, the individuals were sitting at a table and were supplied with neutral reading material (magazines). Afterwards, the *amusement condition* part was carried out. The patients watched a series of eleven entertaining video clips during the amusement condition. Each clip was followed by a five-second neutral segment. The amusement condition lasted 392 seconds in total. During the *stress condition* phase, the subjects were given the well-researched Trier Social Stress Test (TSST) [75], which comprises of a public speaking and a mental arithmetic task. These activities are known to consistently produce stress since they are social evaluative and demand a high mental load on the subjects. In the WESAD study version of the TSST, participants had to first give a five-minute speech about their personal qualities in front of a three-person panel, focusing on their strengths and weaknesses. Following the speech, the committee instructed the audience to count from 2023 to zero in 17-step increments.

Overall, the stress condition part lasted for 10 minutes in total. The amusement and stress conditions, both of which were designed to stimulate the subjects' excitement, were followed by a guided *meditation condition* phase. The goal of this meditation was to 'de-excite' the individuals and return them to a near-neutral affective state. The meditation was built around a controlled breathing exercise that was guided by an audio track. The meditation lasted seven minutes. At the end of the protocol, the sensors were synchronized once again with a double tap gesture and then completely removed from the participants. The experiment lasted approximately two hours in total [74].

The fully synchronized WESAD dataset consisted of physiological and affective annotation data from 15 participants. In the scope of our research we utilized only data that was corresponding to raw ECG signals and annotated labels that were referring to every protocol step of the experiment [74].

3.3.3 Methodology for personalized analysis of the HRV data

Data pre-processing for the CASE dataset

As was discussed previously, data for every participant within the CASE dataset resides in a separate CSV file. Hence, for creating a consistent, pre-processed dataset we firstly extracted raw ECG values and affect annotation labels for every participant and stored them in the intermediate dataframe. Resulting dataset contained information for all original 30 participants, but later on was reduced to 29 individuals, as the ECG-recorded values for participant 24 comprised high presence of noise that made it impossible to obtain accurate HRV parameters, thus it was removed.

As a next step, all rows corresponding to video label values outside the range of [1; 8] were eliminated as they belonged to experimental recordings with non-labelled videos, according to the CASE dataset description [57].

The raw ECG physiological data was partitioned into chunks per video segment and used to detect and remove artifacts, caused by movements. Afterwards, using the Python *heartpy* package [2], the R-peaks were detected and thus R-R intervals were derived, which enabled later HRV parameter generation. Furthermore, for the general exploratory analysis of the HRV values distribution we created an additional separate column with binary labels corresponding to stress and no stress conditions by grouping the original video-stimule labels in a following way: *"No stress"* related to relaxing and boring conditions, while *"Stress"* label corresponded to amusing and scary affective states.

Data pre-processing for the WESAD dataset

WESAD dataset consists of separate files in a semi-structured Pickle format corresponding to each participant's data. Similarly, chest ECG-obtained values and annotated labels were extracted from every subject dictionary and merged into one coherent dataframe. After performing quick data examination, it was revealed that the data related to participant 2 is not consistent and contains high amount of noise, hence those records were removed during pre-processing phase.

In a similar manner, rows having annotation labels outside the range [1; 4] were removed, as they corresponded to intermediate recovery phases of the experiment and did not possess a clear definition of the related affective condition.

Intermediate dataframe was split into separate parts based on the experiment protocol part (e.g. baseline, etc) and processed with a peak detection mechanism from the *heartpy* library [2], that resulted in computing R-peak locations and R-R intervals. Afterwards, data was merged back into one solid dataset that contained information about inter-beat intervals values for all 14 participants and respective condition annotated labels. As for the CASE dataset, an extra column was artificially generated and contained data regarding binary labels of the original affect annotation labels. The grouping was performed in a following way: *"No stress"* corresponding to baseline and meditation conditions, and *"Stress"* label related to amusement and stress emotional states.

HRV parameters generation and analysis

To proceed with the personalized HRV analysis, we utilized previously pre-processed data from CASE and WESAD datasets to compute HRV parameters. According to the meta-analysis on HRV [76], we focused on deriving a limited set of HRV values, specifically: *Mean RRI*, *rMSSD*, *SDNN*, *pNN50* and *LF/HF ratio*, as they were proven to most significantly reflect the physiological changes during induced stress. HRV values were computed from 5-min time-windows. For this purpose, we utilized *hrv-analysis* Python package [5] to compute time-domain HRV features and *Kubios HRV* software for generating frequency-domain HRV metrics using spectral analysis. As a result of HRV generation procedure, we obtained cleaned, transformed and prepared for analysis dataset from 43 participants in total.

As a first step, the general exploratory analysis was carried out by visualizing the distribution of every generated HRV parameter for all participants with regard to the affect annotated group (i.e. "Stress" and "No Stress") for combined datasets, and for CASE and WESAD separately. In order to understand the nature of HRV values change during different emotional conditions, multiple line graphs were generated per every participant separately for CASE and WESAD datasets.

As a following steps, rMSSD and Mean RRI HRV features were used to statistically estimate the *percentage change (PC)* as a ratio of difference between two consecutive HRV values representing previous and current conditions (e.g. baseline and amusement) and an HRV value of the previous condition:

$$PC(\%) = \frac{X_j - X_i}{X_i} \times 100\% \tag{3.1}$$

where X_i represents the old HRV value (i.e. from the previous condition), and X_j denotes the new HRV values (i.e. from the current condition). In case the percentage change result is positive, the values are considered to be increased; otherwise, it represents a decrease.

Resulting percentage change values for rMSSD and Mean RRI were analyzed across all participants and used to identify the trend in HRV dynamics.

[5] R. Champseix, "hrv-analysis: a package to calculate features from RR Interval for HRV analysis". Available at: https://github.com/Aura-healthcare/hrv-analysis

3.3.4 Results of the personalized analysis of the HRV data

Combined data from the pre-processed CASE and WESAD datasets enabled us to perform the general exploratory analysis of the HRV parameters across 43 participants. Violin plots were used to depict the probability density of the data during stress and no-stress conditions, smoothed by kernel density estimator. The box plots integrated within every violin plot illustrate the Interquartile Range (IQR) and mean of the distribution for each particular variable.

Firstly, the violin plots were generated for CASE and WESAD datasets separately (see Fig. 3.2 and A.1). There were noticed some differences in between the distribution of values in WESAD dataset for rMSSD and Mean RRI HRV parameters, although the Interquartile Range was overlapping for both stress and no-stress groups.

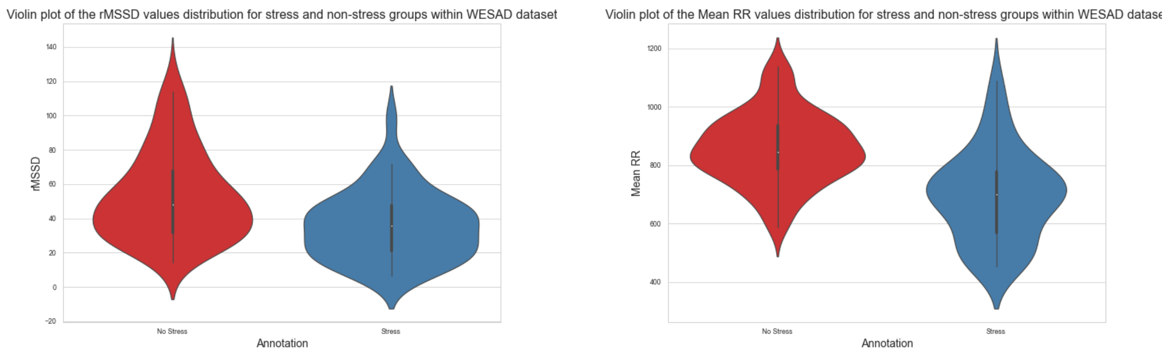

Figure 3.2: “Violin” plots of the distribution of the selected HRV feature values within binary groups of participants from WESAD dataset

As for the visualized density probabilities for the CASE dataset data (see Fig. 3.3 and A.2), we were not able to detect any substantial differences within distributions due to the fact that all HRV features data corresponding to stress and no-stress groups was having the same Interquartile Range and distribution type.

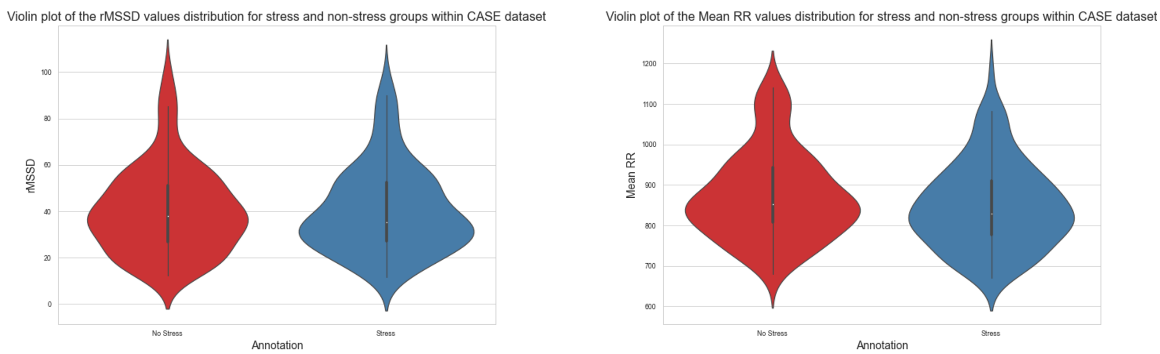

Figure 3.3: “Violin” plots of the distribution of the selected HRV feature values within binary groups of participants from CASE dataset

Afterwards, the violin plots for the merged datasets were visualized (see Fig. 3.4 and A.3). As we can observe, the distribution of values for every HRV variable between stress and no-stress groups has a high level of affinity, so that the values residing between 25th and 75th percentiles are always overlapping. It was noticed that when separated, violin plots depicting distributions of data within WESAD dataset were illustrating a clear differences between stress and no-stress groups, but when combining the aforementioned datasets these differences are lost. Taking this into account, we conclude that generalized group analysis did not prove to be efficient, as there is a need to take into account inter-personal differences in HRV values between the transitions from one affective condition to another.

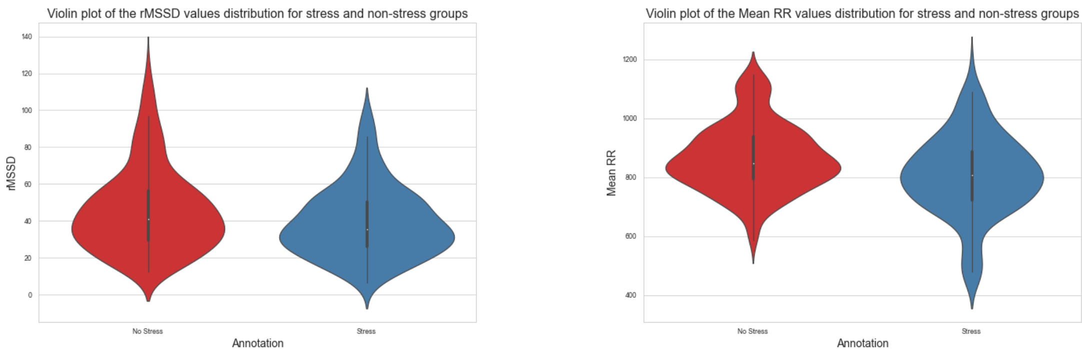

Figure 3.4: "Violin" plots of the distribution of the selected HRV feature values within binary groups of participants from merged datasets

Table 3.2: HRV expected change dynamics under stressful conditions

HRV parameter	Expected behaviour change
rMSSD	decreases under stress
SDNN	decreases under stress
Mean R-R Interval	decreases under stress
pNN50	decreases under stress
LF/HF ratio	increases under stress
Heart Rate	increases under stress

The idea behind generating the multi-line graph was to determine the common pattern in the HRV values change that is observed across all participants. The research conducted by Kim et al. (2018) [76] suggested expected behaviour of different HRV parameters according to the performed meta-analysis. The excerpt of this analysis in outlined in Table 3.2. Thus, the expected visualizations for all analyzed HRV parameters, except LF/HF ratio and Heart Rate, should depict the considerable decrease in values once the participants were enduring stress conditions.

To verify that, the multi-line plots for every participant within CASE dataset and different HRV parameters were visualized, and are illustrated by Fig. 3.5.

Plots in Fig. 3.5 show the changing trend in particular HRV parameter for each subject while shifting from relaxing condition to boring. Although there are clearly depicted changes in the HRV, especially for SDNN and Mean RRI, we were unable to identify any common pattern for all of the participants. For the majority of subjects, there was not noticed any significant decrease in HRV during scary or amusing experiment phases. In addition, based on these plots we cannot conclude which HRV features represent the highest importance for emotional state analysis.

We can conclude that for the purpose of our study CASE dataset turned out to be not useful in providing meaningful and non-ambiguous insights regarding personalized analysis of the HRV. Hence, we will only focus on the WESAD data in the subsequent steps of our analysis. However, the obtained results from visual analytics were conformed to align with the results of the researches within CASE study [57]. Regarding the ambiguity of the insights, we can consider multiple reasons behind it, e.g. the physical activities performed by the participants of the experiment were not sufficient for affecting the emotional state, the inertia of the body after being induced with different stress stimuli was longer than the duration of the experiment phases, or simply the changes in emotional state cannot be accurately reflected by using HRV as a proxy for physiological data source. In general, there is much more in the context of the experience the person is undergoing that can actually influence the emotional state shifts, hence it needs to be further researched in a more profound study.

As a next step, we generated new multi-line graphs for WESAD dataset to understand how real-case HRV values behaviour differs from the expected one from Table 3.2. Visualizations for inter-personal change in different HRV parameters are depicted in Fig. 3.6.

The new visualized plots based on the WESAD data look significantly more promising. We can see that all rMSSD, SDNN, pNN50 and Mean RRI graphs show the considerable decline in HRV values after the baseline condition for the majority of participants, while the LF/HF ratio plot depicts a values increase during the same phase. Furthermore, we can notice that across all time-domain HRV plots the values substantially plummeted not during the actual stress phase, but earlier during the amusement condition. Moreover, during the induced stress period, the HRV values started to actually increase and continued in the same manner during meditation. This could lead to a conclusion that sympathetic branch of the autonomic nervous system is dominating more during the moments of expected potential stress and hence adjusts the modulation of the heart in a respective way. As for the LF/HF ratio graph, the HRV values for almost all of the participants experienced its peak during the actual induced stress condition.

Further analyzing the HRV behaviour plots, we observe that the baseline values are always different for every person (see Fig. 3.6). What's more, for some participants the baseline values are considerably low from the beginning, and are actually lower than the most decreased bottom values for other participants during the amusement phase. This provides an insight that for the inter-personal analysis of the HRV parameters there is a need to focus on the percentage change that can be quite common across different people, rather than absolute values. Bearing this in mind, we proceeded with statistical estimation of percentage change for HRV metrics.

For this purpose we decided to focus on the HRV values of rMSSD and Mean RRI, as according to the plots they showed to be the most prominent in precisely depicting the HRV changes. Percentage change ratios for the rMSSD HRV values are outlined in Table A.1 (see Appendix).

According to the table, the majority of subjects experienced highest decrease in rMSSD during the *baseline to amusement* transition of conditions, which is reflected by the negative percentage change. The largest decrease was observed in case of participant 14, where the percentage change for this shift was equal to -86.87%. The percentage change for transitioning from *amusement to stress* phase was varying across participants. Specifically, for participants 5, 8 and 13 there was noticed a decline in rMSSD, which signifies that the sympathetic branch of their nervous system started to dominate during the induced stress experiment phase, instead of amusement phase. For all the rest of the participants, the transition from amusement to stress resulted in increased rMSSD, where the highest change in percentage ratios was indicated in participant 10 (330.591%). As for the final, stress to meditation transition, all the participants experienced a rise in HRV, which was expected to happen, as during the peaceful and tranquil activities the autonomic nervous system shifts to the "rest and digest" mode, which culminates in increased HRV and decreased HR. The biggest transition ratio for rMSSD was noticed in the case of participant 8 (145.456%).

As for the percentage change in Mean RRI values, we can observe in Table A.2 (see Appendix) that the trend is similar with regard to rMSSD values. During the first transition from *baseline to amusement*, all participants had their Mean RRI HRV values decreased, which more accurately describes the overall HRV behaviour throughout endured stress. The biggest decline was noticed also as in the case of rMSSD comparison in participant 14 and resulted in -47.175% change. During the shifting phase from *amusement to stress*, all subjects aside from participant 8 experienced an increase in Mean RRI indicators, that confirms with the previously discussed hypothesis that most severe physiological reaction with regard to stress happens in the moment of stress anticipation, not the actual stress occasion. Finally, for the last shift in conditions from stress to meditation for the majority of participants (except for 3rd, 5th, 6th and 10th) there was observed a slight increase in Mean RRI, the lowest one being 0.552% and the highest equaling to 19.203%. The absence of drastic change in physiological state from stress to meditation can be interpreted by the weak ability of the autonomic nervous system to quickly adapt to a new tranquil state after experiencing severe stress.

Overall, performing an analysis of the Heart Rate Variability data using multiple datasets led to ambiguous results.

On the one hand, by utilizing CASE dataset we were able to conclude that HRV cannot always be used as a proxy for representing the physiological changes and deriving the insights regarding the emotional state.

Contrary, by exploring the WESAD dataset we were able to confirm that indeed HRV is reflecting the changes in physiology quite accurately and enables to understand the common trends when it comes to inter-personal analysis of the affective conditions. Hence, those conclusions support the idea that Heart Rate Variability can still be used to reflect the emotional state when enduring different acute stimuli.

The necessity of performing only statistical analysis of the Heart Rate Variability data is reasoned by the introduced limitations of the MEX Oncology project that was still in the early stages of development. Hence, we were unable to train any comprehensive Machine Learning model for emotional state recognition since that data collection in the real-life setting was not yet organized. Consequently, in order to run the pilots for evaluating the feasibility of the HRV analysis the aforementioned analytical approach was introduced.

The key outtakes derived from executing the steps of this method are: (1) utilizing multiple unsynchronized datasets for personalized analysis of the HRV data is not feasible, as they do not produce coherent insights when being combined (2) inter-personal physiological differences during artificially induced stimuli significantly affect the possibility of the generalized analysis and reusability of the trained model (3) however, a set of common trends is still noticeable within Heart Rate Variability data for the majority of the analyzed participants.

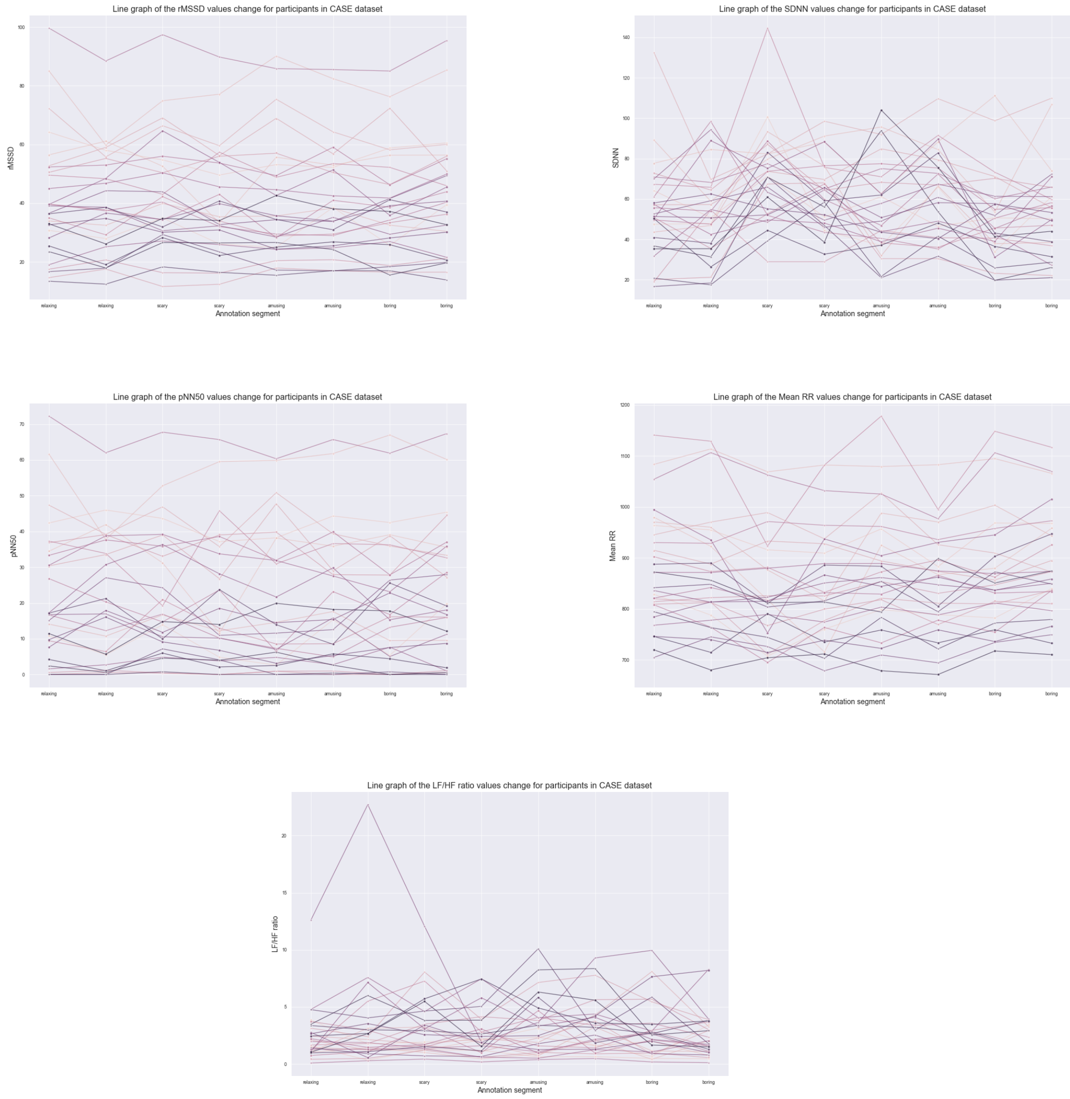

Figure 3.5: Multi-line graphs for selected HRV feature values during sequential artificially induced conditions across all participants from CASE dataset

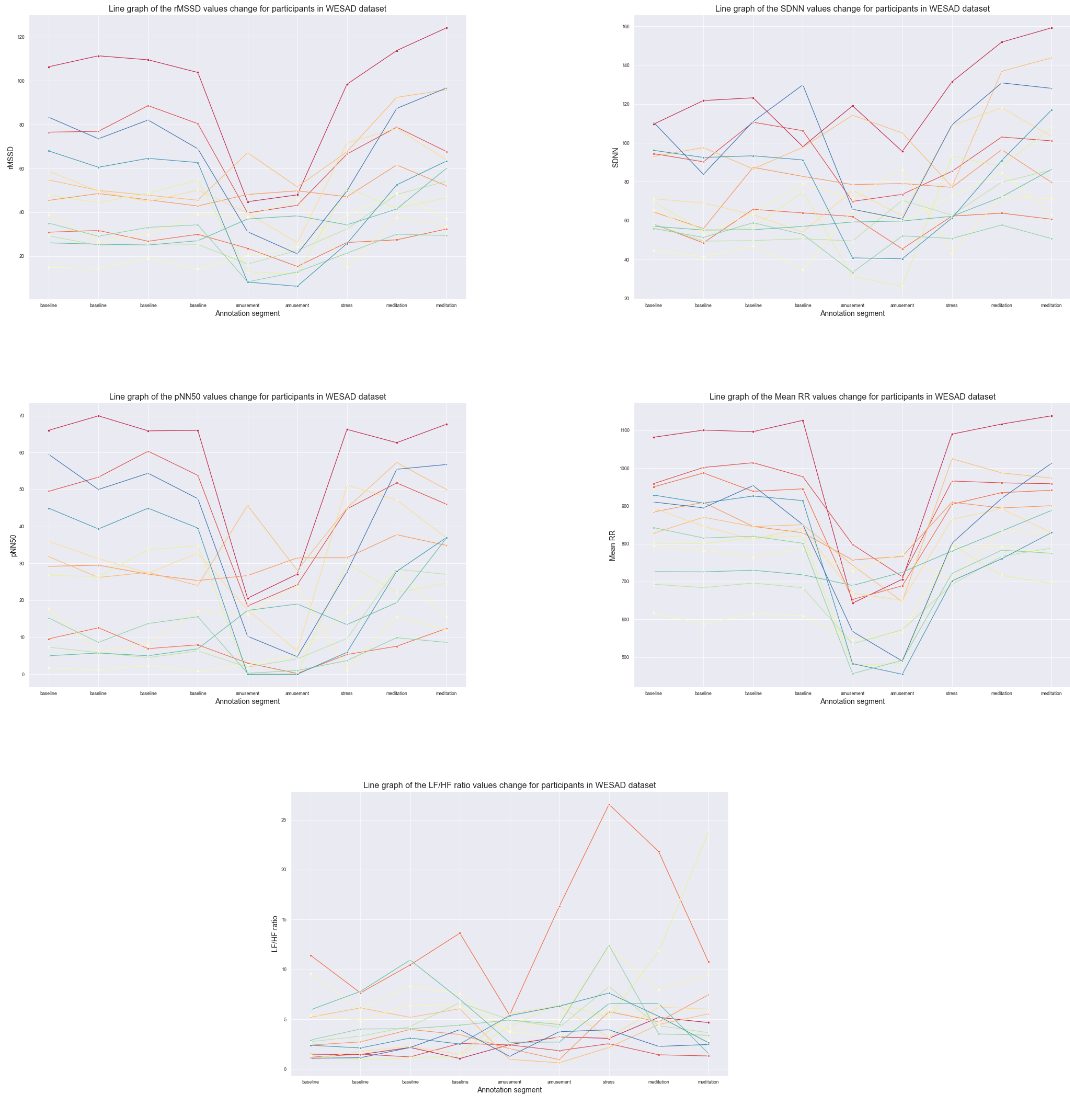

Figure 3.6: Multi-line graphs for selected HRV feature values during sequential artificially induced conditions across all participants from WESAD dataset

3.4 Experimental use cases

3.4.1 Combining post-hoc reflections with physiological signals

As was discussed earlier, there is a need to introduce other input sources combined with physiological signals recorded via Scosche Rhythm24 to discern the context and link the insights with the ground-truth by enabling the post-hoc reflections capability. Our hypothesis was that incorporating the ability to ask participants about their physiological signals would improve our contextual insights. We advised our second participant to use the same Scosche heart rate monitor and the Elite HRV mobile app to collect data for this experiment [69]. Similarly, we used Kubios HRV software for HRV generation and analysis purposes. Utilizing findings on personalized HRV analysis from the Section 3.3.3, we used the same strategy of identifying the physiological states transitioning by computing the percentage change ratios. HRV interval values, corresponding to equivalent ranges of condition shifts from stress to no-stress and vice-a-versa were marked as "moments of interest". Hence, at the end of each working day, the researcher created a visualization of the participant's workday moments of interest, alongside with questions that asked to reflect on them, and shared it via WhatsApp. The subject was able to respond and send the researcher post-hoc reflection data (see Fig. 3.7).

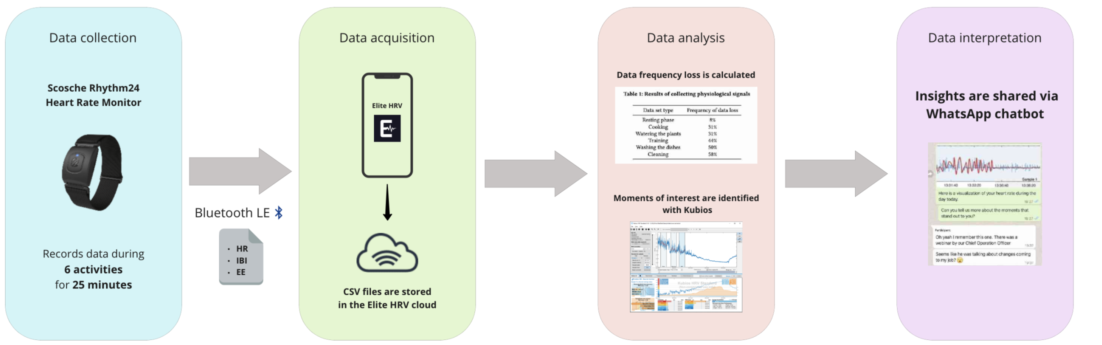

Figure 3.7: The dataflow enabled by combining physiological signals recorded via Scosche Heart Monitor with post-hoc reflections collected via chatbot

The study was carried out during the ten days period, while the subject worked from home. We instructed the participant to collect as much HRV data as possible using the application and to respond to the post-hoc reflection questions immediately at the end of their workday.

At the end of the experiment, we conducted a data-enabled interview with the subject and discussed our key findings. Based on the data obtained and the insights gained through the interview, it was evident that our assumption on participant being able to reflect on their day by looking at their daily visualized *biofeedback* was confirmed. The subject reported that the sensor's data was an excellent predictor of when their Heart Rate Variability deviated from their baseline and clearly depicted the physiological state change. It was feasible to inform our research team about what happened during those moments while the participant was checking his personalized visual biofeedback. Multiple times the participant could not fully recall what happened during the highlighted time intervals of their workday, but in this case, they checked with their work calendar. We also discovered a substantial experience-related insight, specifically that monitoring the Heart Rate Variability caused participants to pay more attention to the stressful or particularly negative events of their day when submitting us post-hoc observations. This discovery has allowed our team to more precisely scope the affect annotation layer of our DED experiment.

3.4.2 Combining post-hoc reflections with event annotations

To integrate post-hoc reflections with event annotations to create one solid and coherent affect annotation layer of the ecosystem, the Flow.ai chatbot was configured to trigger a WhatsApp message at 17.00 after the end of the participants' workday. With this notification, the subjects of the study were requested to evaluate their overall experience throughout the working day via short visual questionnaire. With regard to their response, the chatbot was replying with a message consisting of information about the time-intervals of their positive and negative moments that were logged by two Flic Smart buttons. Finally, the participants were asked to describe their day in detail (see Fig. 3.8).

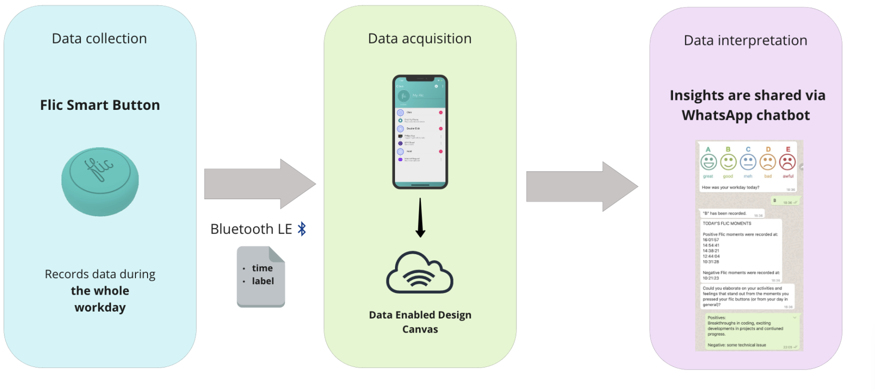

Figure 3.8: The dataflow enabled by combining post-hoc reflections collected via chatbot with experienced events annotated via Flic Smart button

We conducted this experiment over the course of seven days, with two participants working from home. Following the experiment, we organized data-enabled design interviews in which participants were shown their collected data, which allowed them to reflect on it.

3.4.3 Combining all constituents of the eco-system

Within the scope of this Chapter we introduced the concept of micro-intelligence and designed an ecosystem of micro-intelligences. The benefit behind utilizing micro-intelligence is enabled reusability facilitated by the micro-intelligence property of being a self-contained layer of a broader intelligent ecosystem. It was demonstrated in the Chapter by showing that the four experiments incorporating the two micro-intelligences all use common components and infrastructure, and are designed in a way to allow them being reused by other research teams.

Furthermore, using the designed micro-intelligences enables modular evaluation of the intelligent ecosystem components. We were capable of guiding whether the research would focus on testing the experience-related or the functionality-related feature of a component in our experiments. We assessed our design decision to obtain post-hoc reflections via chatbot in four ways (see Fig. 3.9).

As a first step, we used a standalone chatbot component with a purpose of learning about the experience-related aspects regarding post-hoc reflection data collection. The aim of this experiment was to explore experiential aspects of a chatbot like dialogue, personality and timing, and to define its scripting rules. After conducting this experiment we realized that using only chatbot is not sufficient to provide a coherent insights about the personal experience as it is lacking objective data perspective.

Hence, it was decided to combine the chatbot with Heart Rate Variability data (i.e. Scosche heart rate monitor). This allowed us to dedicate our focus on both functionality (can participants submit post-hoc reflections based on their HRV data) and experience-related (what visualizations to provide with chatbot messages) aspects. Moreover, we were also able to test in a separate

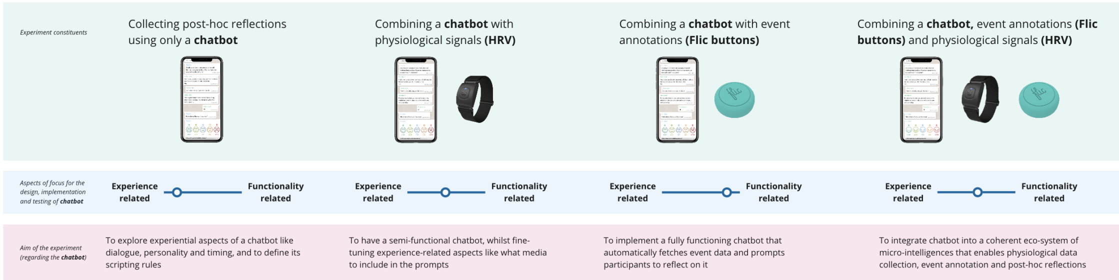

Figure 3.9: Sequential flow of testing the experience-related and functionality-related aspects of elements within the intelligent ecosystem

manner the functionality-related aspect on connecting our chatbot to the affect-annotation layer that enables event annotations. The goal of this experimental step was to have a semi-functional chatbot, whilst fine-tuning experience-related aspects like what media to include in the prompts.

As a next step, we focused on functionality-related aspect of the design, implementation and testing of the chatbot. Thus, the combination of post-hoc reflections via chatbot with event annotations enabled by Flic Smart buttons allowed us to cover all the aspects related to subjective data collection, create a full, coherent affect annotation layer of the ecosystem and bring the ground-truth insights regarding the participant workday experience. The core aim of this testing phase was to implement a fully functioning chatbot that automatically fetches event data and prompts participants to reflect on it.

Finally, the last step of the DED experiment testing process was to combine the autonomous micro-intelligence layers into one intelligent ecosystem, that can be used to collect, analyze and visualize behavioral, contextual, and experimental data from the participants. The designed fully-integrated ecosystem of micro-intelligences will enable performing an emotional state and physiology analysis with regard to the participant's experience by combining the physiological sensing represented by the Heart Rate Variability with self-reflecting annotations facilitated by Flic Smart buttons to annotate positive and negative moments of the workday and a chatbot for post-hoc reflections. The collected physiological signals will be used to correlate the reported moments with objective physiological measurements and understand if there is a compliance between them. The final step after performing this analysis is to collect post-hoc reflections using a chatbot at the end of the working day to fully evaluate the personal experience.

Overall, by designing these micro-intelligences we aimed at testing the required level of automation corresponding to each modular component of the intelligent ecosystem. Throughout our four conducted experiments that were discussed in this Chapter, we were able to flexibly decide whether to employ the components as fully-automated data trackers (a chatbot that can automatically retrieve data from the event annotation component) or in a Wizard of Oz prototyping style (the design researcher manually generated and sent out the HRV graphs to the participant). The experiments we presented allowed us to fine-tune our design decisions while also revealing experience or functional shortcomings in our micro-intelligences. By doing this, we attempted to take a first step in designing a comprehensive intelligent ecosystem for conducting affective computing in a situated manner. It was concluded that the best promising insights that can be captured within the situated affective computing setting occur when combining *all constituents* of the ecosystem into one coherent framework.

As a follow-up in the next chapters we are aiming to investigate the reliability and the accuracy of the device more extensively, to understand the feasibility of employing it as a part of physiological data layer.

Chapter 4

Estimation of the missing signal impact for computing HRV

Due to the increased popularity of wearable devices equipped with heart rate sensors, HRV monitoring and analysis became more accessible to the general public. The main problem of extracting HRV parameters from the data collected by armband devices is that the recording is often influenced by motion artifacts and systematic errors due to the signal loss. Hence, error estimation caused by the significant amount of missing data is essential in order to retrieve reliable HRV parameters from PPG-based sensors. The process of estimating the impact of missing values on the accuracy of HRV parameters often involves investigating the amount and key characteristics of missing data, performing synthetic generation of missing values, deriving the HRV parameters from the data sets with imputed missing data and comparing their deviation from each other through the accuracy metrics. Due to the fact that different approaches of missing data generation can significantly affect the validity of the derived conclusions, the established setup of missing data generation and mechanisms behind it must be thoroughly researched to prevent the invalid and biased results.

The designed micro-intelligences for affective computing in the previous Chapter 3 were aimed to facilitate the accurate data acquisition in a hospital setting. The standardized approach in that scenario would be to use the ECG sensors that are qualified to bring the most reliable HRV measurements. However, the limitation of using ECG devices relies in a fact that they are too obtrusive, which would create issues for the healthcare practitioners to perform their duties. Hence, the assessed PPG sensor is considered to be integrated into the ecosystem of micro-intelligences in the subsequent experiments. As various types of HCPs perform different kinds of activities, that lead to different way of moving, our goal is to provide a method for estimating in advance what sorts of occupations can be recorded accurately and which of them cause a huge data loss. To understand that, we conduced a study involving the data acquisition process within short period of time while performing several dynamic activities.

Methods for evaluating the missing data from the PPG-obtained recording and its impact on HRV have been introduced by multiple researchers with the aim of defining the trade-off between the amount of lost data and reliable results. Baek et al. (2017) explored the effect of missing heart beat interval data on the HRV analysis with a simulation methodology that implied random amputation of the continuously recorded data [70]. A study conducted by Sheridan et al. (2020) focused on determining the impact of small errors on the accuracy of HRV by removing up to 36 % of beat-to-beat time intervals in a random or consecutive manner [71]. In the research [72], the signal loss episodes were artificially inserted in different patterns and percentages to the cardiotocographic recordings in order to evaluate the sensitivity of HRV parameters with regard to lost data. Yu at el. (2015) research was aimed at detecting complexity of HRV by generating synthetic signals that randomly erased data from the original dataset according to a Gaussian distribution and an exponential distribution [77]. In a study by Kim et al. (2007) the effects of consecutive missing RR-interval data on time-domain HRV analysis were investigated by removing randomly selected RR-intervals with the duration from 0 to 100 seconds, and afterwards compared to the actual missing RR-interval data from ECG measurements [78].

In this Chapter we aim to systematically investigate the patterns of missing data and evaluate its effect on the accuracy of computed HRV parameters. To the extent of our knowledge, there does not exist any systematic research focused on assessing the derived HRV parameters after introducing several complex missing data patterns that were identified through the real-time context-specific data collection. Hence, this chapter mainly focuses on addressing the problem of how much missing data and specific patterns of it can affect the reliability and validity of HRV measurements.

In the Section 4.1 we describe the main approach towards the data acquisition within this experiment. Section 4.2 illustrates the preliminary steps conducted to pre-process data before the data exploration phase, namely cleaning the data, transforming the data types and generating a set of new datasets consisting of values corresponding to empty and non-empty recording periods. Within Section 4.3 we explore the collected datasets with the purpose of estimating the total amount of missing data, conditional data loss probabilities and employing visual analytics to understand the circumstances, in which data was missing. In Section 4.4 the missing data mechanisms are explained. Section 4.5 describes the methodology for identifying the missing patterns within collected datasets, specifically what statistical tests are used to detect the patterns and how to assess the goodness of fit using visual analytics as well as outlines the results of this methodology application. In Section 4.6 we unravel the approach towards missing data patterns generation on the open-source dataset to perform the error estimation analysis of the derived HRV values. Sections 4.7 and 4.8 describe the actual HRV parameters analysis methodology that aims at evaluating the accuracy of the HRV features computed from noisy datasets and defining the threshold, that represents an extent beyond which we can consider the HRV parameters to accurately characterize the human physiological condition. Finally, in Section 4.9 we illustrate the use case example of the repeated experiment following the proposed methodology of missing data pattern identification.

4.1 Data collection and extraction

In order to determine the typical patterns of missing data, the data collection process was carried out. Data was recorded during six separate intervals throughout the day, each lasting 25 minutes and consisting of 5 repeated episodes. In this experiment we used only IBI data and corresponding timestamps continuously recorded within the same interval by a Scosche Rhythm24 Heart Rate Monitor [79]. The device is a Bluetooth Low Energy armband equipped with a Valencell PPG-based sensor that was previously validated with respect to ECG to provide moderately accurate inter-beat intervals data [80].

During the data acquisition process, data was collected during a predefined set of tasks that were all including physical activity with different levels of intensity. The idea behind was to record the data that is noisy and contains a lot of missing values due to the motion artifacts and signal loss issues. Specifically, the data collection was performed for the following recorded physical states: resting phase with deep breathing, training, cooking, cleaning, washing the dishes and watering the plants. IBI time series data was extracted using a custom Node.js script that was running on the laptop and enabled the real-time data acquisition from the sensor using the Bluetooth Low Energy connection. The data was sent in packages, containing several characteristics (e.g. Heart Rate Measurement value, Energy Expended, R-R intervals, etc) varying in size from 20 bytes to 236 bytes with a maximum data transfer rate from 222 bytes/sec to 2186 bytes/sec [81]. At the end of each recording session the collected, highly-structured data was stored in a file with a CSV format.

4.2 Initial data pre-processing

First of all, the extracted data sets were preliminarily analyzed before applying any pre-processing steps. It was revealed that the data was only recorded at specific timestamps when the signal was of a decent quality. Therefore, by using the Python *pandas* library [1], the data was rescaled to include every timestamp at the granularity of seconds. For the seconds that were missing any recorded IBI value, NaN values were imputed at corresponding locations. It is also worth mentioning that, as the first part of the aforementioned problem this chapter is targeting is related to missing patterns identification, the initial pre-processing is not aiming to detect and remove the ectopic peaks in the R-R intervals or reconstruct the missing values using interpolation techniques. Hence, the additional step for properly carrying out the subsequent analysis was to generate for each subset a corresponding data series that consisted of aggregated continuous occurrences of missing data intervals. Specifically, while iterating over the dataset the length of each gap including several missing values was counted as an integer value and written to the file. As a result, every collected dataset had a matching dataset with a continuous distribution of accumulated length related to missing data gaps.

4.3 Data exploration

As a first step to identify the missing patterns and understand the specific context in which the collected data has a higher percentage of loss, we estimated the overall frequency of missing values in each of the datasets corresponding to specific activity (Table 4.1).

Table 4.1: Estimation of the overall data loss frequency

Dataset type	Frequency of data loss
Resting phase	8%
Cooking	31%
Watering the plants	31%
Training	44%
Washing the dishes	50%
Cleaning	58%

Analyzing the obtained results of data exploration in Table 4.1, we can conclude that the highest percentage of missing data occurs while performing the cleaning activity (57 %). That can be explained by the significant amount of movement during the data collection process that affects the quality of the signal and subsequently leads to data loss. In the meantime, the lowest percentage of losing the data appears for resting phase data collection (8.04 %) while the participant is located closely to the signal perceiving device and is not involved in any dynamic activities.

Afterwards, datasets were analyzed on the subject of conditional probabilities of data loss, in case the preceding recorded observation was missing or not. It is common that the lost signal occurrence leads to multiple consecutive missing values. Hence, the idea behind this analysis was to understand to which extent the existence of previous observation triggers the missing data mechanism. In Table 4.2 the results of the performed analysis are outlined. The key tendency is following the previous data exploration of total missing values resulting in the dataset of cleaning activity to be affected by the highest missing rate of values in both cases when the previous values are lost and present. Same trend captures the resting phase activity with having the lowest probabilities of data loss. Moreover, the outlined ratios illustrate the stable growth of the frequencies, meaning that for every collected dataset the proportion between frequencies of data loss will be constant. Overall, by making the cross comparison of both conditional probabilities

[1] "pandas: Powerful data structures for data analysis, time series, and statistics". Available at: https://pandas.pydata.org

for every dataset we can conclude that the likeliness of loosing subsequent values in the case of the previous one being lost is proportionally higher across all collected data.

Table 4.2: Estimation of the conditional data loss probabilities

Dataset type	Frequency of data loss if preceding value is NaN	Frequency of data loss if preceding value is Not NaN	Ratio between NaN and Not NaN preceding values data loss frequency
Cleaning	67.088%	45.324%	1.480
Washing the dishes	65.400%	34.214%	1.911
Cooking	60.703%	30.342%	2.00
Watering the plants	61.463%	16.845%	3.649
Training	62.239%	16.652%	3.737
Resting phase	59.139%	3.572%	16.556

In order to better understand the nature of missing data occurrence and analyze the relationship between duration of the periods with data and empty gaps, the visualizations illustrating the correlation between data points representing preceding and following empty gaps of data with regard to non-empty data periods were implemented for every collected data set (Fig. B.1). Gradient color encoding represents the frequency of specific gap length value occurrence.

The general trend across all activity datasets encompasses short durations of the preceding gaps having a highly variable set of values corresponding to the following gap length, while long preceding gap durations tend to have shorter lengths of the following gaps. Table 4.3 represents the computed correlation coefficients for the respective data sets, visualized in Figure B.1. As the majority of the samples did not possess Gaussian distribution, Spearman correlation coefficient was utilized to estimate the correlation between values of preceding and following gaps for every activity data set. Obtained values of the correlation analysis in Table 4.3 indicate that there is no strong correlation between the values of preceding and following gaps. Spearman's ρ for the cooking, cleaning, washing the dishes and watering the plants activity datasets illustrates low to moderate positive correlation, as on the contrary resting phase and training samples tend to have low to moderate negative correlation between the empty gaps.

Table 4.3: Spearman correlation coefficient for the samples representing preceding and following gaps for every activity dataset

Dataset type	Spearman correlation coefficient (ρ)
Resting phase	-0.138
Training	-0.024
Cleaning	0.029
Cooking	0.034
Washing the dishes	0.103
Watering the plants	0.209

The inter-beat interval data for every activity data set was also plotted with respect to the non-empty periods and missing gaps (see Fig. 4.1). The noticeable trend across all data sets consists in IBI duration drastically increasing in case it was short (e.g. from 400 ms to 750 ms) or significantly plummeting if it was long (e.g. from 750 ms to 450 ms) after the missing period had occurred.

It is worth mentioning that during the data collection process, the recorded heart rate of the participant was never falling below 60, which leads to a conclusion that missing data occurrence was only caused by the lost signal or moving artifacts.

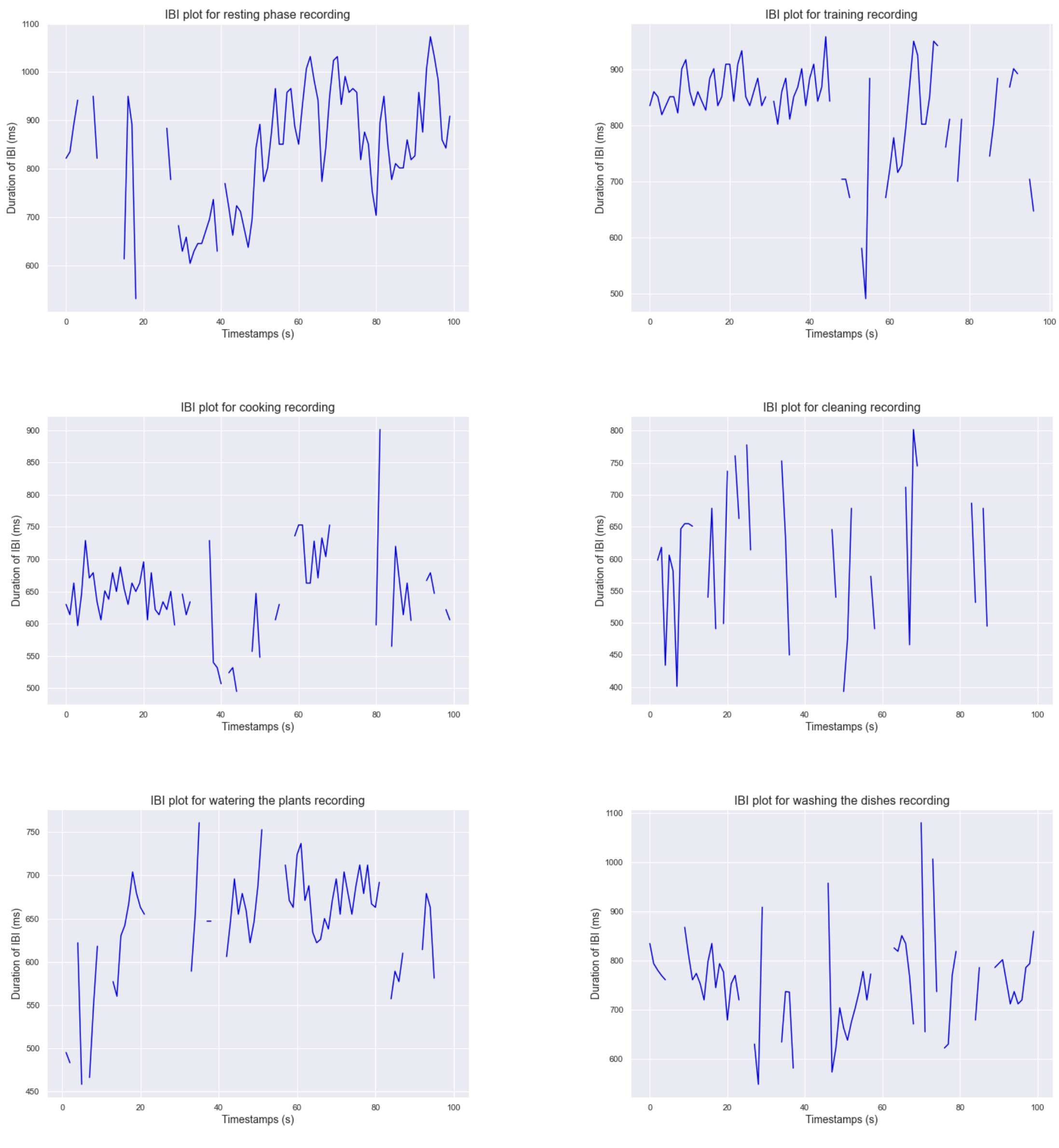

Figure 4.1: IBI plots for every activity dataset

4.4 Missing data mechanisms

Missing data occurs when no data value is assigned to a specific variable in the observation. Datasets can experience different percentages of missing data that are referred to as *missing rates* of the dataset.

In the case of our experiment, the missing values were only analyzed on the univariate type of data with one feature corresponding to the R-R interval values.

Taking into account the results of conditional data loss probabilities in Table 4.2, we conclude that the missingness in our data set is introduced by the *Missing Not At Random* mechanism, as the conditional probability of missing values in case of preceding value being lost is never equal to conditional probability when the preceding value is not missing for every corresponding data set.

Therefore, by not considering the random nature of the missing values occurrence, we are able to identify the specific patterns of missing data that are corresponding to explicit probability distributions.

4.5 Missing patterns discovery

In order to proceed with missing patterns identification a specific approach has to be designed for evaluating the missing data characteristics in a systematic way. In our experiment, a *missing data pattern* corresponds to a particular continuous probability distribution of aggregated missing gaps within the data set, that is generated not by a random mechanism.

For the purpose of missing patterns detection, generated datasets mentioned in the previous section were fitted into a comprehensive set of 90 continuous probability distributions from the Python library *SciPy* [2]. SciPy uses *Maximum Likelihood Estimation* method for estimating the parameters of a fitted distribution. To evaluate the *Goodness of Fit* (GoF) for every distribution, a Kolmogorov–Smirnov statistical test was used that produces test statistic and a p-value. The fitted distribution with the highest p-value and lowest test statistic is considered to be the best fitted distribution to determine the *probability density function* (pdf).

For every dataset consisting of empty gaps (i.e. NaN data sets) and periods with data (i.e. not NaN data sets) we identified the top most fitting probability distributions. Table B.1 (see Appendix) represents the goodness of fit of the most common distributions towards the corresponding NaN data sets in the form of Kolmogorov-Smirnov test statistic values. As we can observe, the most fitting distribution (i.e. distribution with the lowest test-statistic value) for the data set recording during the resting phase and watering the plants activity is *Double-Generalized Gamma* distribution, during the training activity is *Right-skewed Gumbel* distribution, during the cooking and washing the dishes activities is *Von Mises* distribution and finally the *Generalized Normal* distribution is the most fitting for cleaning activity. Although multiple discovered types of probability distribution are fitting particular data sets, we decided to limit our choice to investigate the most recurring ones, namely Double-Generalized Gamma distribution, Right-skewed Gumbel distribution, Von Mises distribution and Generalized Normal distribution.

The *Double-Generalized Gamma* distribution is a continuous probability distribution that happens naturally in systems when the time between events is important. This distribution is used to approximate the time it takes for the k-th (i.e. shape parameter) event to occur [82]. Overall, the Double-Generalized Gamma distribution represents a stochastic process with independent, non-negative increments that has identical scale parameter and a time-dependent shape parameter [83].

The *right-skewed Gumbel* distribution is a popular asymmetric extreme value distribution (EVD) that can be used to simulate operational threats in risk management for nuclear power plant models and predict flood water levels, high winds, earthquakes, and other natural disasters [84, 85].

The *Von Mises* distribution was firstly discovered to model the distribution of atomic weights. Now, it found its application in various scientific fields, such as modelling the spread of disease in

[2] "scipy: SciPy: Scientific Library for Python". Available at: https://www.scipy.org

epidemiology, describing the Brownian motion in physics, training privacy-preserving algorithms in Machine Learning, designing interference alignment in signal processing and even constitutive modeling of soft biological tissues [86, 87, 88].

The *Generalized Normal* distribution (or Generalized Gaussian distribution) is a continuous probability distribution applied in the wide range of tasks, specifically in image compression and filtering, automatic methods for noise variance estimation. Nowadays, this distribution is also commonly used to model processes for embedding watermarks in images, video processing as well as modeling speech signals [89].

With regard to the fitted distributions for datasets containing periods with data, we similarly identified the most common ones (see Table B.2). From the obtained results we derived the top most fitting probability distribution types for every set of collected data, specifically *Log-Laplace* distribution, *Burr* distribution, *Alpha* distribution and *Von Mises* distribution. The aforementioned distributions of non-empty gaps were consequently used in combination with distributions describing empty gaps of data to synthetically generate so-called missing data patterns.

Furthermore, every activity dataset including empty and non-empty periods of data was used to construct *empirical cumulative distribution functions* (ECDFs) and compare it with regard to the theoretical CDFs of the most fitting probability distributions (see Fig. B.2 - B.7).

The key insight about the data distribution derived from the CDF plots depicting durations of the non-empty periods is that the highest portion of the data (from 22 % to 45 % approximately) belongs to the values less than 2. As for the distribution of the data representing durations of empty gaps, the tendency is similar, having data gap equal to 1 constituting 25 % to 40 % of the whole dataset.

The overall trend across all cumulative density functions for every activity dataset encompasses higher variation of values for data representing duration of non-empty periods. Moreover, the frequency of values belonging to certain ranges is significantly higher, which is illustrated by the cumulative probability values for the data representing duration of empty gaps.

In addition, the *kernel density estimator* (KDE) plots were implemented to visualize the underlying probability distribution of the data across all collected data sets using continuous probability density curves constructed from samples with regard to the estimated through parametric tests most fitting distributions (see Fig. B.8 - B.13). From the illustrated plots we can observe that the highest density of the data representing both empty and non-empty periods belongs to the very short durations (e.g. 1 and 2 seconds). In the majority of the cases, the KDE from samples curve more accurately described the distribution of data in comparison with parametrically estimated distribution, having identified correct peaks representing highly dense values.

Overall, through KDE visualizations we observe that all datasets have the unimodal probability distribution, except for the watering the plants activity dataset with empty data periods, whose distribution is multimodal, having most highly dense gap durations equal to 1 and 3 seconds.

We can conclude that there is a difference when collecting data during resting phase and cleaning as there are factors affecting the noise that cannot be predicted in advance. Hence, identifying missing data patterns would enable us to expect apriori which kind of data we are able to collect in advance, as different recorded activities result in different missing data patterns.

4.6 Synthetic generation of missing data patterns

In order to evaluate the precision of the derived HRV parameters based on the different amount of data loss, we had to synthetically impute the full coherent dataset containing inter-beat intervals with a set of missing data patterns. The idea of using concise and accurate dataset was reasoned by our intentions of assessing the affect of missing data on the quality of the HRV derived from the perfectly aligned data.

In this experiment, we used the R-R Interval Time Series data set from the PhysioNet Computing in Cardiology Challenge 2002 as the golden standard for imputing the patterns of missing values and estimating the change in accuracy of the computed HRV parameters. The data set contains 10 hours of continuous Inter-Beat Intervals data from long-term ECG recordings of adults between the ages of 20 and 50 who have no known cardiac abnormalities [90]. The data set is pre-processed and does not include any recordings with significant amounts of noise or ectopic peaks.

To proceed with the missing values imputation process, the most common patterns of missing data outlined in Section 4.5 were used to generate respective probability-density functions (pdfs) per each probability distribution type in every pattern with 5 different sets of input parameters. The pdfs defined multiple scenarios from which the missing values were introduced at different missing rates (5, 10, 25, 50 and 75 %). For every missing rate the imputation of missing values was performed 30 times to eliminate the bias from the derived conclusions. As a result of the missing data synthetic generation, 3000 new datasets were obtained, each comprising missing values according to specific pattern from pdf, set of input parameters, missing rate and iteration of the imputation process.

Figure 4.2 illustrates the KDE visualizations for the inter-beat interval values distributions before and after imputation with missing data was introduced. We can observe that at 5% rate of the introduced missing values, the probability density of the original dataset and the amputed one is still perfectly aligned. However, as the missing rate starts to increase, we notice a slight shift and an emerging skewness in the probability density of the amputed dataset. Taking into account results of these visual analytics we can confirm the importance of using the data whose properties are as realistic as possible (i.e. the shift in distribution should be present once the missing rate increases), hence it will be aligned and in synchronization with the data that would be collected in the real-life setting.

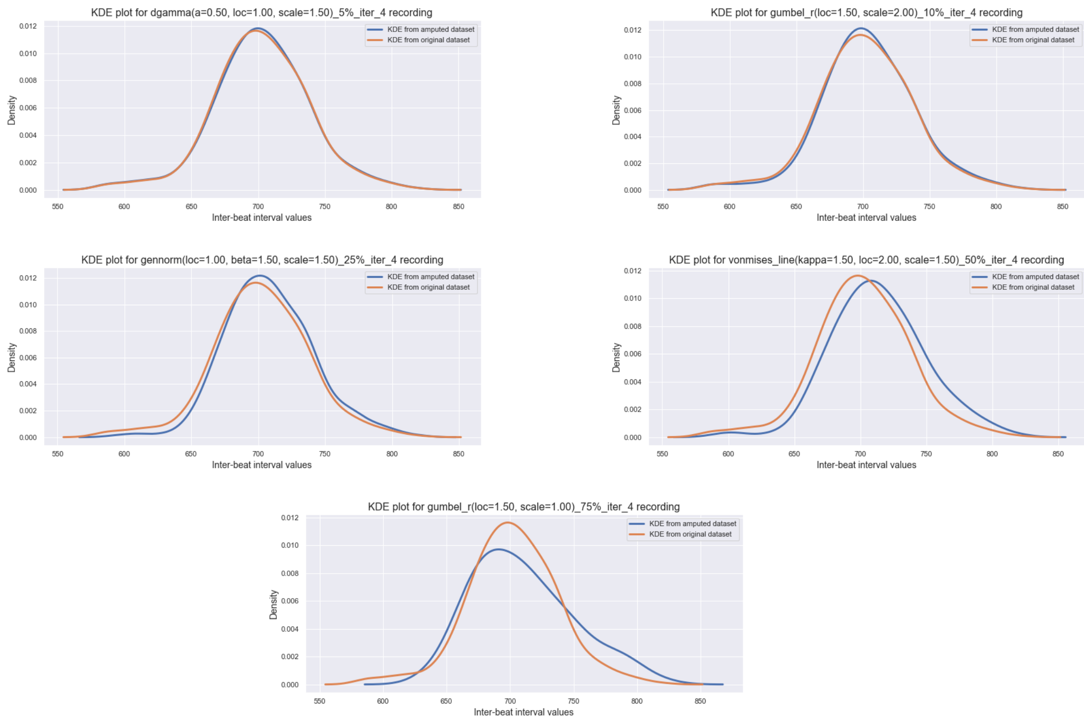

Figure 4.2: KDE plots for IBI values from original and imputed with missing data patterns datasets

4.7 HRV parameters analysis

In order to compute the gold-standard HRV features, the original 10 h R-R intervals time-series as well as data sets with synthetically generated missing values were split into 5-min time-windows [91].

For the aim of our study to evaluate the effect of missing patterns on the HRV parameters accuracy, we focused on the most popular time-domain HRV measurement *rMSSD*, that represents the root mean square of the successive IBI differences estimates short-term components of HRV [92]. By using the Python *hrv-analysis* library [3], the rMSSD metric was calculated and averaged in each 5-min time-window from the original data set and in every data set with imputed missing patterns for every missing rate, input parameters set and iteration.

The error estimation between gold-standard HRV parameters calculated within 5-min time-window and HRV parameters with missing values were assessed by the *Root Mean Squared Error* (RMSE), which is a risk metric corresponding to the expected value of the squared (quadratic) error or loss and represents the precision estimation for regression tasks (i.e. how close the measurements are to each other) [93]. RMSE values closer to 0 serve as more precise HRV parameters. For the accuracy assessment of the derived HRV features (i.e. how close the measurements are to the target "true" value) the coefficient of determination R^2 was used, that represents a measure of the correlation between HRV parameters of the original and imputed with missing values (efficient values should have a value closer to 1) [94]. We define the threshold of HRV values to be inaccurate when there is an increase for more than 5 times in the RMSE metric and decrease below the level of -0.15 for R^2 score metric.

Finally, *residual analysis* was performed to quantify the agreement between the quantitatives of gold-standard rMSSD values and those from partially empty time-series by computing the residuals and carrying out the visual analytics in a form of residual plot graphs. Residuals represent the difference between the observed values (gold-standard rMSSD) and the fitted rMSSD values derived from synthetically amputated data sets. For the computed residuals the *Bias* metric was used, that represents the mean ± standard deviation of the differences.

In order to visually examine the distribution nature of the residuals the residual plots were generated that depicted fitted rMSSD values on the x-axis and computed residuals on the y-axis. Besides the distribution of the residuals the plot also visualized the *lowess curve*, obtained through the bivariate smoother function. Locally, the curve minimizes the variance of the residuals or prediction error [95]. In addition, to check the assumption of normality of the residuals we plotted histograms and ran the *Anderson-Darling* statistical test. This particular test was chosen for testing the normality as it is robust and not sensitive to the ties occurring in the data.

Furthermore, the *Q-Q plots* were utilized to compare the probability distribution of the gold-standard rMSSD values with those derived from the amputated datasets by plotting their quantiles against each other.

4.8 Experimental results and discussion

The obtained statistical results and descriptive statistics of rMSSD and computed error metrics for all probability distribution types (i.e. missing data patterns), types of input parameters used for generating the distributions and missing rate used for imputation are reported in Table B.3 (see Appendix). Each outlined pattern of missing data was imputed by generating empty gap values according to the first distribution and placing them between the interval values of non-empty periods according to the second distribution in column *"Probability distribution types for empty and non-empty gaps", see Table B.3.*

The statistical results are aimed at comparing the gold-standard rMSSD of the original dataset with amputed datasets. The RMSE values detected across different types of missing data patterns do not deviate highly between each other, usually ranging from 1 to 14, which corresponds to the moderate precision (see Table 6). Moreover, there is a noticeable trend in the RMSE values significant growth starting from missing rates of 25 %, having error values varying from 0.9 to 1.8 at the 10 % of introduced missing values and 5-16 at the missing rates 25-75 % accordingly. Hence, we can consider 25 % of missing values and more to be the threshold of significant loss in precision of the HRV parameters, according to the RMSE estimation. Overall, although the

[3] R. Champseix, "hrv-analysis: a package to calculate features from RR Interval for HRV analysis". Available at: https://github.com/Aura-healthcare/hrv-analysis

RMSE values are quite high, they are still moderate according to the scale of rMSSD values and do not considerably alter them, as the 95 % confidence interval for rMSDD is 21–87 [96].

Similarly, the trend of losing the rMSSD accuracy, according to the R^2 score, follows up for all types of probability distributions and input parameters, as the missing rate is rising. Moreover, by estimating the accuracy through the R^2 score metric, we can conclude that 25 % of data loss is also considered to be a threshold above which missing values are highly affecting the quality of HRV features. Regarding the estimation of Bias, we can conclude that with introducing higher rates of missing values into the data set, the values of rMSSD are getting lower and hence the difference with respect to the golden-standard rMSSD is gradually increasing.

Furthermore, the pattern right-skewed Gumbel (loc=1.5, scale=3.00) introduced at 75 % of missing rate was identified as the one having the highest error rate according to RMSE and R^2 score metrics. Hence the HRV values derived from the data sets holding this missing data pattern would be highly affected with regard to accuracy and precision.

Generally analyzing the obtained results of the error estimation metrics, we conclude that overall R^2 score values that represent accuracy tend to be low, while the values of RMSE that stand for precision are moderately high (i.e. the estimated error is low with respect to the scale of the unit). Thus, having low accuracy, but high precision we come to a conclusion that there is a *systematic error* in the HRV features computed from the time-series with more than 25 % of missing rate.

The carried out residual analysis is displayed in Fig. B.14. The residuals were plotted for every missing data pattern and every missing rate, in order to detect the general trend of rMSSD deviation and identify if the loss in accuracy and precision is determined by the random or systematic type of error. In the ideal scenario, residuals are evenly and symmetrically distributed, tending to cluster towards the middle of the plot and around the lower single digits of the y-axis, as well as not forming clear patterns of data [97]. Residual plots for rMSSDs computed with 5 % and 10 % of missing rate (see Fig. B.14) illustrate the high accuracy and precision of the fitted rMSSD values, as residuals are distributed evenly, although only above the x-axis. The results of residual analysis for the fitted rMSSD values with 25 % of missing rate are quite consistent with the previous pattern type, although the values are getting slightly shifted from the x-axis, which signifies accuracy loss. The same trend is followed by the subsequent residual analysis at the missing rate of 50 and 75 %.

The conducted analysis of residuals normality using Anderson-Darling test revealed that for all types of missing patterns and introduced missing rates the residuals hold a normal distribution (see Fig. B.15).

Overall, based on the residual analysis we can again conclude that there is a systematic error for the rMSSD values computed from the data with 25 % of missing rate and higher, as with the slight decrease of accuracy, the precision still remains to be high.

As for the Q-Q plots depicted in Fig. B.16, the rMSSD values up to the missing rate of 10 % are light-tailed and tend to increase alongside with linear relation. However, starting from the 25 % of missing rate the distribution of rMSSD values becomes significantly spread out, although the general linear growth continues. Hence, we can only consider data to be *reliable and usable* if the rate of the missing data is below 25 %.

In general, we can outline the proposed methodology for estimating the missing signal impact for computing HRV in a following way: (1) firstly the collected data has to be pre-processed to generate new datasets consisting of durations of empty and non-empty gaps according to recorded timestamps (2) the frequence of total and conditional data loss has to calculated (3) correlation analysis needs to be performed on the data representing preceding and following empty gaps of data with re-gard to non-empty data periods (4) missing data pattern has to be identified as a combination of two continuous probability distributions of the empty and non-empty gap duration datasets using Maximum Likelihood Estimation 5) the detected missing data pattern needs to be introduced at different missing rates on the clean dataset to evaluate the accuracy of the HRV in case of their repeatability in the subsequent data acquisition 6) HRV parameters must be computed from the clean dataset and the one imputed with missing pattern 7) for performing an error estimation between gold-standard HRV parameters and those from the amputed datasets, the following regression metrics can be utilized: Root mean square error (RMSE), R^2 score and Bias in addition to visualized residual analysis and Q-Q plots.

4.9 Assessment of experiment repeatability

In order to determine if the identified missing data patterns from the previous experience have a tendency for reoccurrence, the second data collection process was carried out. IBI time-series were recorded through the same PPG-sensor mentioned in Section 4.1 and extracted using the same toolkit. The recording was organized during the same set of activities, except for the resting phase with deep breathing as it was not representative for the pattern identification due to the low frequency of data loss, which was discovered in the previous experiment. Similarly, the NaN values were imputed at the locations, where the IBI values were missing at specific timestamps.

The obtained R-R interval time-series were firstly analyzed for the overall amount of lost data (see Table 4.4). In comparison with the results of Table 4.1, the missing rate percentage during the second data collection was similar for the recordings of the training activity (43.68 % and 43.57 %), moderately lower for cleaning and watering the plants activities, and significantly lower for washing the dishes activity (28.65 % and 49.72 %). Overall, the activity with the highest percentage of data loss in the repeated experiment was cleaning (47.61 %), same as in the data from the first data collection.

Table 4.4: Estimation of the overall data loss frequency during repeated experiment

Data set type	Frequency of data loss
Watering the plants	20.24%
Washing the dishes	28.65%
Cooking	41.14%
Training	43.68%
Cleaning	47.61%

The analysis was also performed for estimating the conditional probabilities of data loss. In the case of the second data collection process, the percentage of data loss in both cases when the preceding values are empty and non-empty is moderately lower for the data sets recorded during all outlined activities, except for the training (see Table 4.5). It is worth mentioning that for activities the conditional probabilities of data loss if the preceding values is missing were higher than for the cases when it is not missing. This could be explained by the fact that the occurrence of singular missing values in the data of the aforementioned activities is higher than the occurrence of the sequential chains of missing values as well as shorter periods of lost signal during recording.

Table 4.5: Estimation of the conditional data loss probabilities during repeated experiment

Dataset type	Frequency of data loss if preceding value is NaN	Frequency of data loss if preceding value is Not NaN	Ratio between NaN and Not NaN preceding values data loss frequency
Cooking	57.895%	29.434%	1.967
Cleaning	48.372%	20.728%	2.333
Training	64.689%	27.382%	2.362
Washing the dishes	69.096%	27.947%	2.472
Watering the plants	49.683%	12.691%	3.915

As in the previous experiment, the aggregated missing values data sets of empty and non-empty periods, recorded during the second data collection process were fitted into multiple continuous probability distributions. The most fitting distributions were identified as the corresponding missing data patterns for every data set (see Table B.4 and B.5).

During the second experiment the determined patterns included new top fitting distributions, such as *Normal* distribution, *Logistic* and *Generalized Logistic* distributions, *Generalized Exponential* distribution, *Inverted Gamma*, etc. The most fitting missing data pattern for the cooking activity data set was identified as the same one from the previous experiment and corresponded to *Von Mises* distribution.

As for the other activities, the best fitting distributions were different, specifically training activity data set corresponded to *Double-Generalized Gamma* distribution, washing the dishes activity having *Generalized Logistic* missing data pattern, and cleaning and watering the plants activities possessing *Logistic* distribution. Nevertheless, despite the distinguishing most fitting missing data patterns for each separate data set, discovered probability distributions that are commonly fitting across all the collected data sets still remain to be *Double-Generalized Gamma*, *Right-skewed Gumbel*, *Von Mises* and *Generalized Normal* distributions.

Chapter 5

Reliability assessment of the PPG-based HRV parameters

Heart rate variability (HRV) analysis has experienced a steep increase in popularity through the past recent years as a way to examine the autonomic nervous system's ability to regulate the heart. HRV analysis has been acknowledged as one of the greatest approaches to examine the autonomic nervous system since it is non-invasive, easy, and can quantify and discriminate sympathetic and parasympathetic activity. The most established and common method supported by the scientific community for conducting the HRV analysis is *electrocardiogram* (ECG). The electrocardiogram (ECG) is one of the most reliable indicators of physical health and heart function assessment. It is a graphical representation of the electrical activity of the heart. Currently, the straightforward method for obtaining the HRV time-domain and frequency-domain features from the raw ECG signal data is to firstly apply the suitable QRS detection algorithms to detect the peak of the R-wave, compute the RR intervals based on the peak locations, and perform appropriate interpolation alongside with re-sampling to align the sample frequency in the tachogram. This approach is cumbersome and possesses an additional overhead for the data processing. Besides, as the classic ECG monitor is a stationary device that allows analyzing abnormal cardiac physiology phenomenon, it is limiting the activities performed by a human, hence is not considered as a feasible long-term monitoring device. Furthermore, for ECG measurement two conductive electrodes have to be attached to the limbs and chest, which narrows down the measuring area and causes inconvenience for the participants.

In contrast, an alternative solution would be to use the *photoplethysmography* (PPG) monitor for providing potentially similar measurements. Photoplethysmography (PPG) detects changes in blood volume in the microvascular bed of tissue using a light emitting diode (LED) and a photodiode that illuminate the skin, making it less vulnerable to power supply noise and electromagnetic interference. PPG devices hold the main advantage for enabling the data collection from wider areas of placement and allowing more autonomous monitoring. In addition, the majority of PPG monitors directly provide the recorded R-R intervals data, which significantly simplifies the subsequent HRV analysis.

In this Chapter we aim to present a systematic approach on assessing the quality of HRV features from the PPG-obtained data in comparison with the gold-standard ECG-derived HRV values. By conducting analytical evaluation of the data we would like to understand whether the HRV parameters obtained from the PPG device can be used for substituting the HRV features computed from ECG recorded data. Consequently, this chapter is addressing the problem of how to estimate the influence of the device on the reliability of the derived HRV features.

In Section 5.1 we discuss already existing researches aimed at evaluating the quality of the PPG-based sensors. Section 5.2 describes the data collection setup that was conducted using two heart rate monitors, specifically the PPG device and a gold-standard ECG sensor. Within Section 5.3 we illustrate the methodology for initial exploration of the raw ECG signal and steps performed to pre-process this data for the device benchmarking using analytical tools. Specifically, the Section outlines the following methodological stages: elimination of moving artifacts, filtering algorithms for noise removal, R-peak detection method and inter-beat interval values derivation. In Section 5.4 we present the procedure of the HRV analysis that performs a comparison between ECG and PPG-derived HRV features, namely it outlines specific implemented visualizations, statistical tests and regression error metrics that were used. Section 5.5 provides the experiment results assessment and discusses the validity of them. In Sections 5.6 and 5.7 we delineate the findings of the repeated data experiment aimed at double-checking the insights obtained within the initial experiment. Finally, Section 5.8 summarizes the results of all the conducted analysis according to the proposed methodology and addresses the main research question targeted by this Chapter regarding the device reliability.

5.1 Related work

Methods and frameworks for evaluating the quality of the HRV features from ECG and PPG were previously introduced by the research community. Hsiao et al. (2017) explored the correlation of heart rate variability (HRV) results in different postures between ECG-based and PPG-based cardiac measurement devices. Medical 12-lead Poly-Spectrum-12/E ECG monitor was used as a benchmarking device for comparing the results with data obtained through wearable prototype Z1 Wristband and commercially available Mio Alpha HR watch PPG-based devices. The experimental results illustrated that in different postures, the HRV results of our Z1 prototype are of significant positive correlation with the HRV results of the ECG device ($r \geq 0.861, p \leq 0.01$) [98].

A study conducted by Pinheiro et al. (2016) investigated the hypothesis of using HRV features from PPG-obtained data as surrogates for HRV indexes from ECG-obtained data, in three different contexts: healthy subjects at rest, healthy subjects after physical exercise and subjects with cardiovascular diseases (CVD). With substantial correlations above 82 % for both time and frequency features, the PPG-based HRV can be utilized as an alternative for HRV analysis in healthy subjects, according to the findings. In contrast, time and (most significantly) frequency domain features should be employed with caution in post-exercise and CVD participants (mean correlations ranging from 68 to 88 %) [99].

In an alternative research by Kinnunen et al. (2020), nocturnal HR and HRV were assessed in 49 adults with simultaneous measurements from the Oura ring and the gold standard ECG measurement. It was observed that there is a very high agreement between the ring and ECG for nightly average HR and HRV ($r2 = 0.996$ and 0.980, respectively) with a mean bias of 0.63 bpm and 1.2 ms. Findings from the study indicate high validity of the Oura ring in the assessment of nocturnal HR and HRV in healthy adults [20].

5.2 Data acquisition

In order to investigate the reliability of the PPG-derived data, the data collection process was organized in a way to enable simultaneous recording of the ECG and PPG signals that facilitated accurate control of the samples duration. Data acquisition was carried out during five separate intervals within a day. Each interval of recording had a duration of 30 minutes and consisted of simultaneously repeating activity. Data was collected for the initially predefined set of tasks that were not involving any dynamic movements or intense physical activity. Specifically, data was recorded during following activities: meditation, reading, writing, participating in a video-call and coding session. The reason behind was to provide clear, noise-free data that is not influenced by signal loss issues or motion artifacts.

The experiment involved two heart rate monitors used for recording physiological signals: Aidlab 2.1 and Scosche Rhythm24 Heart Rate Monitor.

Aidlab 2.1 is a wearable chest strap equipped with a 1-Lead ECG sensor, respiration sensor, 9-Axis motion sensor and infrared skin temperature sensor. Aidlab ECG and respiration monitors allow tracking the electric work of the heart and lungs. The chest strap has a basic electrical impedance of approximately 500 Hz for a high frequency (64 kHz) test signal [100]. In the current experiment, Aidlab monitor was recording the heart's ECG signal at a sampling rate of 512 Hz. Raw ECG signal amplitude recorded in millivolts (mV) along with corresponding timestamps was transferred through Bluetooth Low Energy connection from the device to the Aidlab native mobile application. Application allowed extracting the data from the Aidlab Cloud storage in a form of CSV files.

Scosche Rhythm24 Heart Rate Monitor is an armband device equipped with a Valencell PPG sensor [79, 80]. Using the Bluetooth Low Energy connection data containing R-R intervals and corresponding timestamps was sent as packages to the HRV Logger mobile application [101]. HRV Logger stores the collected data about inter-beat intervals as CSV files to the Dropbox Cloud storage.

Figure 5.1: Scosche Rhythm24

Figure 5.2: Aidlab 2.1

Figure 5.3: ECG and PPG devices for data acquisition

5.3 Exploratory data analysis and processing of the raw ECG signal

This section describes the methodology of the raw ECG signal processing steps and outlines some preliminary insights based on the initial visual exploratory analysis. The end goal of the pre-processing is to obtain clean, detrended dataset containing coherent inter-beat interval values without any noise.

5.3.1 Noise filtering and artifacts removal

First of all, the extracted raw ECG signal data from the Aidlab cloud storage was preliminary analyzed for the purpose of identifying noise and outliers. During exploration phase it was concluded that all collected datasets contain a certain amount of noise. Examples of that are illustrated on the Fig. 5.4. Using the Python library *heartpy* [2], Hampel filtering technique was applied on the signal data to detect and remove noise. Hampel filter operates by computing the median of a window composed of the sample and its six surrounding samples, three per side for each sample of data from the input vector. Furthermore, it estimates the standard deviation of each sample with regard to its window median using the median absolute deviation. In case a sample is differing from the median by more than three standard deviations, it is replaced with the median [102].

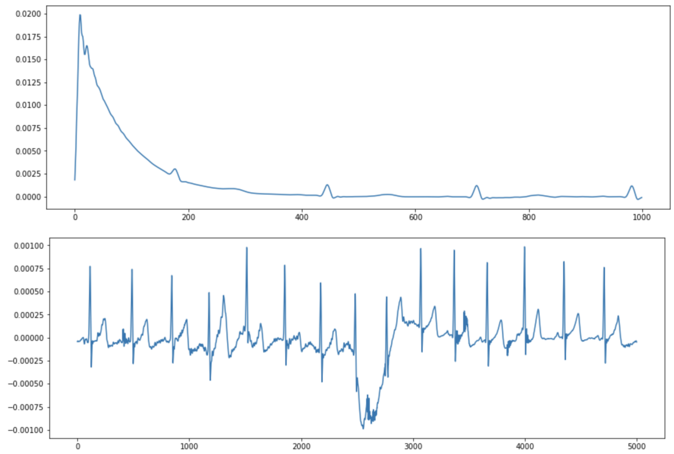

Figure 5.4: Noisy data points in the raw ECG data

Afterwards, a Notch filter with a cutoff frequency of 0.05Hz was applied on the cleaned from outliers data points. Performing this pre-processing step resulted in removing the *baseline wander* from the signal (see Fig. 5.5). Baseline wander is a low-frequency artefact in the ECG signal recordings of a subject. The benefit behind using a Notch filter resides in a fact that it does not introduce any phase shift, thus the QRS complexes and peaks remain at the same position [103].

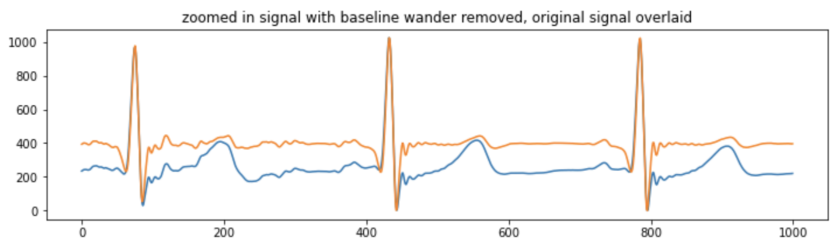

Figure 5.5: Baseline wander removed from the original raw ECG signal

Overall, the ECG data was of a good quality, hence there was no need to perform any subsequent peak correction.

5.3.2 Peak detection and IBI computation

In order to proceed with the HRV ECG-obtained data computation, datasets should be transformed into R-R intervals. Hence, the pre-processed and cleaned raw ECG data from the Section 5.2 was used as an input source to proceed with R-peaks detection. Peak detection was performed using the aforementioned *heartpy* Python library. R-peak detection algorithm developed within heartpy library is built on top of the Pan-Tompkins peak detection algorithm [38].

The idea behind peak detection phase is to locate the amplitude variation and morphology changes of the QRS complexes by using an adaptive peak detection threshold, that is followed by multiple steps of outlier detection and rejection. A moving average is computed on both sides of each data point to identify the heartbeats, with a window of 0.75 seconds on both sides. The signal's mean value is used to populate the first and last 0.75 seconds of the signal, hence no moving average is calculated for these sections. Regions of interest (ROI) are defined as two locations of intersection where the signal amplitude is greater than the moving average. R-peaks are identified at the highest point of each region of interest. During the peak detection phase, the algorithm incrementally adjusts the amplitude of the calculated threshold. The standard deviation between consecutive differences (SDSD) is minimized and the signal's BPM is verified to identify the best fit [2] (see Fig. 5.7).

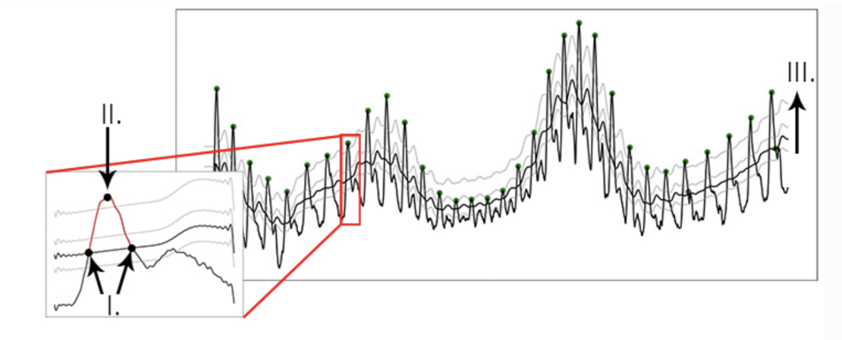

Figure 5.6: Peak extraction phase within the heartpy algorithm [2]

Once the fitting phase is finished, multiple incorrectly detected peaks residing in the data are tested and rejected based on threshold defined from specific section. Thresholds are calculated based on the mean of the RR-interval values in the segments (see Fig. 5.7) [104].

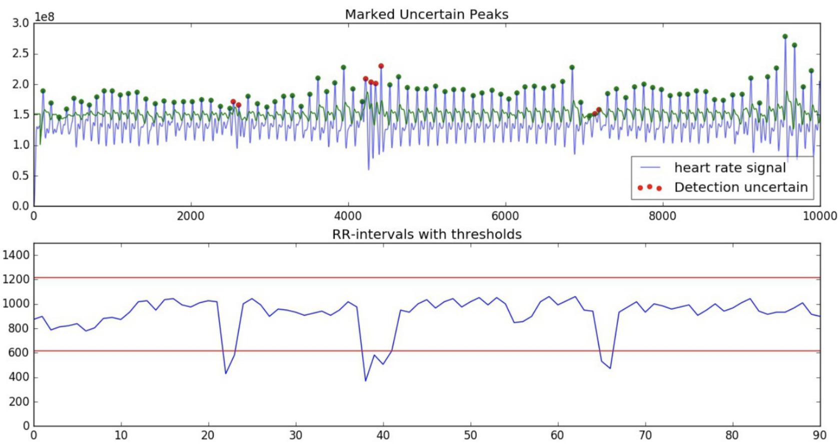

Figure 5.7: Peak rejection phase within the heartpy algorithm [2]

As a next step, correctly detected and verified peak locations from every ECG dataset were used to compute the R-R intervals using the in-build method within heartpy package (see Fig. 5.8). It can be observed that the algorithm managed to accurately identify the peaks in the signal, which is illustrated by green dots, as well as precisely reject the minor noise in the data, which is not considered to be a peak.

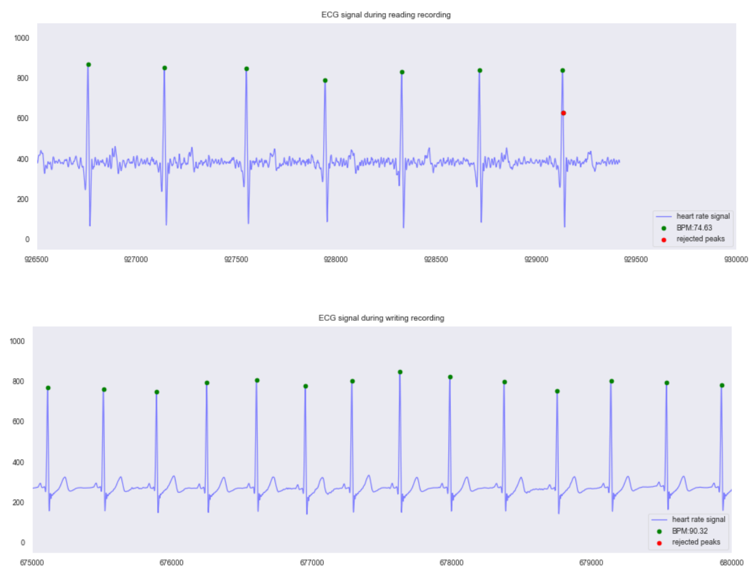

Figure 5.8: Pre-processed segments of ECG signal with correctly detected R-peaks

5.4 HRV analysis of the ECG and PPG data

Data acquired from the recording through the PPG sensor as well as pre-processed data from the ECG device as criterion measurement were used to compute the HRV features. Time-series representing R-R intervals from both devices were split into 5-min time-windows and used as an input to the Kubios HRV software. *Kubios HRV* software is a device agnostic desktop application, supporting a variety of regularly used HR monitors, ECG devices, and PPG monitors. Kubios HRV delivers precise and extensive HRV analysis for short- and long-term data, as well as for all types of study procedures. Kubios HRV software has reached gold-standard reputation in scientific research, with over 1200 universities in 128 countries using it [1], Elite HRV [69]. By using Kubios HRV, top four most common and relevant for this study HRV features were calculated in each 5-min time-windows from every ECG dataset and corresponding PPG dataset: *rMSSD*, *SDNN*, *pNN50* and *Mean RRI*. During the HRV computation, filtering was applied on the data from both recording devices. Filtering represented low beat-correction to eliminate potentially existing noise within R-R intervals.

In order to perform general exploratory analysis of the R-R intervals derived from ECG and PPG recordings we utilized the distribution histograms, depicting the frequency of the values within specific ranges in the datasets. The distributions were checked for normality using Anderson-Darling test. In addition, Kolmogorov-Smirnov non-parametric test was carried out to quantify the distance between the empirical distribution functions of two samples, corresponding to ECG and PPG datasets.

The error evaluation between HRV parameters derived from the ECG dataset representing gold-standard measurements and HRV computed from PPG-obtained data was assessed by calculating the Root mean square error (RMSE), R^2 score, Bias, Mean absolute error (MAE) and Mean relative error (MRE). *Bias* represents mean difference (MD) and standard deviation of the mean difference (SDD), MD $\pm$ SDD. *Mean absolute error* and *Mean absolute percentage error* are risk metrics, corresponding to the expected value of the absolute error loss or *l1*-norm loss [93]. The *mean relative error* was used to estimate how good the HRV values from PPG data were relative to the gold-standard HRV values obtained from the ECG data. The MRE is defined in a following way:

[1]"Kubios HRV Analysis Software - Heart Rate Variability". Available at: https://www.kubios.com

$$MRE(\%) = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{X_k^i - X_{orig}^i}{X_{orig}^i} \right| \tag{5.1}$$

where X_{orig} represents gold-standard HRV derived from the ECG data, and X_k denotes HRV value from the PPG dataset.

For every used regression error metric we defined a criteria for considering the produced HRV values reliable. Specifically, the RMSE values should not go above the value of 15 for rMSSD parameter, 10 for the SDNN parameter, 25 for the Mean RR and 10 for pNN50 values. The reason because why threshold are different per multiple HRV parameters is that the standard deviation range varies per different HRV feature. As for the R^2 score, we consider values that do not reach below -10 to be reliable, taking the nature highly variable HRV values into account. For the rest of absolute metrics we will primarily focus on comparing the changes and identifying the minimum as defining a strict threshold is not feasible with regard to absolute values.

In order to compare the correlation between the quantitatives of golden-standard HRV from the ECG and those derived from PPG data, the *Spearman correlation coefficient* (ρ) was computed and scatter plots illustrating the HRV values from all conducted recordings for both ECG and PPG were visualized. The idea behind using Spearman correlation coefficient was due to the reason that the data samples of HRV values are small in size, which results in having a non-Gaussian distribution.

Finally, *Bland-Altman* plots were constructed for the ECG and PPG derived HRV values. Mean difference and limits of agreement (LoA, MD±1.96*SDD) were plotted on the Bland-Altman plots. Bland-Altman (BA) plot is used to describe the agreement between two quantitative measurements. The resulting graph is an XY scatter plot, with the Y axis representing the difference between the two paired measurements (A-B) and the X axis representing the average of these measurements ((A+B)/2). To put it another way, the difference between the two paired measurements is plotted against the mean of the two measurements. Bland-Altman plots advise that 95 percent of the data points should reside within two standard deviations of the mean difference [105]. To rule out the systematic error, generated Bland-Altman plots were visually examined.

5.5 Initial experiment results and discussion

The statistical analysis is aimed at comparing the gold-standard HRV values from the ECG device with regard to results from PPG device. As a first part of exploratory analysis, the raw non-converted to HRV R-R intervals values were visualized as frequency distribution plots. It can be observed in the Fig. 5.9 that the overall amount of data coming from the PPG device is significantly lower, than the recorded data through the ECG device. However, the general shapes of the distributions corresponding to ECG and PPG-obtained data are similar with regard to unimodality observed in meditation, reading and coding datasets, that potentially leads to promising results in comparing converted into HRV values data points and small percentage of discrepancy.

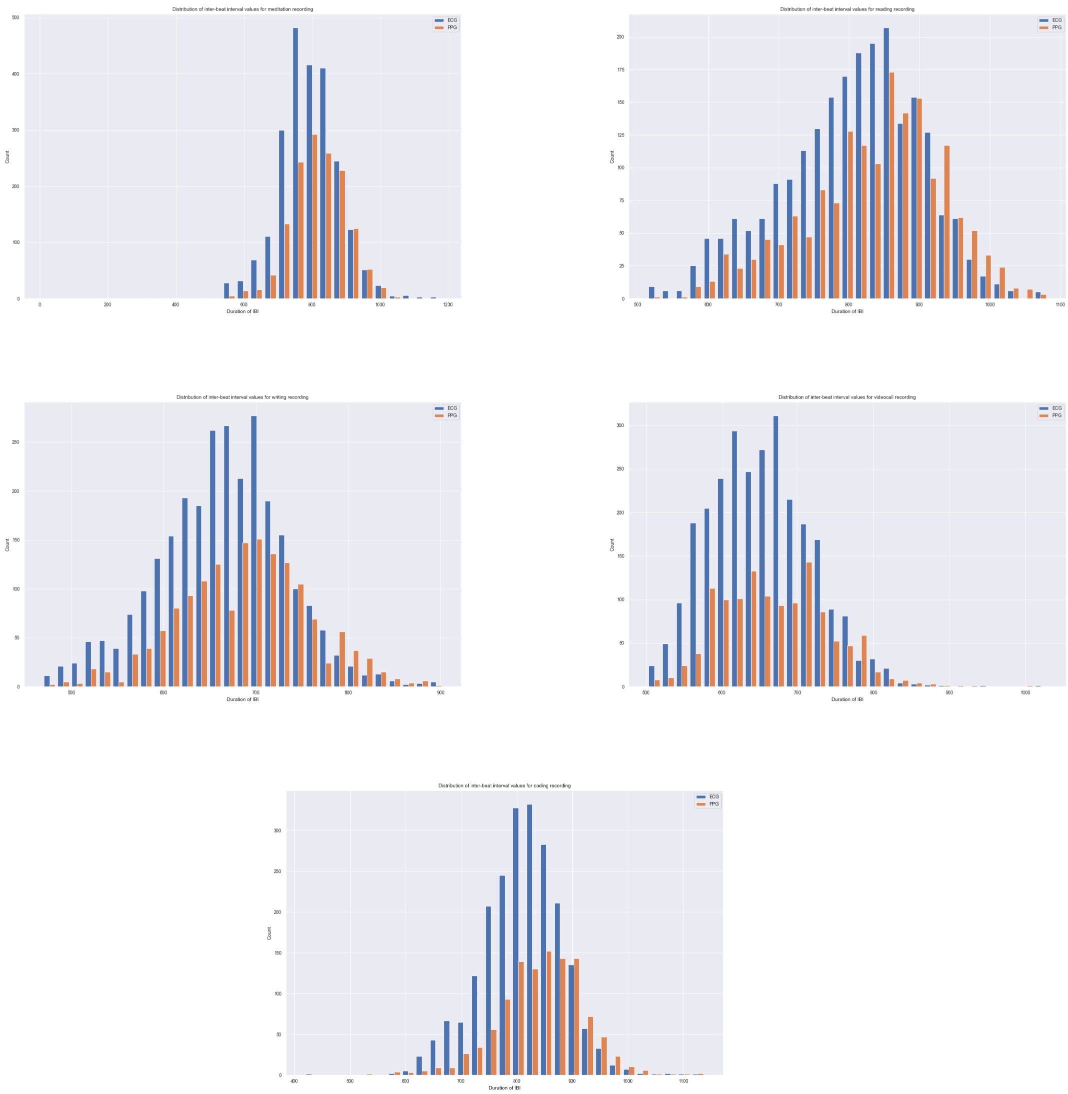

Figure 5.9: Frequency distribution for ECG and PPG recordings

As a next step, the distributions of the R-R intervals data coming from ECG and PPG were checked for normality. In order to precisely access if the data comes from normal distribution an appropriate normality test should be conducted. Hence, for this purpose the data sets were visualized as histogram with the unique values in a data set plotted against its count. The idea behind this visualization was to depict the amount of ties in the data thus choosing the most relevant normality test. The Fig. C.1 - C.5 illustrate the frequency distributions of ties in the data. Due to the fact that the amount of data recorded by the PPG sensor is considerably smaller, the bin sizes in the PPG datasets distributions are larger. It can be concluded that both recordings obtained from ECG and PPG devices for all performed activities contain a lot of ties. Moreover, the amount of ties in the distributions corresponding to the ECG-obtained datasets is significantly higher, than for PPG, which can be explained by a heavy data loss during recording through PPG device.

Taking into account the aforementioned visual ties analysis, Anderson-Darling test was chosen as an omnibus and most powerful test for detecting all departures from normality in the datasets. The Anderson-Darling test produced p-values for all analyzed datasets, which were put into comparison with the level of significance (i.e. alpha) that was equal to 0.05. Resulting p-values lower than significance level were used to reject the null-hypothesis for particular dataset to have a normal distribution (see Fig. C.6 - C.10). As we can observe, neither of the datasets acquired from ECG and PPG recording has a normal distribution.

Performing the Kolmogorov-Smirnov test on the data enabled us to check the hypothesis if data samples from ECG recording and data samples from PPG recording come from the same distribution. The rejection level used for the aforementioned test equaled to 5 %.

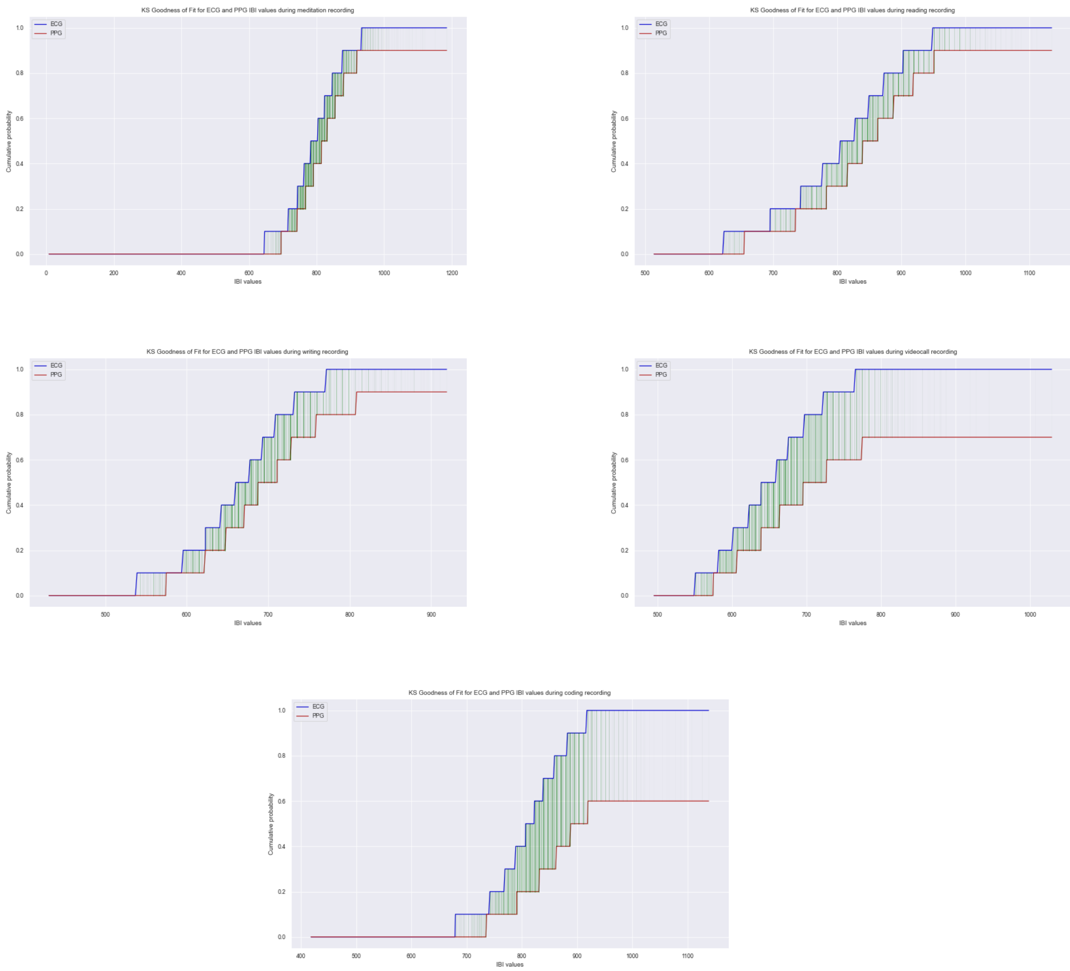

Figure 5.10: Cumulative probability plots for ECG and PPG datasets

The results depicted in Fig. 5.10 turned out to be ambiguous, as the datasets obtained during meditation, reading and coding activities were confirmed to have different types of the distribution. This can be explained by omni-present data loss issue within the PPG sensor as well as a potential unreliability of the ECG device that was chosen as a benchmark. Although, the null hypothesis suggested by the Kolmogorov-Smirnov test was not rejected with regard to datasets recorded during writing and videocall activities. Hence, the empirical cumulative distribution functions (ECDFs) of those datasets is considered to be quite similar. To further understand the extend of deviation of PPG-obtained data from the ECG, the HRV parameters analysis was carried out.

Scatter plots used for visually depicting the correlation between ECG and PPG HRV values are outlined in the Fig. 5.11. Overall, all derived HRV values (e.g. rMSSD, SDNN, pNN50 and

Mean RRI) are higher for the recording obtained through PPG sensor, which indicates the high variation in values within those datasets. We can see that there is a trend in HRV parameters proportionally increasing, that suggests strong correlation between values. However, there was noticed a slight misalignment of values for all computed HRV parameters except the Mean RRI, illustrated by left skewness of the distribution.

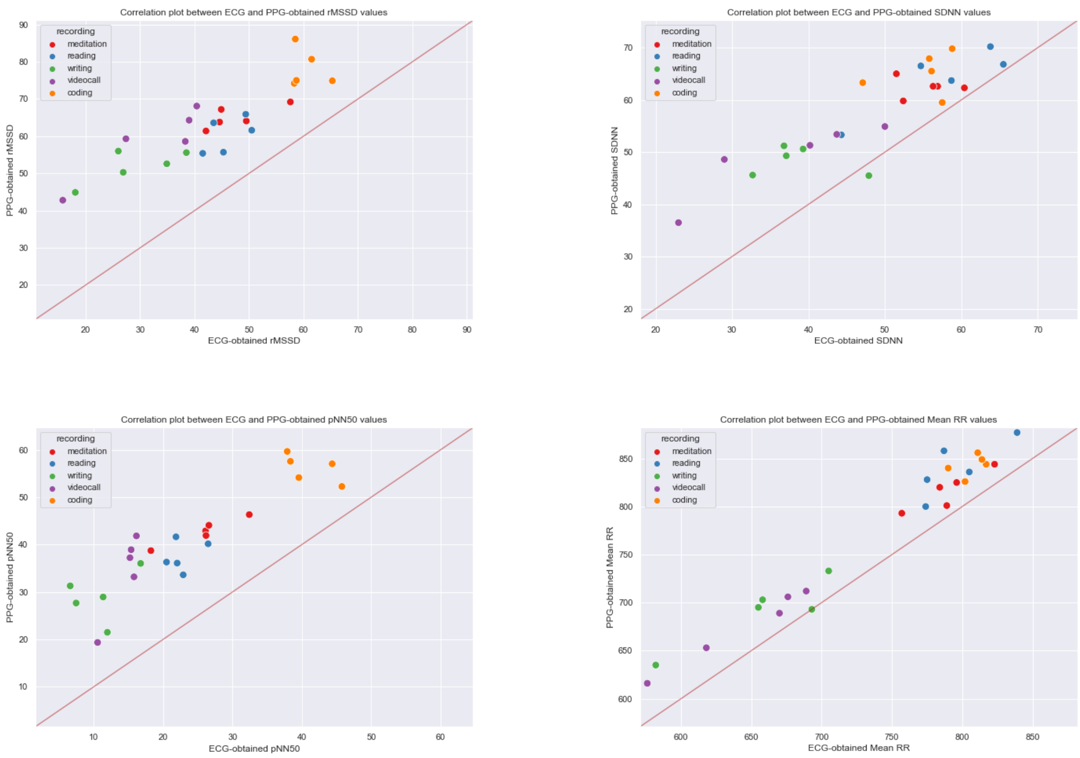

Figure 5.11: Correlation of HRV parameters derived from ECG and PPG datasets

The Bland-Altman plots visualizing HRV parameters from ECG and PPG data are illustrated in Fig. 5.12 and depict the comparison between mean values on the x-axis ([ECG + PPG]/2) with the difference of the values on the y-axis (PPG - ECG). Limits of agreement and Bias are outlined as horizontal lines. All plots depict that the values are spread uniformly horizontally between two-planes within limits of agreement with a good correlation of measurements, which suggests that there is a systematic error in the HRV values from PPG-obtained recordings. Furthermore, for the SDNN, Mean RRI and pNN50 HRV values Bland-Altman plots show data points that reside outside limits of agreement, which could indicate a potential noise.

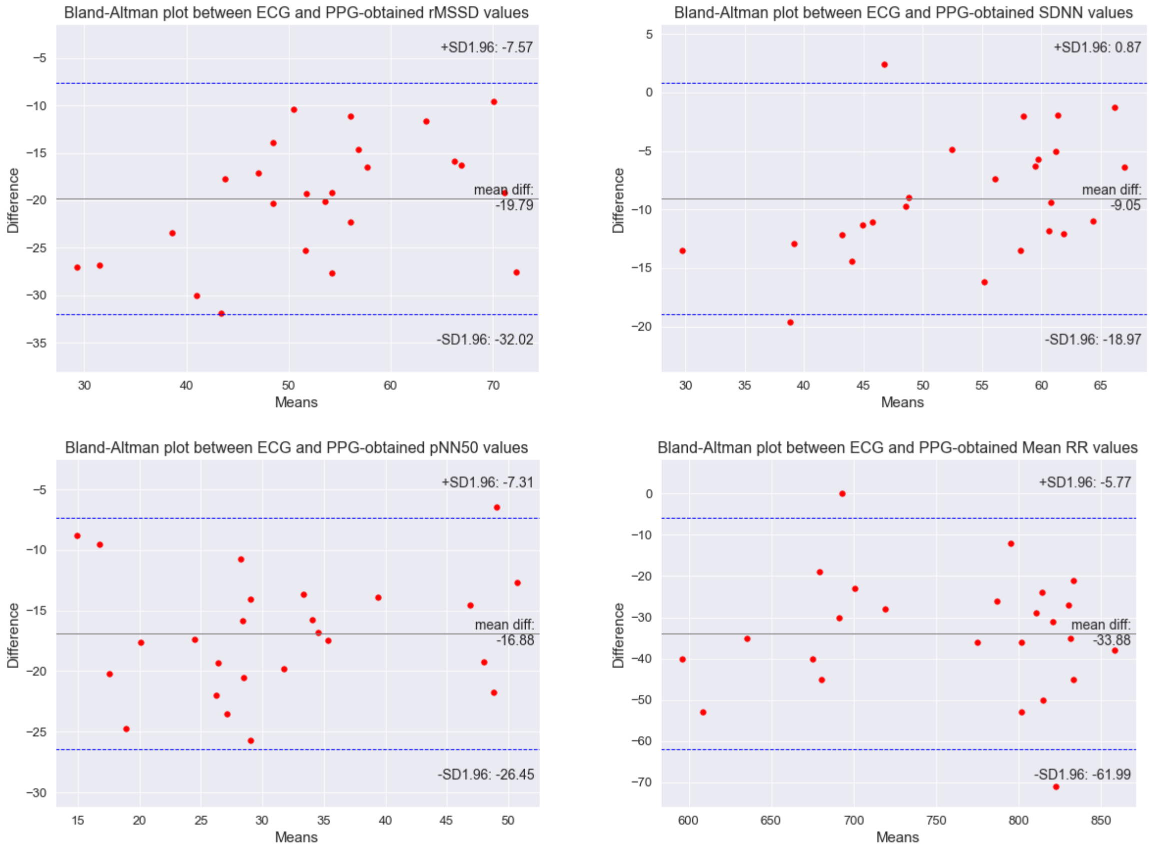

Figure 5.12: Bland-Altman plots for comparing HRV measurements from PPG with golden-standard ECG

Finally, the computed HRV values were assessed using a range of error metrics: RMSE, R2 score, Bias, Mean Absolute Error, Mean Absolute Percentage Error, Mean Relative Error and Spearman correlation coefficient. The obtained results are outlined in the Tables C.1 - C.4 (see Appendix).

The lowest RMSE values representing the precision of the results between ECG and PPG HRV were observed for the reading recording across all derived HRV parameters except for Mean RR. In particular, SDNN values for this recording were having the minimum RMSE rate equal to 7.588 (see Table C.2). The highest discrepancy between ECG and PPG values according to the RMSE were observed for the videocall recording.

As for the R^2 score error metric, which defines the accuracy of the compared values to the gold-standard ones, it was revealed that meditation recording rMSSD and pNN50 values possess the lowest R^2 score, while reading recording showed the most accurate values for SDNN parameter. It is worth to be noted that R2 score across all four computed HRV parameters turned out to be significantly low, which signifies poor accuracy of the collected PPG data. The lowest R2 scores were identified in pNN50 values.

Evaluating Bias, MAE and MRE error metrics results we can conclude that the lowest absolute discrepancy between ECG and PPG values was noticed during the reading recorded activity.

Finally, the highest correlation between ECG and PPG collected data was observed during the meditation, videocall and reading by taking into account Spearman rho results for rMSSD, SDNN, Mean RR and pNN50 values.

In general, the conducted error evaluation suggests controversial insights with regard to the level of discrepancy between ECG and PPG data. Hence, to precisely understand whether the values are realistically deviating due to inaccuracy of the PPG device or the conformed gold-standard ECG device is unreliable to be used as a benchmark, the data collection experiment had to be repeated with an alternative ECG device.

5.6 Repeated data acquisition

In order to conform the reliability of the Scosche Rhythm24 PPG device with regard to the gold-standard ECG measurements, the data collection process was repeated using additional device Polar H10 Heart Rate Sensor, representing a benchmark. This device was proved to be an accurate gold-standard ECG data recorder in a wide range of previous studies [8, 22, 106].

Figure 5.13: Polar H10 Heart Rate Sensor

Polar H10 Heart Rate Sensor is a wearable chest strap equipped with an ECG sensor and internal memory for heart rate data from one training session. Device has an operating time of 400 hours and enables simultaneous connection with sports watches, smartwatches, fitness equipment and devices using Bluetooth Low Energy and ANT+ technology. ECG sensor has a 5000 Hz transmission, that allows non-interrupted recordings in the water. Aside from raw ECG signal, Polar H10 additionally supports the recording of timestamped inter-beat interval data [107].

Data acquisition was performed during five separate intervals within the same day. Every recording consisted of 65 minutes duration repeated activity. By simultaneously utilizing Scosche Rhythm24 and Polar H10 devices, data was collected throughout similar predefined set of tasks: yoga, reading, writing, participating in a video-call and coding session.

To facilitate the data extraction process, both devices were connected to the computer running a custom Node.js script for real-time R-R intervals data acquisition. Extracted data from the sensors was stored in a CSV format files.

5.7 HRV analysis of the repeated experiment data

As in the previous experiment, collected inter-beat interval data from ECG and PPG devices was used for computing HRV parameters. Using the Kubios HRV software, first 30 minutes of the whole recording time-series with R-R intervals values were split into 5-min time-windows to calculate the gold-standard HRV time-domain features: rMSSD, SDNN, pNN50 and Mean RRI.

Similarly, the general exploratory analysis of the ECG and PPG data was carried out using frequency distribution histograms. In addition, the frequency distribution of each 5-min inter-beat interval series sample for every recording type and activity was visualized. Distributions of every dataset were also checked for normality using Anderson-Darling test. To understand the level of discrepancy between R-R interval values belonging to ECG and PPG datasets, the Kolmogorov-Smirnov non-parametric test was applied both to the fully recorded datasets and 5-min time-window samples that were used to compute HRV values. Furthermore, each of the 5-min data sample corresponding to respective ECG and PPG recordings was fit into an extensive set of continuous probability distributions from the Python *SciPy* package [2]. The best fitting distributions were evaluated with regard to Goodness of Fit metric using Kolmogorov-Smirnov statistical test.

As a next step before starting the error evaluation of the HRV metrics, every 5-min sample of inter-beat intervals from ECG and PPG datasets was statistically explored to contain any outliers. Firstly, every 5-min data sample corresponding to ECG was used to compute the inter quantile range (IQR) of the data distribution. IQR was computed as a difference between the 75th and the 25th percentiles of the data. Afterwards, the calculated IQR was utilized to identify the outliers within ECG data sample by defining limits on the sample values that equal to factor 1.5 of the IQR below the 25th percentile or above the 75th percentile. Having detected the outliers within the ECG samples as a reference, we proceeded with outlier detection among PPG samples. For this purpose, a One-Class Classification model was trained on the partition of every PPG-recorded sample [108].

A one-class classifier was trained using only instances from the normal class in its training dataset. Once prepared, the model was used to categorize additional samples as normal or abnormal. As a one-class classification method we utilized *local outlier factor* (LOF) technique. LOF is an approach that strives to use the concept of nearest neighbors to find outliers. Based on the size of its local neighborhood, each instance is assigned with a score indicating how isolated or probable it is to be an outlier. Outliers are more likely to be found in the instances with the highest scores [109]. Considering the aforementioned method, outliers that were identified within the PPG data samples using LOF trained model, were afterwards checked for the matter of overlapping with outliers detected using IQR within ECG data. In case the specific data instance in the PPG data sample was complying with all those conditions, it was considered as a true positive outlier and eliminated from the dataset.

The error evaluation between the HRV features was conducted using the same methodology as in the previous experiment. We utilized correlation plots to perform the visual exploratory analysis regarding the accurate alignment of the ECG and PPG derived HRV values. In order to support the findings from the visualizations, Spearman correlation coefficient was computed between ECG and PPG-obtained HRV values. To understand the limits of agreement between two HRV measurements Bland-Altman plots were constructed. The error evaluation between the HRV parameters was performed on the data samples before and after outlier removal procedure using Root mean square error (RMSE), R^2 score, Bias, Mean absolute error (MAE), Mean absolute percentage error (MAPE) and Mean relative error (MRE).

5.8 Repeated experiment results evaluation and discussion

Datasets collected through the ECG and PPG sensors recording were firstly visualized as a frequency distribution histograms to visually assess the amount of lost data in the PPG dataset as well as similarity of the distributions. In the Fig. 5.14 we can observe that data acquired during the repeated experiment is significantly more aligned with each other. Although the there is still a minor discrepancy in terms of certain amount of lost PPG-obtained data, in most of the cases the distributions have the same shape. The most reliable PPG recording based on the visualizations is the writing activity. This can be explained by the fact that during the writing recording there was no dynamic movements of the arm, which resulted in a low amount of noise within dataset as well as there was no performed movements around the place, which did not lead to lost sensor connection.

[2] "scipy: SciPy: Scientific Library for Python". Available at: https://www.scipy.org

Taking the aforementioned into account, we can assume that the inter-beat intervals data acquired through the PPG device becomes more reliable with longer duration of the recording (i.e. at least one hour) and during passive activities.

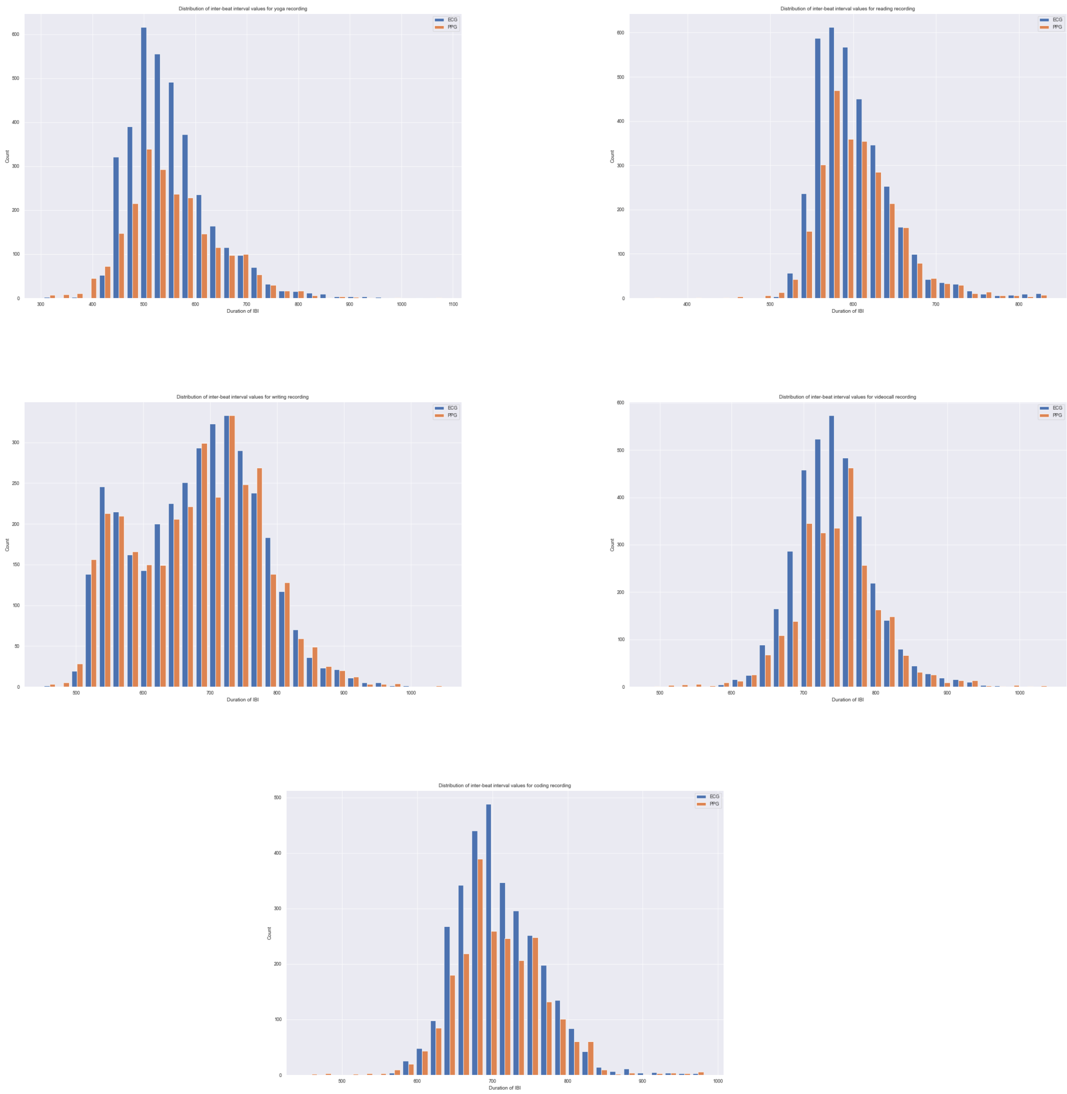

Figure 5.14: Frequency distribution for ECG and PPG recordings during repeated experiment

Despite the fact that long-duration PPG recordings have a high level of alignment, the HRV parameters are still being computed from the short 5-min time-windows. Hence, to understand whether the partitioned into 5-min intervals datasets are coherent in the same manner as the full datasets, distributions of the ECG and PPG 5-min samples were visualized (see Fig. C.11 - C.15). Observing the outlined distributions of short samples from every recording activity, we conclude that the PPG inter-beat intervals within the partitioned datasets have a moderate level of discrepancy with regard to the corresponding ECG R-R intervals.

To further proceed with the exploratory analysis of the raw R-R intervals from ECG and PPG recordings, the frequency distributions of ties (i.e. unique values) within every activity dataset were visualized (see Fig. C.16 - C.20). Similarly as in the previous experiment we conclude that both ECG and PPG recordings contain a significant amount of ties.

Following the methodology of the initial experiment the datasets were checked for normality using Anderson-Darling test. From the outlined plots in Fig. C.21 - C.25 we presume that none of the ECG nor PPG collected datasets have normal distribution.

Conducting a Kolmogorov-Smirnov test on the full datasets from the repeated experiment proved that data samples from ECG and PPG recording come from the same distribution, and thus are coherent (see Fig. 5.15). The reliability of the long-term recording via the PPG sensor within this analysis was confirmed by the fact that neither of the Kolmogorov-Smirnov tests run on the data allowed to reject the null hypothesis.

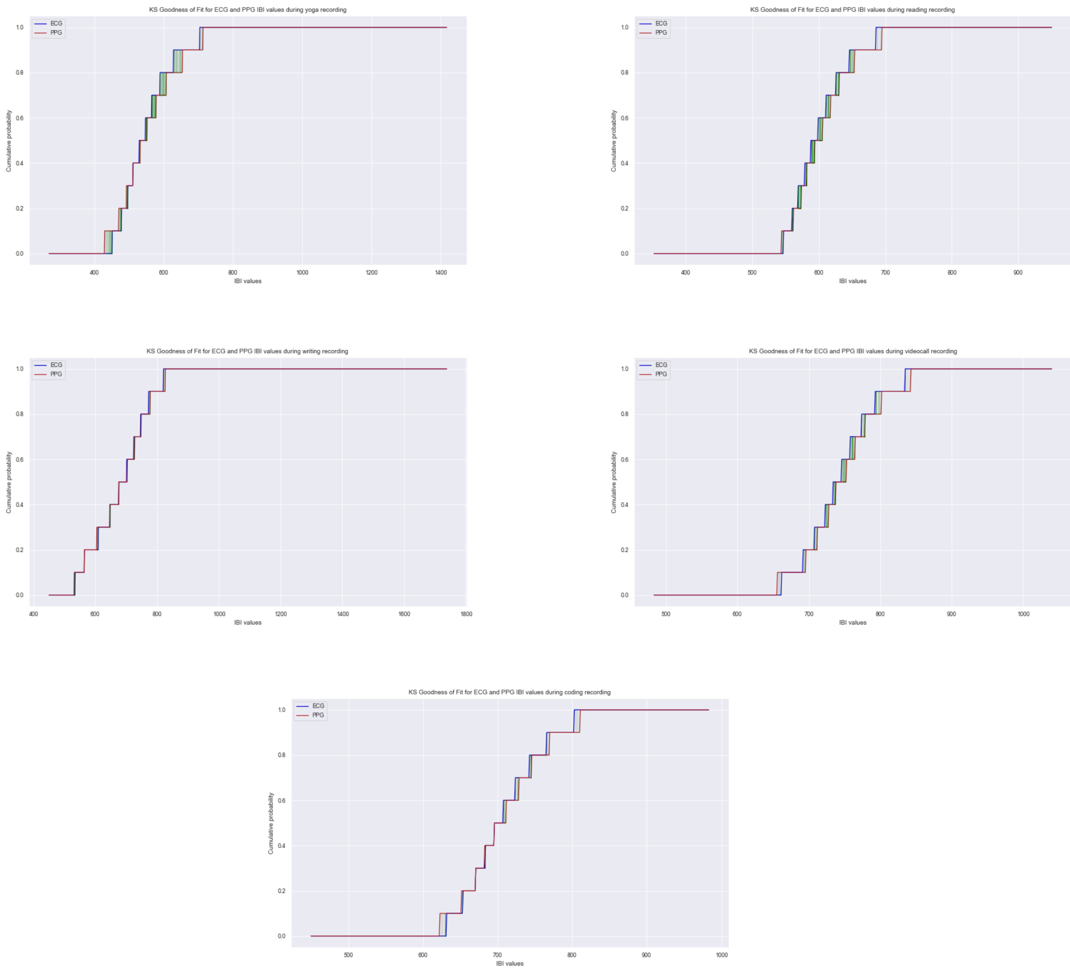

Figure 5.15: Cumulative probability plots for ECG and PPG datasets collected during repeated experiment

As a follow-up, we applied the Kolmogorov-Smirnov test on the partitioned datasets, each sample having duration of 5-min (see Fig. C.26 - C.30). Performing this statistical procedure confirmed our insights derived from the previous steps regarding the moderate level of discrepancy between the ECG and PPG inter-beat intervals for short duration samples. Hence, the outtake of this preliminary exploratory analysis is that inter-beat interval values from the PPG sensor can only be of sufficient reliability in the case of long-duration recording.

To investigate the nature of each of the 5-min interval sample from every ECG and PPG-obtained datasets during several types of activity recording, the data was assessed on the matter of best fitting distribution. The results of the Goodness of Fit analysis are outlined in Tables C.5 and C.6 (see Appendix). As we can see, the most common types of the distribution both ECG and PPG data samples possess are *Generalized Normal*, *Exponential Normal*, *Johnson SU*, *Double Weibull* and *Logistic*.

After the general assessment of the inter-beat intervals data, the HRV values were computed from the activity datasets. Scatter plots illustrated in Fig. 5.16 depict the correlation between different HRV metrics derived from the repeated experiment collected ECG and PPG data. As in the previous experiment, HRV values derived from PPG are slightly higher than those derived from ECG data. Overall, all PPG-based HRV parameters except for pNN50 seem to be better aligned with gold-standard ECG ones, which increases the reliability of the Scosche Rhythm24 Heart Monitor. The lowest discrepancy between ECG and PPG values was observed in the Mean RRI HRV metric, and the highest discrepancy was noticed between the pNN50 HRV values, as the values formed a cluster and any trend of growth is absent.

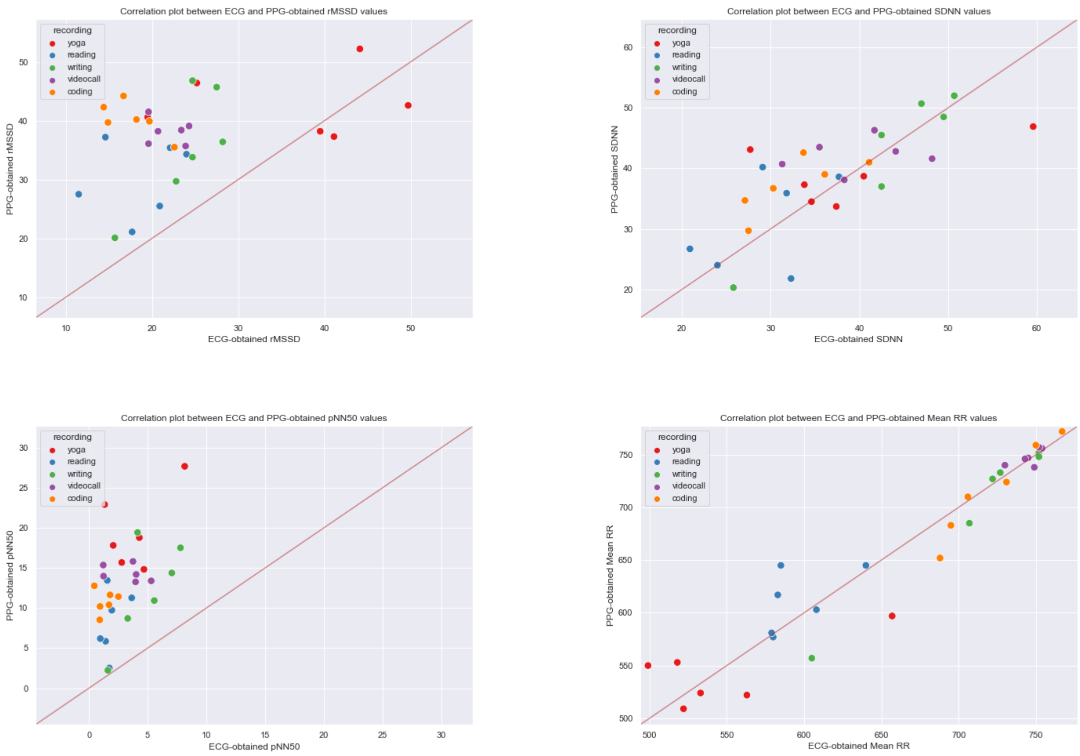

Figure 5.16: Correlation of HRV parameters derived from ECG and PPG datasets

According to the Bland-Altman plots, the lowest Bias between ECG and PPG data was present in the SDNN and Mean RR HRV values, and equaled to -1.66 and 1.1 respectively (see Fig. 5.17). As for the pNN50 and rMSSD values, their distribution across planes is not uniformal. Furthermore, there was noticed a trend in pNN50 values Bias increase from 0 to -20 as the average increases. Overall, the Bias between the values is significantly lower for the data acquired during the repeated data collection.

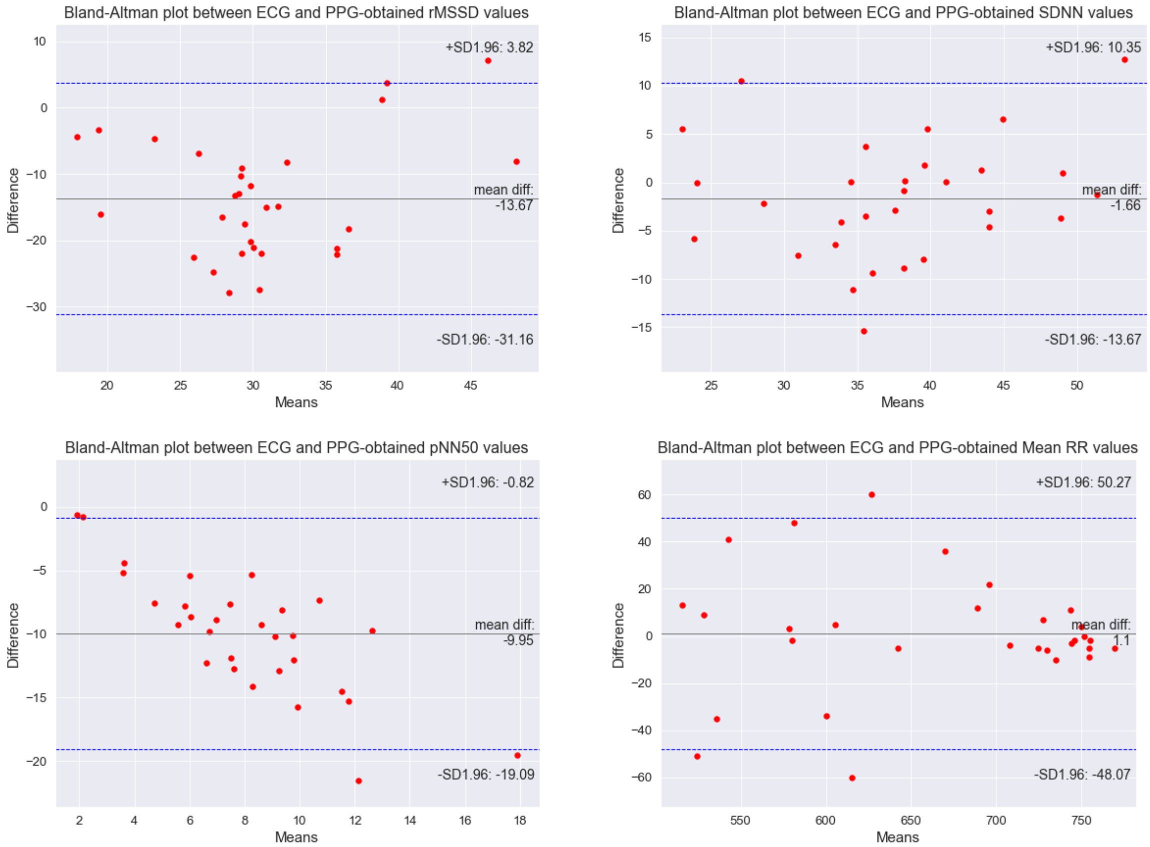

Figure 5.17: Bland-Altman plots for comparing HRV measurements from PPG with golden-standard ECG within repeated experiment

The error evaluation between the ECG and PPG data collected during the repeated experiment is outlined in the Table C.7 - C.10 (see Appendix).

Generally analyzing the error metrics results across multiple HRV parameter values we can conclude that the repeated data collection revealed to be significantly *more aligned* with regard to the ECG than the initial one, with Aidlab used as a gold-standard ECG measurement device.

According to the RMSE error metric, the highest precision of the PPG data was noticed during the yoga recording within the computed rMSSD values, writing within SDNN values, videocall within the Mean RRI values and reading within pNN50 results. As for the accuracy of the obtained PPG measurements, the lowest discrepancy was observed during the writing recording according to R^2 score results within SDNN and Mean RRI values. Overall, aside from pNN50 results for R^2 score, the accuracy of the new measurement within the scope of the repeated experiment revealed to be considerably higher in comparison with the previous data acquisition.

The lowest inconsistency in absolute and relative values according to the Bias, MAE and MRE metrics was detected during yoga recording for rMSSD values, writing recording for SDNN, videocall recording for Mean RRI and reading recording for pNN50. The most correlated ECG and PPG datasets were recorded during writing activity, according to the Spearman's ρ metric across all HRV parameters.

As a next step, we performed the procedure for detecting and removing outliers from the PPG 5-min data samples using trained One-Class Classification model. The methodology of this step is described in the Section 5.7.

To visually assess the efficiency of the outlier elimination procedure, cleaned from noise 5-min data samples were visualized (see Fig. C.31 - C.35). We can observe that at the general level the data became more consistent, without any deviating patterns in the context of PPG values distribution. However, we can still notice the significant presence of the lost data values coming from the PPG sensor, especially for the moderately high range of R-R intervals. Hence, we conclude that for the values lying within the 50th percentile the values are semi-reliable, while for the edge ranges of inter-beat intervals the R-R intervals data can be considered reliable.

Following the reliability assessment methodology, the correlation plots for newly computed HRV values from the cleaned data were constructed. In the Fig. 5.18 we can clearly see that the same trend continues, having the highest coherence and lowest discrepancy in SDNN and Mean RRI HRV values. Graph containing data for the pNN50 HRV metric illustrates poor alignment between ECG and PPG values, as they form a cluster within 25th percentile of the ECG distribution. In addition, plot outlining Mean RR values conforms with the conclusion made during the exploratory analysis of the 5-min frequency distribution plots after outlier elimination. Thus, it can confirmed that values residing on the higher edge of the distribution are significantly more reliable, than those within 50th percentile, as they are more divergent from each other.

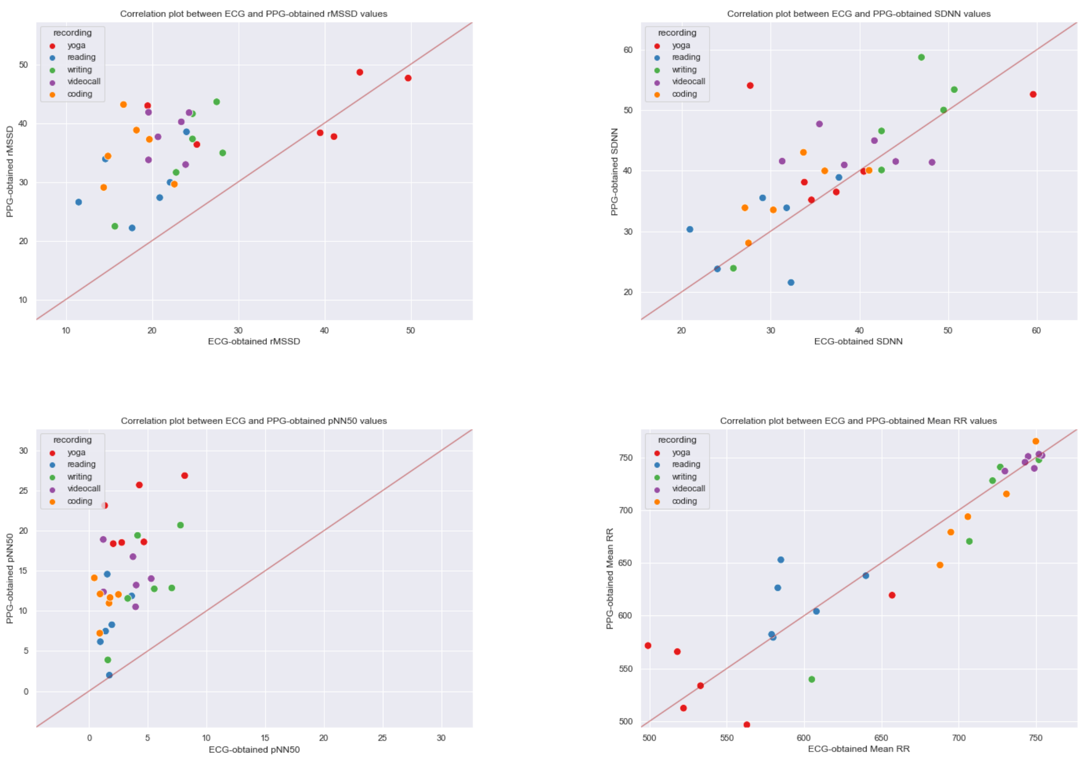

Figure 5.18: Correlation of HRV parameters derived from cleaned ECG and PPG datasets

For the data without outliers, Bland-Altman plots seem to be more consistent, as the overall Bias across all plots became lower (see Fig. 5.19). Plot visualizing pNN50 illustrates the familiar trend of the increased mean difference values as the average is increasing. Mean RRI and SDNN HRV metrics again prove to be the most accurate in characterizing the coherence between the ECG and PPG data, as their values are uniformly scattered across horizontal Bias line.

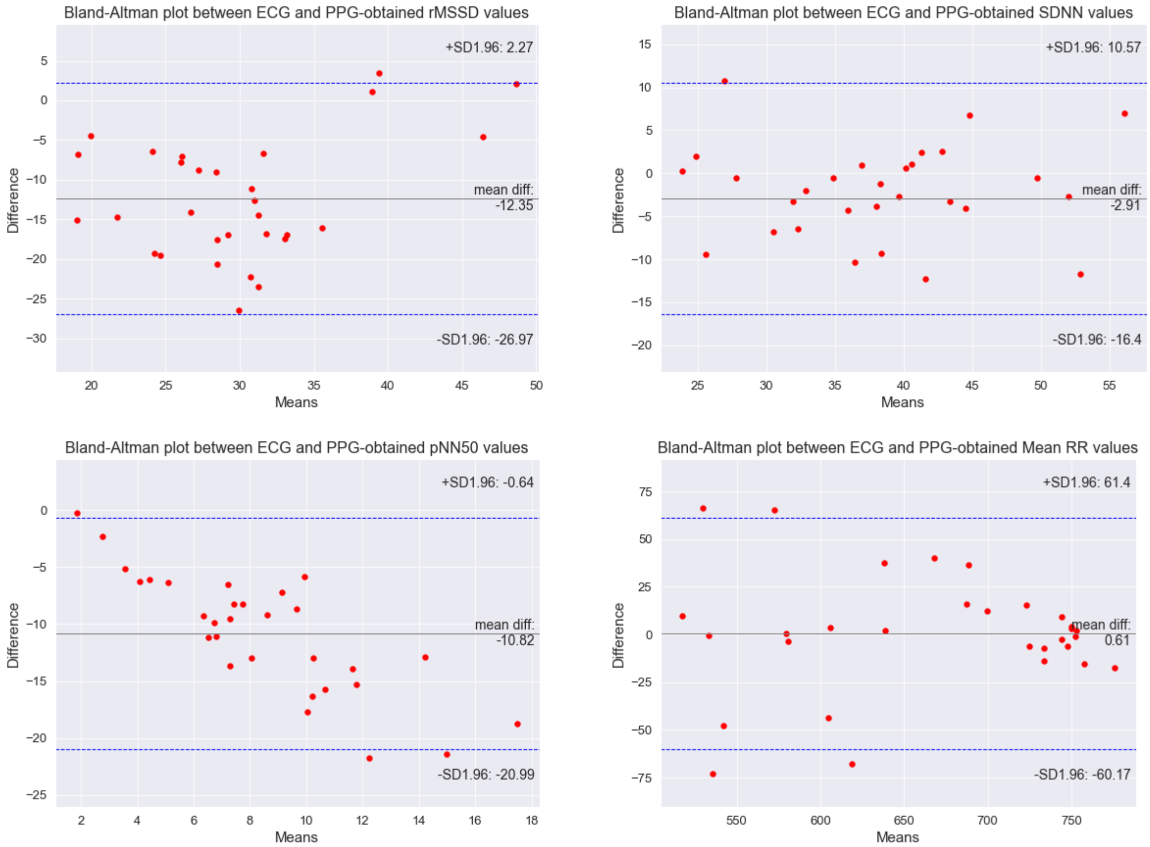

Figure 5.19: Bland-Altman plots for comparing HRV measurements from cleaned PPG datasets with golden-standard ECG within repeated experiment

Finally, the error evaluation using regression metrics was performed to statistically assess how significantly the quality of the data improved after outlier elimination (see Tables C.11 - C.14 in Appendix).

According to the RMSE metrics, we were not able to notice a significant improvement with regard to the precision of the computed HRV parameters from the cleaned PPG data. Similarly, the results of the R^2 score metric did not change considerably comparing with the PPG datasets containing outliers. Nevertheless, the accuracy of the HRV values derived from PPG still remained substantially high, except for the pNN50 parameter.

The absolute and relative errors illustrated by Bias, MAE and MRE showed that yoga recording contains the lowest discrepancy in PPG-obtained rMSSD data with regard to the gold-standard rMSSD. This can be explained by the fact that during the yoga recording session the sensor was located in a short distance to the data receiving device. As for the SDNN and pNN50 values, the lowest divergence of absolute and relative values was observed during reading and writing recordings. The most aligned HRV values according to Mean RRI were obtained during the videocall recording. Within the cleaned PPG datasets all derived HRV values are still highly correlated with the ECG gold-standard values. The highest correlation was observed during the writing recording across all HRV parameters.

Overall, taking into account the results obtained from the repeated experiment data analysis we observe that for the majority of the recorded datasets the regression metric results were meeting the requirements introduced earlier with regard to the accuracy threshold. We can conclude that the HRV values obtained from the Scosche ECG heart rate monitor can be considered to be *reliable* with regard to gold-standard ECG values obtained from Polar H10.

Chapter 6

Conclusions

This thesis work is primarily focused on exploring the constraints of the Heart Rate Variability data to understand the minimum requirements in terms of quality that affects the feasibility of using it for measuring the emotional state of the healthcare practitioners within the hospital setting. The proposed methodology helps to understand whether we are able to cope with the significant amount of data loss, if the specific device can be used for the accurate collection of physiological data and subsequent integration as an autonomous component into the intelligent system. Following the guidelines of the data-enabled design approach we were able to design an ecosystem that is sustainable and opens a room for extended researches. In addition, the implemented analytical methodology is reusable, hence the people who will be involved into the future data-enabled design researches would be able to replicate the analysis steps and gain insights from the data recorded in a similar way. Specifically, the work conducted within this project aimed at answering the three following research questions:

1. How the combination of physiological sensing and self-reflecting annotations can be utilized for emotional state and physiology analysis?
2. How much missing data and specific patterns of it can affect the reliability and validity of HRV measurements?
3. How to estimate the influence of the device on the reliability of the derived HRV features?

6.1 Contributions of the Thesis

The previously mentioned research questions were addressed by implementing three research solutions that together constitute a coherent systematic framework for Heart Rate Variability analysis. The first research solution is directed towards creating iterative prototypes that enable data-driven ecosystem for affective computing. The second one focuses on the extensive analysis of the missing data patterns and its affect on the accuracy of the HRV results. The third task was mainly aimed at verifying the reliability of using the PPG device as a main input source of physiological data layer within the previously designed ecosystem.

6.1.1 Ecosystem of micro-intelligences

To achieve the first research goal, inspired by the micro-service architecture the concept of micro-intelligences was introduced within Chapter 3. By incorporating the Data-Enabled Design approach, we were able to construct comprehensive design decisions that allowed to develop a system for collecting behavioural, contextual and experiential data. Based on the evaluation of four experiments conducted within the design research team's work-from-home context, their core characteristics were analyzed when designing and developing the experiments as preparation for a contextual DED phase.

Micro-intelligence approach enabled us to create separate, self-contained and autonomous layers, where every component is reusable and inter-changeable to boost the loose-coupling within the ecosystem. By focusing on every standalone data layer that refers to the behavioural, contextual or experiential aspect of the input, we were able to evaluate them individually. Moreover, following the micro-intelligence approach made it possible to design, test and refine modular components within the intelligent ecosystem as well as assess the effect of combining them in a pair-wise manner or as a consistent toolkit.

Addressing the need for subjective data input, we designed the affect annotation layer, that was comprising of post-hoc reflections and self-annotations from the participant side. However, as affect annotation layer was playing the role of supporting the insights with the ground-truth, the physiological layer needed more investigation to create a systematic methodology for personalized analysis.

For the aforementioned purpose, we utilized an open-source datasets as an extended source of input for our experiment. Visual analytics as well as statistical tests, such as violin plots, multi-line graphs of HRV dynamics when induced with multiple affect-stimuli, percentage change evaluation were performed to understand whether the Heart Rate Variability data can be a good estimator of human emotional state reflector and if there is any consistency in the derived insights. It was revealed that although the absolute values in the HRV trend are not aligning during inter-personal analysis, the general common trends in HRV changes still exist for the majority of the participants.

In the experimental use cases, we were able to demonstrate that by combining physiological signals data, self-annotations and post-hoc reflections it is feasible to design a full coherent system for conducting affective computing in a situated manner.

6.1.2 Missing data analysis

To address the second research question, the data collection was carried out using the PPG device during a set of activities that included a lot of movement. Performing this step allowed us to test in Chapter 4, whether this device is an appropriate choice for recording Heart Rate Variability data when the participant is involved in the dynamic actions.

As an insight during the data exploration, we discovered that once the intensity of the activities increases, the frequency of data loss starts to rise significantly. Additionally by computing the conditional data loss probabilities, it was revealed that the likeliness of loosing subsequent values in the case of the previous one being lost is proportionally higher across all collected datasets.

Further, the visual analytics and Spearman correlation test were performed to understand the correlation between the distribution of values corresponding to durations of preceding and following gaps for every activity dataset. It was concluded that there was no strong correlation between them. The general trend across all activity data sets encompassed short durations of the preceding gaps having a highly variable set of values corresponding to the following gap length, while long preceding gap durations tend to have shorter lengths of the following gaps. Moreover, taking into account that the heart rate of the participant involved in the experiment was never reaching below 60, we concluded that missing data occurrence was only caused by the lost signal or moving artifacts.

Within the research, the concept of missing data pattern was defined as a particular continuous probability distribution of aggregated missing gaps within the data set, that is generated not by a random mechanism.

By applying the Kolmogorov-Smirnov statistical test we were able to identify the most relevant missing data patterns corresponding to different activity datasets. The obtained continuous probability distributions were firstly visually examined for every collected dataset through the Cumulative Distribution Function and Kernel Density Estimate plots to understand the alignment between the empirical distribution of the real data and the one derived by the statistical test. It was confirmed that all the identified continuous probability distributions were consistent with the related empirical distributions.

The identified most fitting distributions were then used to generate *probability density functions* (pdf) that defined different scenarios at which missing data was introduced in the open-source dataset [90]. After computing Heart Rate Variability features from the amputed datasets, we utilized several regression metrics to evaluate the accuracy change, when induced with different missing data patterns. It was revealed that the there is a systematic error for the HRV values computed from the data with 25 % of missing rate and higher. Hence, we can only consider data to be reliable and usable if the rate of the missing data is below 25 %.

6.1.3 Device reliability evaluation

To achieve the third research goal, we performed a simultaneous data collection using PPG and gold-standard ECG heart rate monitors. As ECG sensor is considered to be the most accurate for conducting HRV measurements, it was chosen to be a benchmark for the PPG device reliability evaluation.

Before conducting the analysis on the Heart Rate Variability features, the raw ECG data was pre-processed and transformed to remove the moving artifacts, filter the existing noise in the signal and detect the R-peaks.

Further, the visual exploratory analysis was conducted to understand the level of discrepancy between the gold-standard ECG data and the collected PPG physiological signals. It was concluded that the overall amount of data coming from the PPG device was significantly lower, than the recorded data through the ECG device, although the general shape of the distributions corresponding to the ECG and PPG-obtained data were aligned. All matching datasets contained a lot of ties in the data and did not possess a normal probability distribution. After applying the Kolmogorov-Smirnov test on the datasets we concluded that only values from the two matching datasets out of five come from the same distribution.

Performed correlation analysis revealed that HRV values derived from the data collected by ECG and PPG devices were highly correlated. However, there was noticed a slight misalignment in distribution values for all types of computed HRV parameters except the Mean RRI. Furthermore, it was discovered that there is a systematic error in the Heart Rate Variability values computed from the PPG sensor recordings with regard to the ECG ones, according to the applied Bland-Altman statistical test.

Error estimation tests of the HRV features derived from the ECG and PPG revealed that although the precision of the PPG-computed Heart Rate Variability was quite high, the accuracy of those values was substantially low. It was concluded that the conducted error evaluation possessed controversial insights with regard to the level of discrepancy between the ECG and PPG Heart Rate Variability data, hence the repeated data collection experiment was carried out using the alternative ECG device.

Performing the visual analytics on the datasets from repeated experiments showed that the overall distributions of the values corresponding to ECG and PPG datasets were considerably aligned. Nonetheless, by comparing the distribution of values within within partitioned into 5-min interval samples it was deduced that the PPG-based inter-beat intervals values within the partitioned datasets had a moderate level of discrepancy with regard to the corresponding ECG R-R intervals.

Thus, to bridge the gap between the ECG and PPG values, the PPG datasets were used as an input for outlier detection algorithm. After performing this procedure, the data became more consistent and cleaned from any deviating patterns in the context of PPG values distribution.

Performing the repeated error estimation procedure on the HRV features computed from the cleaned PPG datasets, provided us with the insights that obtained precision and accuracy were quite high in this case, while the bias hence the level of discrepancy was significantly lower. As a result of this research, it was concluded that the data collected by the particular PPG device used in the experiment can be considered reliable with regard to the gold-standard ECG data.

To summarize, there are three *key novelties* introduced in this thesis, which will be covered next:

1. *Introducing a micro-intelligence approach for designing and testing the ecosystem for situated affective computing:* the prototype, implemented by following the Data-Enabled Design

guidelines, facilitated collecting behavioural, experimental and contextual data within multi-layered data acquisition and analysis architecture. The conducted experiments, discussed previously, allowed us to fine-tune our design decisions while also revealing experiential or functional shortcomings in our micro-intelligences. They enabled us to understand whether the designed settings of current study can be applicable for integration into a larger research.

2. *Making the missing data pattern analysis more specific with regard to the Heart Rate Variability context:* we have uncovered a new way of identifying the missing patterns within a univariate datasets comprised of inter-beat intervals values. By evaluating the obtained results of regression metrics from Heart Rate Variability values we were able to identify whether the recording, non-obtrusive device would be an adequate choice for data collection during highly dynamic activities.

3. *Validating the PPG sensor accuracy with regard to the two ECG devices:* we assessed the data reliability of a particular PPG device and proved its feasibility to be incorporated into an ecosystem of micro-intelligences with a purpose of collecting physiological signals.

6.2 Limitations and Future Work

The COVID-19 pandemic that was happening during the execution of this project led to some unprecedented limitations and challenges that remain open for the future work.

First of all, every experiment related to this research was conducted not in the hospital setting as was planned by the MEX Oncology research team at Philips Design, but within working-from-home context. Hence, the amount of data collected from the participants was limited and not comprehensive enough for the extensive analysis. As a future work, the designed ecosystem for situated affective computing has to be utilized for carrying out the real-life experiment involving healthcare professionals in the hospital environment.

Secondly, the missing data pattern identification process was performed on the limited amount of data collected from only one participant (i.e. team-member of the MEX Oncology project). Hence, for the future research, there is a need to conduct a study with a larger sample of people that will facilitate bringing new insights from the error estimation analysis. Furthermore, the effect of only discovered within our experiment missing data patterns was assessed with regard to the Heart Rate Variability accuracy. Thus, there is a room for the next studies to extend the set of evaluated missing data patterns representing different types of continuous probability distributions.

Lastly, the reliability of the data collection by the PPG device was assessed using only two alternative ECG sensors within a limited set of activities. Therefore, there is a scientific gap for a follow-up research that can be addressed by involving more than two ECG devices used as a benchmark for evaluation and carrying out long-term recordings during the whole working day.

Bibliography

[1] "FibriCheck Beat-to-Beat Accuracy Compared With Wearable ECG in Broad Dynamic Range - FibriCheck." [Online]. Available: https://www.fibricheck.com/nl/fibricheck-beat-to-beat-accuracy-compared-with-wearable-ecg-in-broad-dynamic-range/ ix, 9

[2] P. v. Gent, "heartpy: Heart Rate Analysis Toolkit." [Online]. Available: https://github.com/paulvangentcom/heartrate_analysis_python ix, ix, 16, 25, 51, 53

[3] S. P. Contributor, "Heart Rate Vs. Heart Rate Variability | South Pasadena Review." [Online]. Available: https://southpasadenareview.com/heart-rate-vs-heart-rate-variability/ 1

[4] R. K. Dishman, Y. Nakamura, M. E. Garcia, R. W. Thompson, A. L. Dunn, and S. N. Blair, "Heart rate variability, trait anxiety, and perceived stress among physically fit men and women," *International Journal of Psychophysiology*, vol. 37, no. 2, pp. 121–133, Aug. 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167876000000854 1, 7, 11, 19

[5] J. V. Kollenburg and S. Bogers, "A situated design approach that uses data as creative material when designing for intelligent ecosystems," p. 342. 2, 3, 20

[6] A. F. Hussein, N. A. kumar, M. Burbano-Fernandez, G. Ramírez-González, E. Abdulhay, and V. H. C. De Albuquerque, "An Automated Remote Cloud-Based Heart Rate Variability Monitoring System," *IEEE Access*, vol. 6, pp. 77 055–77 064, 2018, conference Name: IEEE Access. 2

[7] L. Haoyu, L. Jianxing, N. Arunkumar, A. F. Hussein, and M. M. Jaber, "An IoMT cloud-based real time sleep apnea detection scheme by using the SpO2 estimation supported by heart rate variability," *Future Generation Computer Systems*, vol. 98, pp. 69–77, Sep. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X18326980 2

[8] J. Zhou, J. Riekki, W. Zhang, and T. Qiu, "iHRV: Cloud-Based Mobile Heart Rate Variability Monitoring System," Dec. 2016, pp. 328–333. 2, 60

[9] G. Ernst, "Hidden Signals—The History and Methods of Heart Rate Variability," *Frontiers in Public Health 5:265*, 2017, publisher: Frontiers. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpubh.2017.00265/full 3

[10] F. Shaffer and J. Ginsberg, "An Overview of Heart Rate Variability Metrics and Norms," *Frontiers in Public Health*, vol. 5, p. 258, Sep. 2017. 7, 8

[11] R. McCraty and F. Shaffer, "Heart Rate Variability: New Perspectives on Physiological Mechanisms, Assessment of Self-regulatory Capacity, and Health risk," *Global Advances in Health and Medicine*, vol. 4, no. 1, pp. 46–61, Jan. 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4311559/ 7

[12] T. A. VanDusseldorp and L. Kravitz, "Heart Rate Variability and Overtraining," *IDEA Fitness Journal*, p. 8, Jan. 2015. 7

[13] L. Mourot, M. Bouhaddi, S. Perrey, S. Cappelle, M.-T. Henriet, J.-P. Wolf, J.-D. Rouillon, and J. Regnard, "Decrease in heart rate variability with overtraining: assessment by the Poincaré plot analysis," *Clinical Physiology and Functional Imaging*, vol. 24, no. 1, pp. 10–18, Jan. 2004. 7

[14] K. Li, H. Rüdiger, and T. Ziemssen, "Spectral Analysis of Heart Rate Variability: Time Window Matters," *Frontiers in Neurology*, vol. 10, p. 545, 2019. 7, 10

[15] F. T. Ltd., "Stress and Recovery Analysis Method Based on 24-hour Heart Rate Variability," 2014. [Online]. Available: https://assets.firstbeat.com/firstbeat/uploads/2015/11/Stress-and-recovery_white-paper_20145.pdf 7, 12

[16] G. Lu, F. Yang, J. A. Taylor, and J. F. Stein, "A comparison of photoplethysmography and ECG recording to analyse heart rate variability in healthy subjects," *Journal of Medical Engineering & Technology*, vol. 33, no. 8, pp. 634–641, 2009. 7, 8, 9

[17] A.-F. The, I. Reijmerink, M. van der Laan, and F. Cnossen, "Heart rate variability as a measure of mental stress in surgery: a systematic review," *International Archives of Occupational and Environmental Health*, vol. 93, no. 7, pp. 805–821, Oct. 2020. 8, 11, 12

[18] A. Sutarto, N. Wahab, and N. Mat Zin, "Heart Rate Variability (HRV) biofeedback: A new training approach for operator's performance enhancement," *Journal of Industrial Engineering and Management*, vol. 3, Jun. 2010. 8, 11, 13

[19] H.-Y. Jan, M.-F. Chen, T.-C. Fu, W.-C. Lin, C.-L. Tsai, and K.-P. Lin, "Evaluation of Coherence Between ECG and PPG Derived Parameters on Heart Rate Variability and Respiration in Healthy Volunteers With/Without Controlled Breathing," *Journal of Medical and Biological Engineering*, vol. 39, no. 5, pp. 783–795, Oct. 2019. [Online]. Available: https://doi.org/10.1007/s40846-019-00468-9 8, 9

[20] H. Kinnunen, A. Rantanen, T. Kenttä, and H. Koskimäki, "Feasible assessment of recovery and cardiovascular health: accuracy of nocturnal HR and HRV assessed via ring PPG in comparison to medical grade ECG," *Physiological Measurement*, vol. 41, no. 4, p. 04NT01, May 2020. 9, 50

[21] W.-H. Lin, D. Wu, C. Li, H. Zhang, and Y.-T. Zhang, "Comparison of Heart Rate Variability from PPG with That from ECG," in *The International Conference on Health Informatics*, ser. IFMBE Proceedings, Y.-T. Zhang, Ed. Cham: Springer International Publishing, 2014, pp. 213–215. 9

[22] D. J. Plews, B. Scott, M. Altini, M. Wood, A. E. Kilding, and P. B. Laursen, "Comparison of Heart-Rate-Variability Recording With Smartphone Photoplethysmography, Polar H7 Chest Strap, and Electrocardiography," *International Journal of Sports Physiology and Performance*, vol. 12, no. 10, pp. 1324–1328, Nov. 2017. 10, 11, 60

[23] J. L. Hagad and K.-i. Fukui, "Modelling Naturalistic Work Stress Using Spectral HRV Representations and Deep Learning | SpringerLink." [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-39878-1_24 10

[24] K.-C. Lan, P. Raknim, W.-F. Kao, and J.-H. Huang, "Toward Hypertension Prediction Based on PPG-Derived HRV Signals: a Feasibility Study," *Journal of Medical Systems*, vol. 42, no. 6, p. 103, Apr. 2018. 10

[25] T. Bhowmik, J. Dey, and V. N. Tiwari, "A novel method for accurate estimation of HRV from smartwatch PPG signals," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2017, pp. 109–112, Jul. 2017. 11

[26] M. Wu, H. Cao, H.-L. Nguyen, K. Surmacz, and C. Hargrove, "Modeling perceived stress via HRV and accelerometer sensor streams," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2015, pp. 1625–1628, Aug. 2015. 11

[27] V. Jindal, "Integrating mobile and cloud for PPG signal selection to monitor heart rate during intensive physical exercise," in *Proceedings of the International Conference on Mobile Software Engineering and Systems*, ser. MOBILESoft '16. New York, NY, USA: Association for Computing Machinery, May 2016, pp. 36–37. [Online]. Available: https://doi.org/10.1145/2897073.2897132 11

[28] "Heart Rate Variability Biofeedback Improves Emotional and Physical Health and Performance: A Systematic Review and Meta Analysis | Request PDF." [Online]. Available: https://www.researchgate.net/publication/341254440_Heart_Rate_Variability_Biofeedback_Improves_Emotional_and_Physical_Health_and_Performance_A_Systematic_Review_and_Meta_Analysis 11

[29] S. Koldijk, M. Sappelli, and S. Verberne, "The SWELL Knowledge Work Dataset for Stress and User Modeling Research | Proceedings of the 16th International Conference on Multimodal Interaction," 2014. [Online]. Available: https://dl.acm.org/doi/10.1145/2663204.2663257 12

[30] A. Oskooei, S. M. Chau, J. Weiss, A. Sridhar, M. R. Martínez, and B. Michel, "DeStress: Deep Learning for Unsupervised Identification of Mental Stress in Firefighters from Heart-rate Variability (HRV) Data," *arXiv:1911.13213 [cs, eess, stat]*, Nov. 2019, arXiv: 1911.13213. [Online]. Available: http://arxiv.org/abs/1911.13213 12

[31] L. Schmitt, J. Regnard, M. Desmarets, F. Mauny, L. Mourot, J.-P. Fouillot, N. Coulmy, and G. Millet, "Fatigue shifts and scatters heart rate variability in elite endurance athletes," *PloS One*, vol. 8, no. 8, p. e71588, 2013. 12

[32] E. Butkevičiūtė, M. Eriņš, and L. Bikulčienė, "Mobile Platform for Fatigue Evaluation: HRV Analysis," in *Information and Software Technologies*, ser. Communications in Computer and Information Science, R. Damaševičius and G. Vasiljevienė, Eds. Cham: Springer International Publishing, 2019, pp. 541–552. 12

[33] V. Balakarthikeyan, A. Sebastin, P. Sp, J. Joseph, and M. Sivaprakasam, *HRV based Stress Assessment of Individuals in a Work Environment*, Jun. 2020, pages: 6. 12

[34] J. B. Lemaire, J. E. Wallace, A. M. Lewin, J. de Grood, and J. P. Schaefer, "The effect of a biofeedback-based stress management tool on physician stress: a randomized controlled clinical trial," *Open Medicine: A Peer-Reviewed, Independent, Open-Access Journal*, vol. 5, no. 4, pp. e154–163, 2011. 13

[35] G. Baselli, S. Cerutti, S. Civardi, F. Lombardi, A. Malliani, M. Merri, M. Pagani, and G. Rizzo, "Heart rate variability signal processing: A quantitative approach as an aid to diagnosis in cardiovascular pathologies," *International Journal of Bio-Medical Computing*, vol. 20, no. 1, pp. 51–70, Jan. 1987. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0020710187900146 13

[36] G. D. Clifford, "Signal Processing Methods for Heart Rate Variability," Ph.D. dissertation, St. Cross College, University of Oxford, 2002. 13, 14

[37] M. Bolanos, H. Nazeran, and E. Haltiwanger, "Comparison of heart rate variability signal features derived from electrocardiography and photoplethysmography in healthy individuals," *Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2006, pp. 4289–4294, 2006. 14

[38] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985. [Online]. Available: http://ieeexplore.ieee.org/document/4122029/ 14, 52

[39] G. Moody and R. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, Jun. 2001. [Online]. Available: http://ieeexplore.ieee.org/document/932724/ 14

[40] R. A. Álvarez, A. J. M. Penín, and X. A. V. Sobrino, "A Comparison of Three QRS Detection Algorithms Over a Public Database," *Procedia Technology*, vol. 9, pp. 1159–1165, Jan. 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212017313002831 14

[41] P. S. Hamilton and W. J. Tompkins, "Quantitative Investigation of QRS Detection Rules Using the MIT/BIH Arrhythmia Database," *IEEE Transactions on Biomedical Engineering*, vol. BME-33, no. 12, pp. 1157–1165, Dec. 1986, conference Name: IEEE Transactions on Biomedical Engineering. 14

[42] J. Kim and H. Shin, "Simple and Robust Realtime QRS Detection Algorithm Based on Spatiotemporal Characteristic of the QRS Complex," *PLOS ONE*, vol. 11, no. 3, p. e0150144, 2016, publisher: Public Library of Science. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150144 14

[43] "Ectopic heartbeat: Causes, treatment, and risks," Sep. 2018. [Online]. Available: https://www.medicalnewstoday.com/articles/323202 14

[44] M. Malik, "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use," *Eur Heart J*, vol. 17, no. American Heart Association Inc., European Society of Cardiology, p. 28, 1996. 14

[45] H. Mølgaard, "Evaluation of the Reynolds Pathfinder II system for 24 h heart rate variability analysis," *European Heart Journal*, vol. 12, no. 11, pp. 1153–1162, Nov. 1991. 14

[46] M. Karlsson, R. Hörnsten, A. Rydberg, and U. Wiklund, "Automatic filtering of outliers in RR intervals before analysis of heart rate variability in Holter recordings: a comparison with carefully edited data," *BioMedical Engineering OnLine*, vol. 11, p. 2, Jan. 2012. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3268104/ 14

[47] F. Wen and F.-t. He, "An efficient method of addressing ectopic beats: new insight into data preprocessing of heart rate variability analysis," *Journal of Zhejiang University. Science. B*, vol. 12, no. 12, pp. 976–982, Dec. 2011. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3232430/ 15

[48] M. A. Chikh and F. Bereksi–Reguig, "Classification of ventricular ectopic beats (veb's) using neural networks," *Journal of Mechanics in Medicine and Biology*, vol. 04, no. 03, pp. 333–340, Sep. 2004, publisher: World Scientific Publishing Co. [Online]. Available: https://www.worldscientific.com/doi/abs/10.1142/S0219519404001089 15

[49] V. Pichot, F. Roche, S. Celle, J.-C. Barthélémy, and F. Chouchou, "HRVanalysis: A Free Software for Analyzing Cardiac Autonomic Activity," *Frontiers in Physiology*, vol. 7, p. 557, Nov. 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5118625/ 15

[50] R. Bartels, L. Neumamm, T. Peçanha, and A. R. S. Carvalho, "SinusCor: an advanced tool for heart rate variability analysis," *BioMedical Engineering OnLine*, vol. 16, no. 1, p. 110, Sep. 2017. [Online]. Available: https://doi.org/10.1186/s12938-017-0401-4 15

[51] "Kubios hrv analysis software - heart rate variability." [Online]. Available: https://www.kubios.com/ 15

[52] M. P. Tarvainen, J.-P. Niskanen, J. A. Lipponen, P. O. Ranta-Aho, and P. A. Karjalainen, "Kubios HRV–heart rate variability analysis software," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 210–220, 2014. 15

[53] T. Kaufmann, S. Sütterlin, S. M. Schulz, and C. Vögele, "ARTiiFACT: a tool for heart rate artifact processing and heart rate variability analysis," *Behavior Research Methods*, vol. 43, no. 4, pp. 1161–1170, Dec. 2011. [Online]. Available: https://doi.org/10.3758/s13428-011-0107-7 15

[54] R. Champseix, "hrv-analysis: a package to calculate features from Rr Interval for HRV analyses." [Online]. Available: https://github.com/Aura-healthcare/hrvanalysis 16

[55] R. Bartels, "hrv: A python package for heart rate variability analysis." [Online]. Available: https://github.com/rhenanbartels/hrv 16

[56] P. Gomes, "pyHRV - OpenSource Python Toolbox for Heart Rate Variability Documentation," p. 162. 16

[57] K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, and F. Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," *Scientific Data*, vol. 6, no. 1, p. 196, Oct. 2019. [Online]. Available: https://www.nature.com/articles/s41597-019-0209-0 17, 25, 28

[58] R. W. Picard, "Affective computing: challenges," *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 55–64, 2003. 17

[59] J.-M. Jansen, K. Niemantsverdriet, A. W. Burghoorn, P. Lovei, I. Neutelings, E. Deckers, and S. Nienhuijs, "Design for Co-responsibility: Connecting Patients, Partners, and Professionals in Bariatric Lifestyle Changes," in *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. New York, NY, USA: Association for Computing Machinery, Jul. 2020, pp. 1537–1549. [Online]. Available: https://doi.org/10.1145/3357236.3395469 17, 18

[60] S. Bogers, J. v. Kollenburg, E. Deckers, J. Frens, and C. Hummels, "A Situated Exploration of Designing for Personal Health Ecosystems through Data-enabled Design," ser. the 2018 DIS conference, 2018, pp. 109 – 120. 17

[61] P. Lovei, E. Deckers, M. Funk, and S. Wensveen, "The marios and luigis of design: Design plumbers wanted!" in *Companion Publication of the 2020 ACM Designing Interactive Systems Conference*, 2020, pp. 197–201. 17, 18

[62] S. Bogers, J. Frens, J. Kollenburg, E. Deckers, and C. Hummels, *Connected Baby Bottle: A Design Case Study Towards a Framework for Data-Enabled Design*, Jun. 2016, pages: 311. 17

[63] P. Lovei, E. Deckers, M. Funk, and S. Wensveen, "Field lab sleep and energy: A system for longitudinal remote sleep tracking and prototyping," in *CHIIoT 1*, 2021. 17

[64] CHIIoT 2 - 2nd workshop on computer human interaction in iot applications. [Online]. Available: https://chiiot.org/ 18

[65] J. Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel, "Influence of Mental Stress on Heart Rate and Heart Rate Variability," in *4th European Conference of the International Federation for Medical and Biological Engineering*, ser. IFMBE Proceedings, J. Vander Sloten, P. Verdonck, M. Nyssen, and J. Haueisen, Eds. Berlin, Heidelberg: Springer, 2009, pp. 1366–1369. 19

[66] A. Uusitalo, T. Mets, K. Martinmäki, S. Mauno, U. Kinnunen, and H. Rusko, "Heart rate variability related to effort at work," *Applied Ergonomics*, vol. 42, no. 6, pp. 830–838, Nov. 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003687011000147 19

[67] A. Rieger, R. Stoll, S. Kreuzfeld, K. Behrens, and M. Weippert, "Heart rate and heart rate variability as indirect markers of surgeons' intraoperative stress," *International Archives of Occupational and Environmental Health*, vol. 87, no. 2, pp. 165–174, Feb. 2014. [Online]. Available: https://doi.org/10.1007/s00420-013-0847-z 19

[68] D. Namiot and M. Sneps-Sneppe, "On micro-services architecture," *International Journal of Open Information Technologies*, vol. 2, no. 9, pp. 24–27, 2014. 20

[69] A. S. Perrotta, A. T. Jeklin, B. A. Hives, L. E. Meanwell, and D. E. Warburton, "Validity of the elite hrv smartphone application for examining heart rate variability in a field-based setting," *The Journal of Strength & Conditioning Research*, vol. 31, no. 8, pp. 2296–2302, 2017. 20, 33, 54

[70] H. J. Baek and J. Shin, "Effect of Missing Inter-Beat Interval Data on Heart Rate Variability Analysis Using Wrist-Worn Wearables," *Journal of Medical Systems*, vol. 41, no. 10, p. 147, Aug. 2017. [Online]. Available: https://doi.org/10.1007/s10916-017-0796-2 21, 37

[71] D. C. Sheridan, R. Dehart, A. Lin, M. Sabbaj, and S. D. Baker, "Heart Rate Variability Analysis: How Much Artifact Can We Remove?" *Psychiatry Investigation*, vol. 17, no. 9, pp. 960–965, Sep. 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7538246/ 21, 37

[72] J. Wrobel, D. Roj, J. Jezewski, K. Horoba, T. Kupka, and M. Jezewski, "Evaluation of the Robustness of Fetal Heart Rate Variability Measures to Low Signal Quality," *Journal of Medical Imaging and Health Informatics*, vol. 5, no. 6, pp. 1311–1318, Nov. 2015. 21, 37

[73] K. Sharma, C. Castellini, E. L. van den Broek, A. Albu-Schaeffer, and F. Schwenker, "A dataset of continuous affect annotations and physiological signals for emotion analysis," *Scientific Data*, vol. 6, no. 1, p. 196, Oct. 2019. [Online]. Available: https://www.nature.com/articles/s41597-019-0209-0 23

[74] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, ser. ICMI '18. New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 400–408. [Online]. Available: https://doi.org/10.1145/3242969.3242985 24, 25

[75] C. Kirschbaum, K. M. Pirke, and D. H. Hellhammer, "The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1-2, pp. 76–81, 1993. 24

[76] H.-G. Kim, E.-J. Cheon, D.-S. Bai, Y. H. Lee, and B.-H. Koo, "Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature," *Psychiatry Investigation*, vol. 15, no. 3, pp. 235–245, Mar. 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5900369/ 25, 28

[77] Y. Li, J. Wang, J. Li, W. Xu, P. Zhang, and X. Zhang, "Detecting complexity changes for heart rate variability in the case of missing data," *SCIENTIA SINICA Informationis*, vol. 45, no. 8, pp. 1015–1024, Aug. 2015, publisher: Science China Press. [Online]. Available: /doi/10.1360/N112014-00112 37

[78] K. K. Kim, Y. G. Lim, J. S. Kim, and K. S. Park, "Effect of missing RR-interval data on heart rate variability analysis in the time domain," *Physiological Measurement*, vol. 28, no. 12, pp. 1485–1494, Dec. 2007. 38

[79] "Rhythm24 Heart Rate Monitor - Black." [Online]. Available: https://scosche.eu/shop/521-heart-rate-monitor/1569-rhythm24-heart-rate-monitor---black/ 38, 51

[80] V. Inc., "Accuracy of heart rate variability metrics calculated using R-R Interval series from Valencell Benchmark™ Wrist 1.2," 2017. [Online]. Available: https://valencell.com/wp-content/uploads/2020/05/001402-01.00-HRV_PerformanceEvaluationPaper_BW1_2.pdf 38, 51

[81] "Bluetooth LE (BLE)," 2021. [Online]. Available: https://docs.particle.io/tutorials/device-os/bluetooth-le/ 38

[82] S. Theodoridis, "Chapter 2 - Probability and Stochastic Processes," in *Machine Learning (Second Edition)*, S. Theodoridis, Ed. Academic Press, Jan. 2020, pp. 19–65. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128188033000118 42

[83] M. Mahmoodian and C. Q. Li, "Chapter 11 - Stochastic failure analysis of defected oil and gas pipelines," in *Handbook of Materials Failure Analysis with Case Studies from the Oil and Gas Industry*, A. S. H. Makhlouf and M. Aliofkhazraei, Eds. Butterworth-Heinemann, Jan. 2016, pp. 235–255. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780081001172000145 42

[84] E. Calixto, "Chapter 5 - Human Reliability Analysis," in *Gas and Oil Reliability Engineering (Second Edition)*, E. Calixto, Ed. Boston: Gulf Professional Publishing, Jan. 2016, pp. 471–552. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128054277000051 42

[85] D. G. Havard, C. J. Pon, and G. S. Watt, "Clearances for Galloping on Overhead Lines," in *Probabilistic Methods Applied to Electric Power Systems*, S. G. Krishnasamy, Ed. Pergamon, Jan. 1987, pp. 503–512. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780080318745500577 42

[86] Y. Wang, S. Son, S. M. Swartz, and N. C. Goulbourne, "A mixed Von Mises distribution for modeling soft biological tissues with two distributed fiber properties," *International Journal of Solids and Structures*, vol. 49, no. 21, pp. 2914–2923, Oct. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020768312001515 43

[87] E. Grafarend and J. Awange, *Applications of Linear and Nonlinear Models: Fixed Effects, Random Effects, and Total Least Squares.* Berlin Heidelberg: Springer-Verlag, 2012. [Online]. Available: https://www.springer.com/gp/book/9783642222405 43

[88] R. Von Mises, "Über die 'Ganzzahligkeit' der Atomgewichte und verwandte Fragen," *Physikalische Zeitschrift*, vol. 19, pp. 490–500, Jan. 1918. 43

[89] A. A. Roenko, V. V. Lukin, I. Djurović, and M. Simeunović, "Estimation of parameters for generalized Gaussian distribution," in *2014 6th International Symposium on Communications, Control and Signal Processing (ISCCSP)*, May 2014, pp. 376–379. 43

[90] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. E215–220, Jun. 2000. 44, 71

[91] “Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology,” *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar. 1996. 44

[92] D. Liao, J. Cai, W. D. Rosamond, R. W. Barnes, R. G. Hutchinson, E. A. Whitsel, P. Rautaharju, and G. Heiss, “Cardiac autonomic function and incident coronary heart disease: a population-based case-cohort study. The ARIC Study. Atherosclerosis Risk in Communities Study,” *American Journal of Epidemiology*, vol. 145, no. 8, pp. 696–706, Apr. 1997. 45

[93] “Metrics and scoring: quantifying the quality of predictions — scikit-learn 0.24.1 documentation.” [Online]. Available: https://scikit-learn.org/stable/modules/model_evaluation.html#mean-squared-error 45, 54

[94] J. Pompeu Soares, M. Seoane Santos, P. Henriques Abreu, H. Araújo, and J. Santos, “Exploring the Effects of Data Distribution in Missing Data Imputation,” in *Advances in Intelligent Data Analysis XVII*, ser. Lecture Notes in Computer Science, W. Duivesteijn, A. Siebes, and A. Ukkonen, Eds. Cham: Springer International Publishing, 2018, pp. 251–263. 45

[95] R. Dias, “Nonparametric Regression: Lowess/Loess,” *GEOG 414/514*. [Online]. Available: https://www.ime.unicamp.br/~dias/loess.pdf 45

[96] K. Umetani, D. H. Singer, R. McCraty, and M. Atkinson, “Twenty-Four Hour Time Domain Heart Rate Variability and Heart Rate: Relations to Age and Gender Over Nine Decades,” *Journal of the American College of Cardiology*, vol. 31, no. 3, pp. 593–601, Mar. 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0735109797005548 46

[97] “Interpreting Residual Plots to Improve Your Regression.” [Online]. Available: https://www.qualtrics.com/support/stats-iq/analyses/regression-guides/interpreting-residual-plots-improve-regression/ 46

[98] C.-C. Hsiao, F.-W. Hsu, R.-G. Lee, and R. Lin, “Correlation analysis of heart rate variability between PPG and ECG for wearable devices in different postures,” Oct. 2017, pp. 2957–2962. 50

[99] N. Pinheiro, R. Couceiro, J. Henriques, J. Muehlsteff, I. Quintal, L. Goncalves, and P. Carvalho, “Can PPG be used for HRV analysis?” *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2016, pp. 2945–2949, Aug. 2016. 50

[100] “Aidlab. Most Advanced Activity Tracker,” 2021. [Online]. Available: https://www.aidlab.com/aidlab/ 51

[101] “Heart Rate Variability (HRV) Logger for Android and iPhone.” [Online]. Available: http://www.hrv4training.com/1/post/2021/03/heart-rate-variability-hrv-logger-for-android.html 51

[102] “Outlier removal using Hampel identifier - MATLAB hampel.” [Online]. Available: https://www.mathworks.com/help/signal/ref/hampel.html 52

[103] P. Gupta, K. K. Sharma, and S. D. Joshi, “Baseline wander removal of electrocardiogram signals using multivariate empirical mode decomposition,” *Healthcare Technology Letters*, vol. 2, no. 6, pp. 164–166, Nov. 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4678436/ 52

[104] P. van Gent, "Algorithm functioning," 2018. [Online]. Available: https://python-heart-rate-analysis-toolkit.readthedocs.io/en/latest/algorithmfunctioning.html 53

[105] D. Giavarina, "Understanding Bland Altman analysis," *Biochemia Medica*, vol. 25, no. 2, pp. 141–151, Jun. 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4470095/ 55

[106] L. Weaver, T. Wooden, and J. Grazer, "_USValidity of Apple Watch Heart Rate Sensor Compared to Polar H10 Heart Rate Monitor [Georgia College and State University]," *_USJournal of Student Research*, Apr. 2019. [Online]. Available: https://www.jsr.org/index.php/path/article/view/662 60

[107] "Polar H10 | Hartslagmeter met borstband." [Online]. Available: https://www.polar.com/nl/producten/accessoires/h10_hartslagsensor 60

[108] Fernández, *Learning from Imbalanced Data Sets*, 1st ed. New York, NY: Springer, Nov. 2018. 61

[109] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, ser. SIGMOD '00. New York, NY, USA: Association for Computing Machinery, May 2000, pp. 93–104. [Online]. Available: https://doi.org/10.1145/342009.335388 61

Appendix A

Ecosystem of micro-intelligences for affective computing

A.1 Results of the personalized analysis of the HRV data

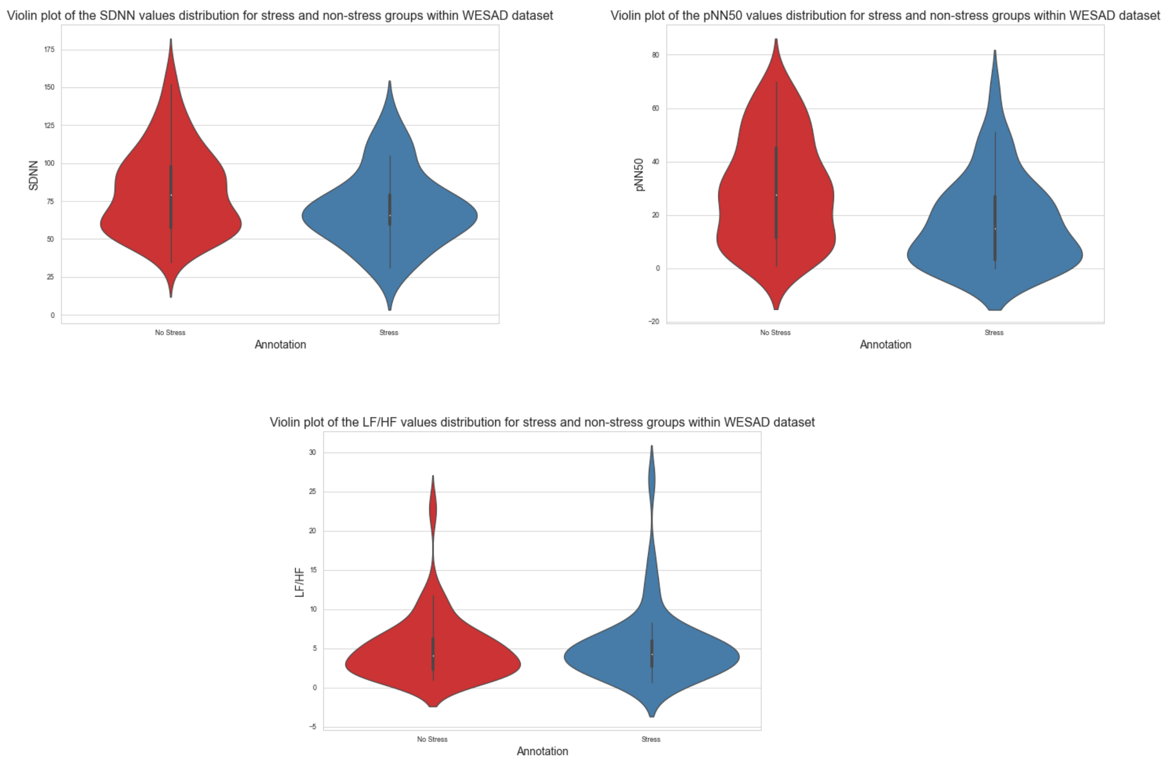

Figure A.1: "Violin" plots of the distribution of the selected HRV feature values within binary groups of participants from WESAD dataset

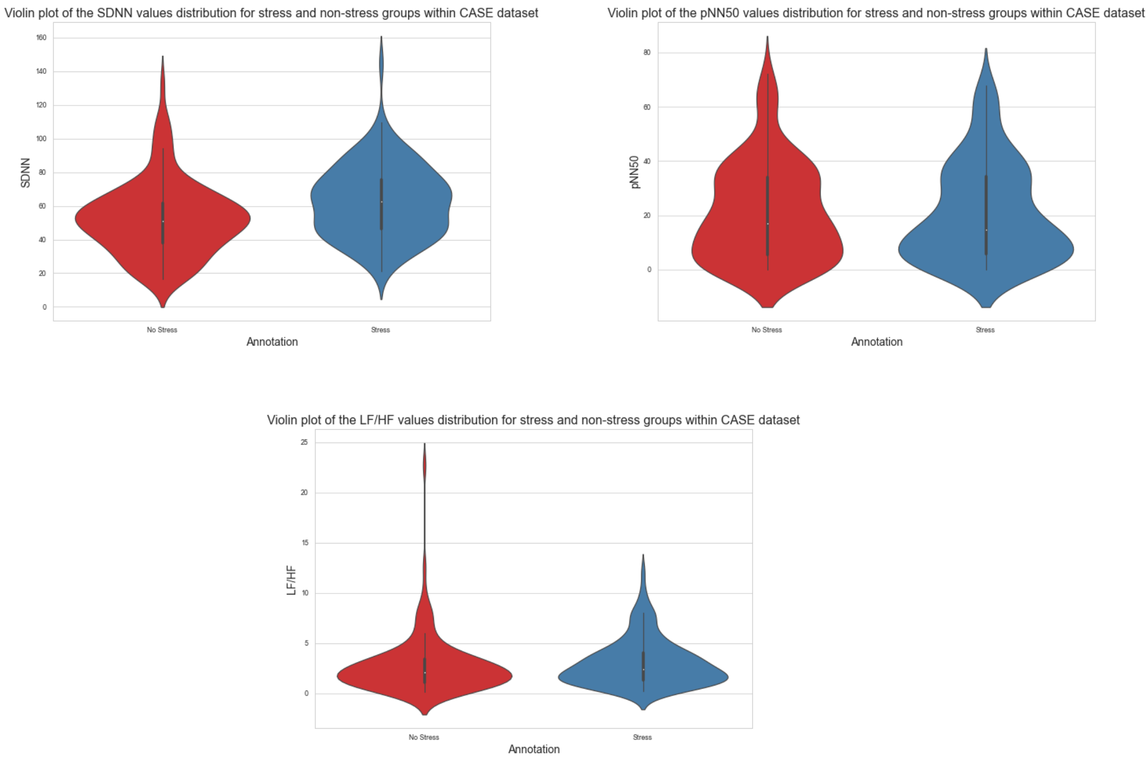

Figure A.2: “Violin” plots of the distribution of the selected HRV feature values within binary groups of participants from CASE dataset

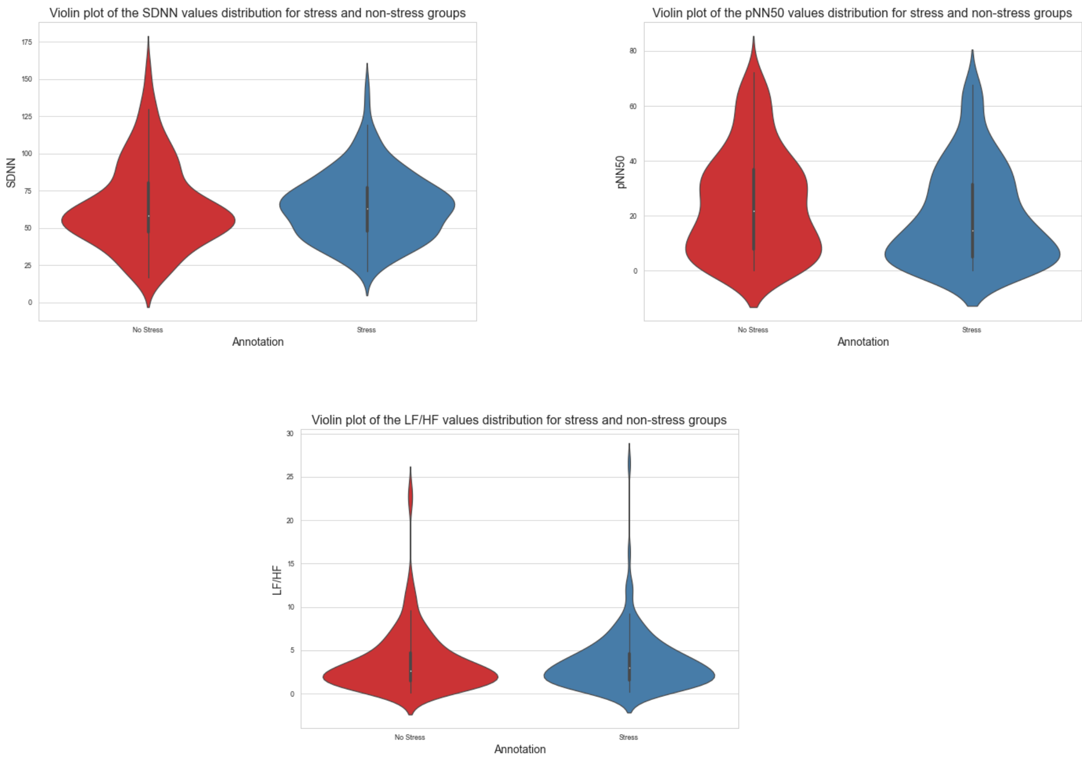

Figure A.3: "Violin" plots of the distribution of the selected HRV feature values within binary groups of participants from merged datasets

Table A.1: Percentage change in rMSSD for phase transitions across participants

Subject	**Baseline - Amusement**	**Amusement - Stress**	**Stress - Meditation**
1	-56.768 %	105.132 %	15.582 %
3	-50.600 %	54.033 %	18.416 %
4	-21.230 %	71.051 %	4.649 %
6	47.583 %	31.105 %	36.598 %
7	-20.806 %	174.157 %	9.261 %
8	-3.076 %	-66.822 %	145.456 %
9	145.456 %	59.542 %	25.684 %
10	-76.431 %	330.591 %	-18.457 %
11	-35.400 %	43.492 %	48.360 %
12	-75.704 %	75.704 %	40.298 %
13	36.569 %	-10.700 %	21.017 %
14	-86.870 %	308.311 %	103.595 %
15	-54.819 %	136.184 %	75.422 %

Table A.2: Percentage change in Mean RRI for phase transitions across participants

Subject	**Baseline - Amusement**	**Amusement - Stress**	**Stress - Meditation**
1	-42.928 %	54.377 %	2.444 %
3	-18.339 %	35.494 %	-0.498 %
4	-30.939 %	31.331 %	3.417 %
5	-8.682 %	18.749 %	-1.690 %
6	-12.610 %	58.413 %	-3.650 %
7	-20.470 %	32.944 %	3.264 %
8	-7.503 %	-9.601 %	19.203 %
9	-10.535 %	39.205 %	0.552 %
10	-41.595 %	65.135 %	-10.289 %
11	-21.715 %	20.862 %	10.819 %
12	-43.025 %	46.717 %	8.511 %
13	-3.950 %	7.712 %	6.799 %
14	-47.175 %	54.314 %	8.351 %
15	-33.221 %	63.721 %	14.810 %

Appendix B

Estimation of the missing signal impact for computing HRV

B.1 Data exploration

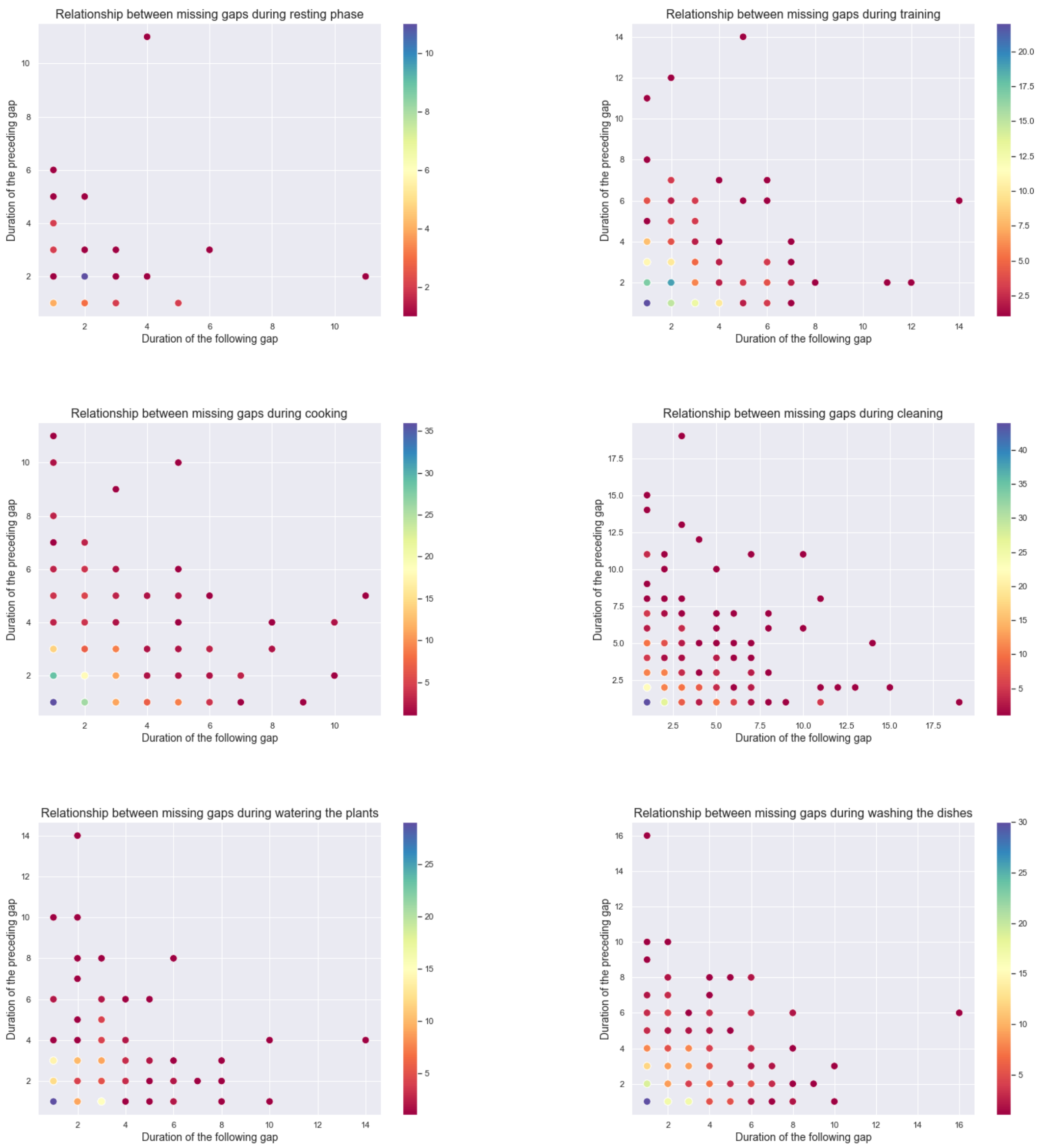

Figure B.1: Correlation plots between durations of preceding and following gaps for every activity dataset

Table B.1: Goodness of fit evaluation of the fitted probability distributions for NaN datasets

Distribution type \ Dataset type	**Resting phase**	**Training**	**Cooking**	**Cleaning**	**Washing the dishes**	**Watering the plants**
Double-Generalized Gamma	0.211	0.213	0.221	-	0.209	0.204
Right-skewed Gumbel	-	0.203	-	0.221	0.198	-
Von Mises	-	-	0.208	0.229	0.187	-
Generalized Normal	0.223	0.223	-	0.210	-	-
Laplace	0.211	-	-	-	-	-
Log-Laplace	0.211	-	-	0.217	-	-
Double Weibull	0.212	0.213	0.222	-	-	0.207
Gauss hypergeometric	-	-	0.216	-	-	-
Hyperbolic Secant	0.228	-	-	-	-	0.216
Logistic	-	0.208	0.222	0.243	0.194	0.209
Generalized Logistic	-	0.204	0.226	0.221	0.197	0.236
Student's t	-	-	-	-	0.199	0.211

Table B.2: Goodness of fit evaluation of the fitted probability distributions for not NaN datasets

Distribution type \ Dataset type	**Resting phase**	**Training**	**Cooking**	**Cleaning**	**Washing the dishes**	**Watering the plants**
Double-Generalized Gamma	-	-	0.241	0.238	-	-
Right-skewed Gumbel	-	-	-	-	0.245	-
Von Mises	-	0.231	0.237	0.224	-	-
Generalized Normal	-	-	0.199	-	0.245	-
Laplace	-	-	-	0.244	-	0.159
Log-Laplace	-	-	0.174	-	0.246	-
Double Weibull	-	-	-	-	0.214	-
Non-central Student's t	-	-	0.227	-	-	0.158
Gauss hypergeometric	-	-	-	0.241	-	-
Non-central F	-	-	0.208	-	-	0.149
Hyperbolic Secant	-	-	-	0.247	-	-
Alpha	-	0.162	-	-	-	0.155
Beta Prime	-	0.168	-	-	-	-
Logistic	-	-	-	0.235	-	-
Gilbrat	-	0.171	-	-	-	-
Erlang	-	0.231	-	-	-	-
Student's t	-	-	-	-	0.263	-
Mielke Beta-Kappa	0.116	-	-	-	-	-
Burr	0.129	0.209			0.259	0.154
Levy	0.130	-	-	-	-	-
Wald	-	-	-	-	-	0.158
inverted Weibull	0.130	-	-	-	-	-
Generalized Extreme Value	0.131	-	-	-	-	-
Power-Function	0.132	-	-	-	-	-

B.2 Missing patterns discovery

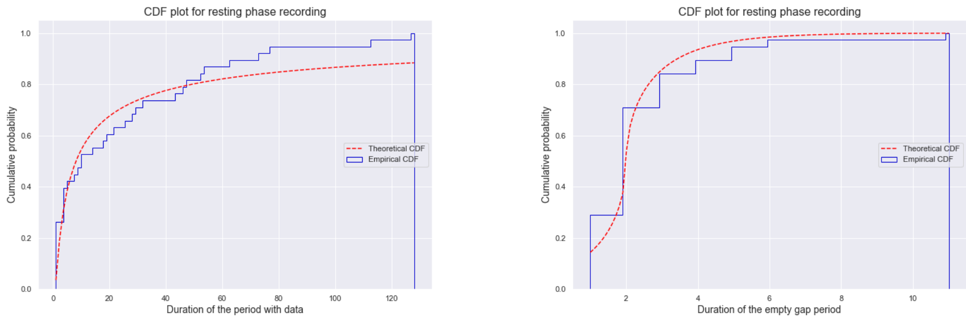

Figure B.2: CDF plots for resting phase recording datasets with empty and non-empty data periods

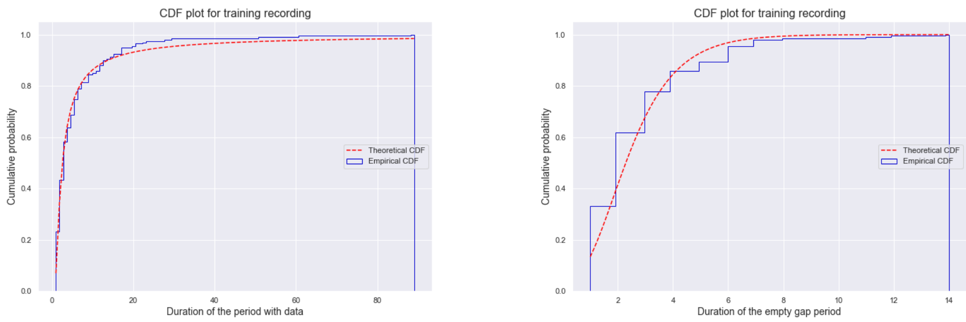

Figure B.3: CDF plots for training recording datasets with empty and non-empty data periods

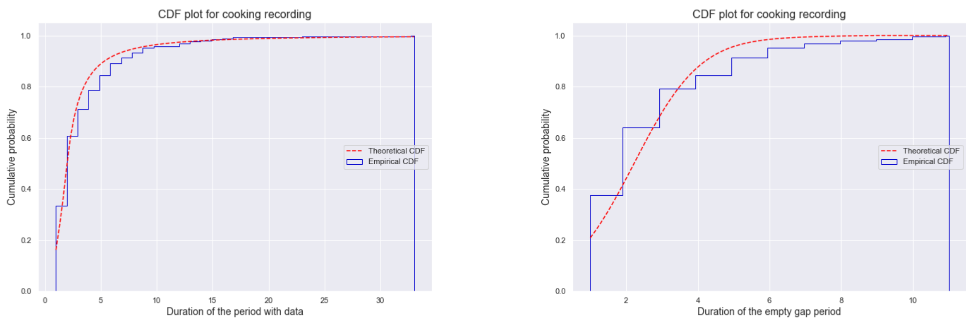

Figure B.4: CDF plots for cooking recording datasets with empty and non-empty data periods

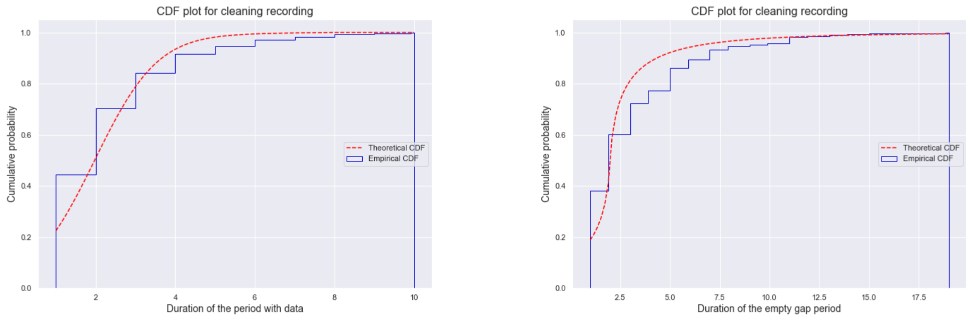

Figure B.5: CDF plots for cleaning recording datasets with empty and non-empty data periods

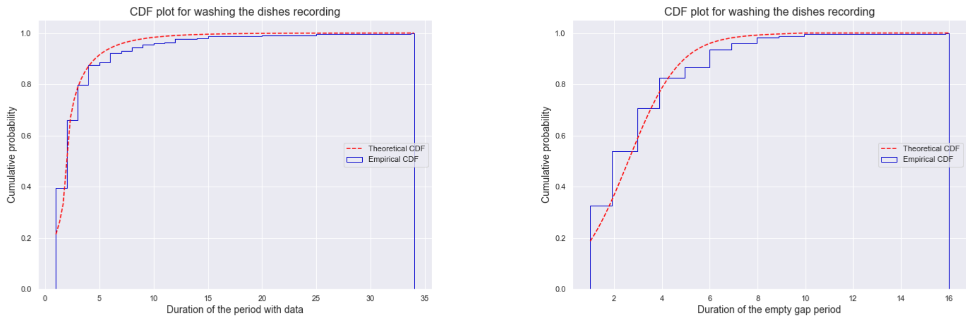

Figure B.6: CDF plots for washing the dishes recording datasets with empty and non-empty data periods

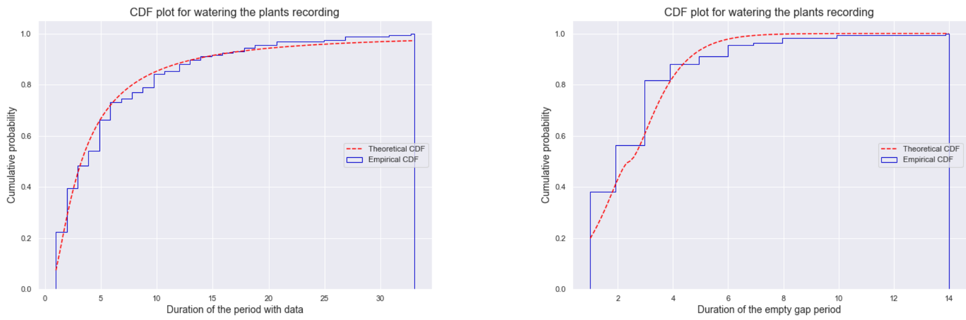

Figure B.7: CDF plots for watering the plants recording datasets with empty and non-empty data periods

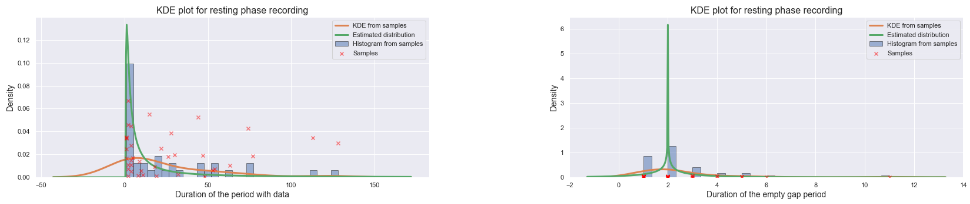

Figure B.8: KDE plots for resting phase recording data sets with empty and non-empty data periods

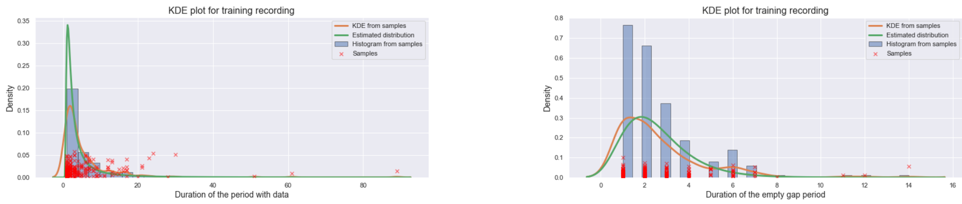

Figure B.9: KDE plots for training recording data sets with empty and non-empty data periods

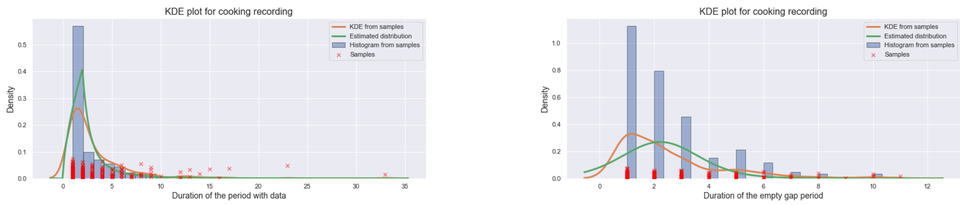

Figure B.10: KDE plots for cooking recording data sets with empty and non-empty data periods

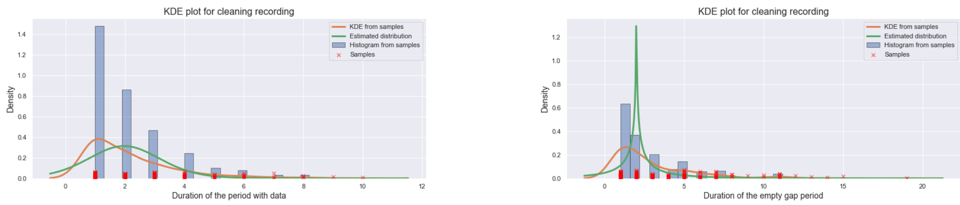

Figure B.11: KDE plots for cleaning recording data sets with empty and non-empty data periods

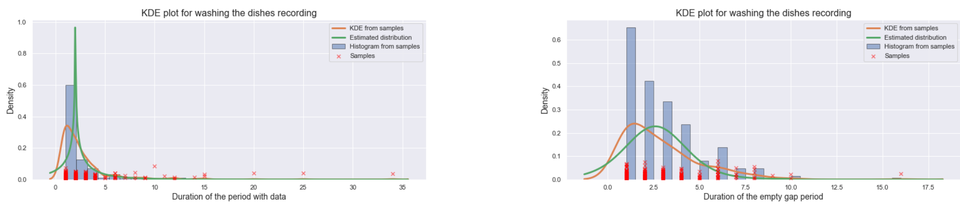

Figure B.12: KDE plots for washing the dishes recording data sets with empty and non-empty data periods

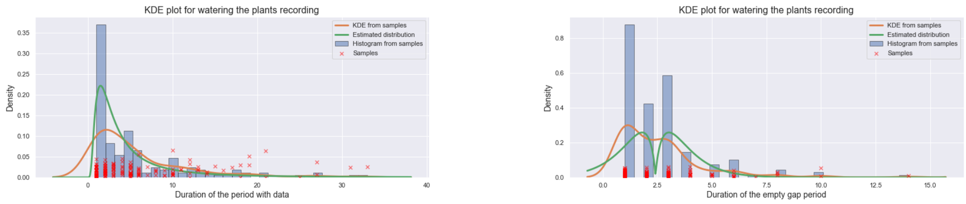

Figure B.13: KDE plots for watering the plants recording data sets with empty and non-empty data periods

B.3 Experimental results and discussion

Table B.3: Statistical analysis and descriptive statistics of rMSSD for all missing data patterns

Pattern type	Probability distribution types for empty and non-empty gaps	Missing rate	rMSSD	Bias	R^2 score	RMSE
-	-	-	28.410 ± 4.928	-	-	-
Pattern 1	"Von Mises (kappa=3.99 loc=1.0 scale=1.0) Log-Laplace (loc=1.00 c=3.5 scale=1.00)"	5 %	27.564 ± 4.670	0.846 ± 0.546	0.9572291222	1.00195744
		10 %	26.855 ± 4.573	1.555 ± 0.959	0.8591419174	1.818302261
		25 %	23.981 ± 4.437	4.429 ± 2.549	-0.1034502162	5.089223977
		50 %	19.149 ± 3.842	9.262 ± 3.106	-3.051837914	9.752169128
		75 %	13.641 ± 2.644	14.770 ± 3.751	-8.873228295	15.22315895
	"Von Mises (kappa=1.5 loc=1.5 scale=1.0) Log-Laplace (loc=1.5 c=2.0 scale=1.00)"	5 %	27.644 ± 4.723	0.766 ± 0.500	0.964707391	0.9101585952
		10 %	27.003 ± 4.650	1.407 ± 0.841	8.87E-01	1.632120344
		25 %	24.067 ± 4.535	4.344 ± 1.910	0.04585669131	4.732406008
		50 %	19.418 ± 3.721	8.993 ± 2.433	-2.689202342	9.305536788
		75 %	13.922 ± 2.954	14.488 ± 2.827	-8.271674382	14.75211462
	"Von Mises (kappa=1.5 loc=2.0 scale=1.5) Log-Laplace (loc=1.5 c=2.0 scale=1.5)"	5 %	27.572 ± 4.699	0.839 ± 0.511	0.9592767771	0.9776789474
		10 %	26.898 ± 4.597	1.512 ± 0.875	0.8710250267	1.739914534
		25 %	23.833 ± 4.388	4.577 ± 2.092	-0.07284979205	5.01816172
		50 %	18.956 ± 3.681	9.455 ± 2.646	-3.096899156	9.806247033
		75 %	13.424 ± 2.724	14.986 ± 2.979	-8.934024328	15.26995658
	"Von Mises (kappa=2.5 loc=2.0 scale=1.5) Log-Laplace (loc=2.0 c=1.5 scale=1.5)"	5 %	27.606 ± 4.705	0.804 ± 0.489	0.962589006	0.9370760029
		10 %	26.912 ± 4.588	1.499 ± 0.796	0.8782462603	1.690504563
		25 %	24.043 ± 4.560	4.368 ± 2.006	0.02165775938	4.79204177
		50 %	19.199 ± 3.731	9.212 ± 2.804	-2.938987256	9.615402763
		75 %	13.912 ± 3.064	14.499 ± 2.645	-8.244059918	14.73012962
	"Von Mises (kappa=3.0 loc=2.0 scale=2.0) Log-Laplace (loc=2.5 c=4 scale=1.0)"	5 %	27.606 ± 4.689	0.805 ± 0.500	0.9621205317	0.9429249552
		10 %	26.871 ± 4.581	1.539 ± 0.888	0.8665738531	1.769683757
		25 %	23.820 ± 4.385	4.590 ± 2.240	-0.1043277501	5.091247212
		50 %	19.008 ± 3.831	9.402 ± 2.817	-3.092994785	9.801573212
		75 %	13.428 ± 2.692	14.983 ± 3.413	-9.043422401	15.35380628
Pattern 2	"Double-Generalized Gamma (a=1.10 loc=1.00 scale=1.00) Burr (loc=1.00 c=10.50 scale=1.50 d=4.50)"	5 %	27.515 ± 4.666	0.896 ± 0.566	0.9526111066	1.054662432
		10 %	26.819 ± 4.561	1.591 ± 1.005	0.8505064396	1.87320986
		25 %	23.589 ± 4.419	4.822 ± 2.639	-0.2773967723	5.475680502
		50 %	18.674 ± 4.065	9.737 ± 3.112	-3.43788963	10.20618511
		75 %	12.948 ± 2.805	15.463 ± 3.949	-9.828629727	15.94270259
	"Double-Generalized Gamma (a=1.10 loc=1.5 scale=1.00) Burr (loc=1.0 c=8.50 scale=1.0 d=2.50)"	5 %	27.565 ± 4.691	0.845 ± 0.520	0.9584125494	0.9879986283
		10 %	26.846 ± 4.554	1.565 ± 0.968	0.8571510322	1.831107097
		25 %	23.851 ± 4.457	4.559 ± 2.606	-0.1651784569	5.229635447
		50 %	18.892 ± 4.111	9.519 ± 3.091	-3.253694793	9.992136434
		75 %	12.925 ± 2.838	15.485 ± 4.149	-9.925207183	16.01363898
	"Double-Generalized Gamma (a=2.0 loc=1.5 scale=1.5) Burr (loc=1.5 c=6.50 scale=1.0 d=3.5)"	5 %	27.537 ± 4.671	0.874 ± 0.552	0.9549157819	1.028696977
		10 %	26.882 ± 4.614	1.529 ± 0.963	0.8622406493	1.798190654
		25 %	23.894 ± 4.515	4.517 ± 2.583	-0.1440875235	5.182088401
		50 %	19.014 ± 4.123	9.396 ± 2.994	-3.130390302	9.846247284
		75 %	13.051 ± 2.877	15.359 ± 3.990	-9.706340107	15.85242511
	"Double-Generalized Gamma (a=3.0 loc=2.5 scale=1.5) Burr (loc=1.5 c=6.0 scale=1.5 d=5.5)"	5 %	27.363 ± 4.624	1.047 ± 0.676	0.93449093	1.240012061
		10 %	26.720 ± 4.550	1.690 ± 0.948	8.41E-01	1.930171336
		25 %	23.743 ± 4.400	4.667 ± 2.485	-0.1823406897	5.268009027
		50 %	18.954 ± 3.991	9.456 ± 2.926	-3.162146952	9.884026415
		75 %	12.833 ± 2.524	15.577 ± 3.911	-9.967589882	16.04467017
	"Double-Generalized Gamma (a=0.5 loc=1.0 scale=1.5) Burr (loc=2.5 c=4.5 scale=1.0 d=7.5)"	5 %	27.610 ± 4.708	0.800 ± 0.494	0.9626810447	0.9359225935
		10 %	26.930 ± 4.629	26.930 ± 4.629	0.8753824464	1.710270494
		25 %	24.052 ± 4.589	4.359 ± 2.212	-0.01077550656	4.870825233
		50 %	19.345 ± 4.007	9.065 ± 2.661	-2.792660004	9.435113924
		75 %	13.945 ± 3.067	14.465 ± 2.964	-8.276563985	14.75600402
Pattern 3	"right-skewed Gumbel (loc=1.00 scale=1.00) Alpha (a=3.5 loc=1.00 scale=1.00)"	5 %	27.568 ± 4.662	0.842 ± 0.526	0.9584068904	0.9880658469
		10 %	26.787 ± 4.540	1.623 ± 1.005	0.84613444	1.900403858
		25 %	23.777 ± 4.457	4.634 ± 2.677	-0.2099362159	5.329131424
		50 %	18.729 ± 4.014	9.682 ± 3.094	-3.387749928	10.14836611
		75 %	12.857 ± 2.719	15.553 ± 4.024	-9.972862861	16.04852667
	"right-skewed Gumbel (loc=1.5 scale=1.00) Alpha (a=3.0 loc=1.5 scale=1.50)"	5 %	27.572 ± 4.686	0.838 ± 0.531	0.9584570547	0.9874698277
		10 %	26.775 ± 4.554	1.636 ± 1.032	0.8421455379	1.92487984
		25 %	23.812 ± 4.476	4.598 ± 2.686	-0.1980057196	5.30279256
		50 %	18.863 ± 4.179	9.548 ± 3.160	-3.294924383	10.04044482
		75 %	12.803 ± 2.914	15.608 ± 4.191	-10.10173312	16.14249215

	"right-skewed Gumbel (loc=1.5 scale=2.00) Alpha (a=2.5 loc=1.0 scale=1.50)"	5 %	27.586 ± 4.688	0.825 ± 0.524	0.9597174813	0.9723743654
		10 %	26.770 ± 4.553	1.640 ± 1.030	0.8416331163	1.928001547
		25 %	23.865 ± 4.312	4.545 ± 2.337	-0.1050785531	5.09297762
		50 %	18.764 ± 3.836	9.646 ± 2.896	-3.309615284	10.05760196
		75 %	13.296 ± 2.562	15.115 ± 3.539	-9.248744577	15.50995463
	"right-skewed Gumbel (loc=1.3 scale=2.00) Alpha (a=2.0 loc=1.5 scale=2.50)"	5 %	27.423 ± 4.646	0.987 ± 0.624	0.9424236104	1.162511899
		10 %	26.845 ± 4.572	1.566 ± 0.974	0.8564881113	1.835350999
		25 %	23.925 ± 4.462	4.485 ± 2.385	-0.09134303491	5.061227197
		50 %	19.022 ± 3.825	9.388 ± 2.791	-3.075779277	9.780938337
		75 %	13.436 ± 2.584	14.974 ± 3.359	-9.01782275	15.33422616
	"right-skewed Gumbel (loc=1.5 scale=3.00) Alpha (a=1.5 loc=2.0 scale=1.5)"	5 %	27.455 ± 4.621	0.956 ± 0.644	0.9440387965	1.146089979
		10 %	26.872 ± 4.556	1.538 ± 0.966	0.860749442	1.807896919
		25 %	23.766 ± 4.286	4.645 ± 2.204	-0.1192453782	5.125518995
		50 %	18.821 ± 3.621	9.589 ± 2.846	-3.251224165	9.989234199
		75 %	12.684 ± 2.210	15.726 ± 3.716	-10.10526237	16.1450578
Pattern 4	"Generalized normal (loc=1.00 beta=1.30 scale=1.00) Von Mises (kappa=3.99 loc=1.0 scale=1.0)"	5 %	27.579 ± 4.696	0.832 ± 0.497	0.9603516985	0.9646893549
		10 %	26.861 ± 4.577	1.550 ± 0.885	0.8653889509	1.777524302
		25 %	23.864 ± 4.463	4.546 ± 2.398	-0.1173317974	5.121135555
		50 %	19.099 ± 3.893	9.311 ± 2.954	-3.052967439	9.753528333
		75 %	13.364 ± 2.859	15.047 ± 3.260	-9.083466082	15.38438407
	"Generalized normal (loc=1.00 beta=1.50 scale=1.50) Von Mises (kappa=1.5 loc=1.5 scale=1.0)"	5 %	27.607 ± 4.673	0.803 ± 0.522	0.9613026087	0.9530507712
		10 %	26.925 ± 4.564	1.486 ± 0.869	0.8748214383	1.714115851
		25 %	24.193 ± 4.440	4.218 ± 1.473	0.1528561884	4.459167272
		50 %	19.607 ± 3.789	8.803 ± 2.116	-2.485945556	9.045560704
		75 %	14.361 ± 3.422	14.050 ± 2.943	-7.766331266	14.34445761
	"Generalized normal (loc=1.50 beta=1.50 scale=1.50) Von Mises (kappa=1.50 loc=2.0 scale=1.5)"	5 %	27.591 ± 4.696	0.820 ± 0.503	0.9609320374	0.9576031682
		10 %	26.889 ± 4.593	1.521 ± 0.874	0.8699792683	1.746954104
		25 %	23.941 ± 4.427	4.469 ± 1.882	0.003117462059	4.837235003
		50 %	19.505 ± 3.502	8.906 ± 2.469	-2.630198311	9.230821814
		75 %	14.422 ± 2.996	13.988 ± 2.591	-7.612731648	14.21823377
	"Generalized normal (loc=1.0 beta=2.0 scale=2.0) Von Mises (kappa=2.5 loc=2.0 scale=1.5)"	5 %	27.620 ± 4.724	0.790 ± 0.464	0.9645304104	0.9124378123
		10 %	26.900 ± 4.626	1.511 ± 0.774	0.878090587	1.69158495
		25 %	24.142 ± 4.647	4.268 ± 2.141	0.03502840448	4.759183579
		50 %	19.744 ± 3.989	8.667 ± 2.481	-2.453407977	9.003246408
		75 %	14.502 ± 3.512	13.908 ± 2.476	-7.493489372	14.11946582
	"Generalized normal (loc=2.5 beta=4 scale=2.0) Von Mises (kappa=3.0 loc=2.0 scale=2.0)"	5 %	27.556 ± 4.709	0.854 ± 0.492	0.9589449349	0.9816542818
		10 %	26.803 ± 4.586	1.607 ± 0.906	0.856089825	1.837896045
		25 %	24.084 ± 4.264	4.327 ± 1.588	0.09859083271	4.59977065
		50 %	19.292 ± 3.618	9.119 ± 2.524	-2.805001974	9.4504532
		75 %	13.681 ± 2.830	14.729 ± 3.362	-8.708158919	15.09536543

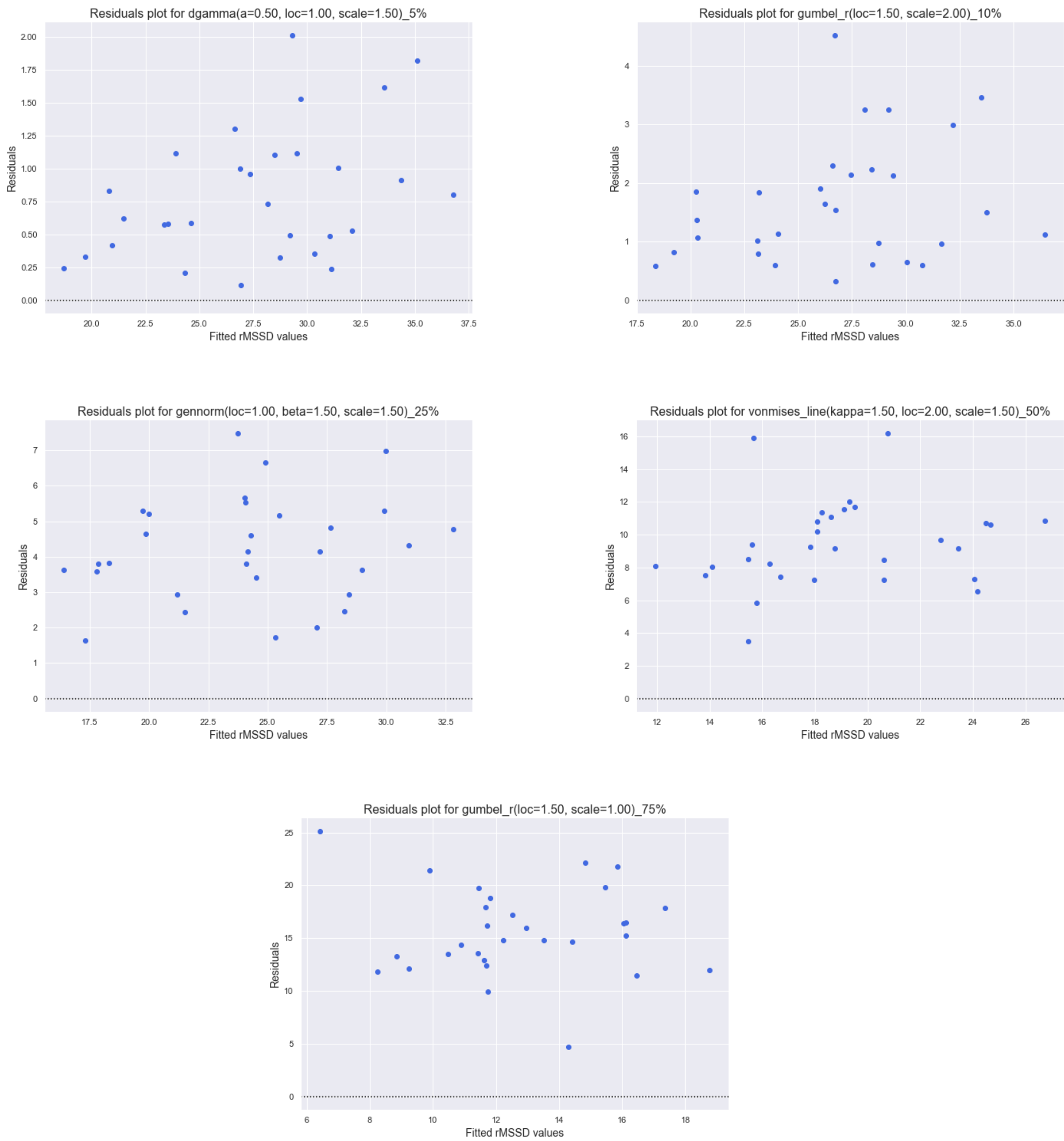

Figure B.14: Residual plots for the fitted rMSSD values computed at different missing rates

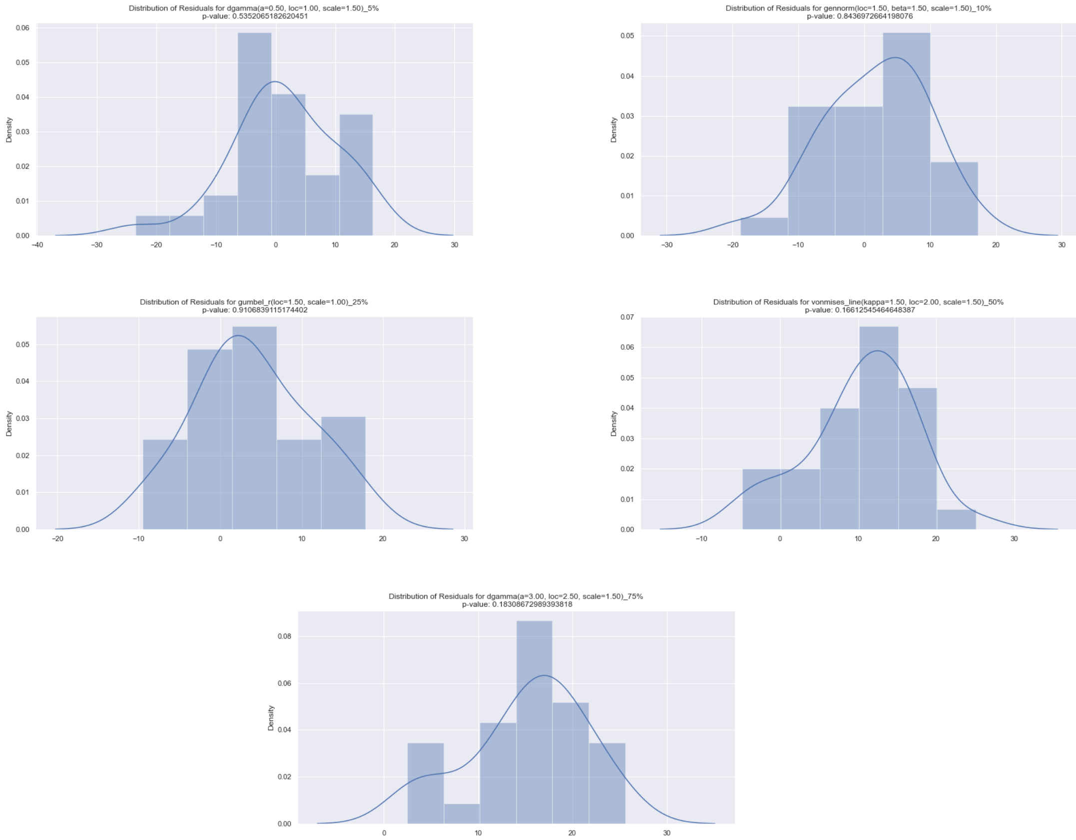

Figure B.15: Normal distributions of the residuals

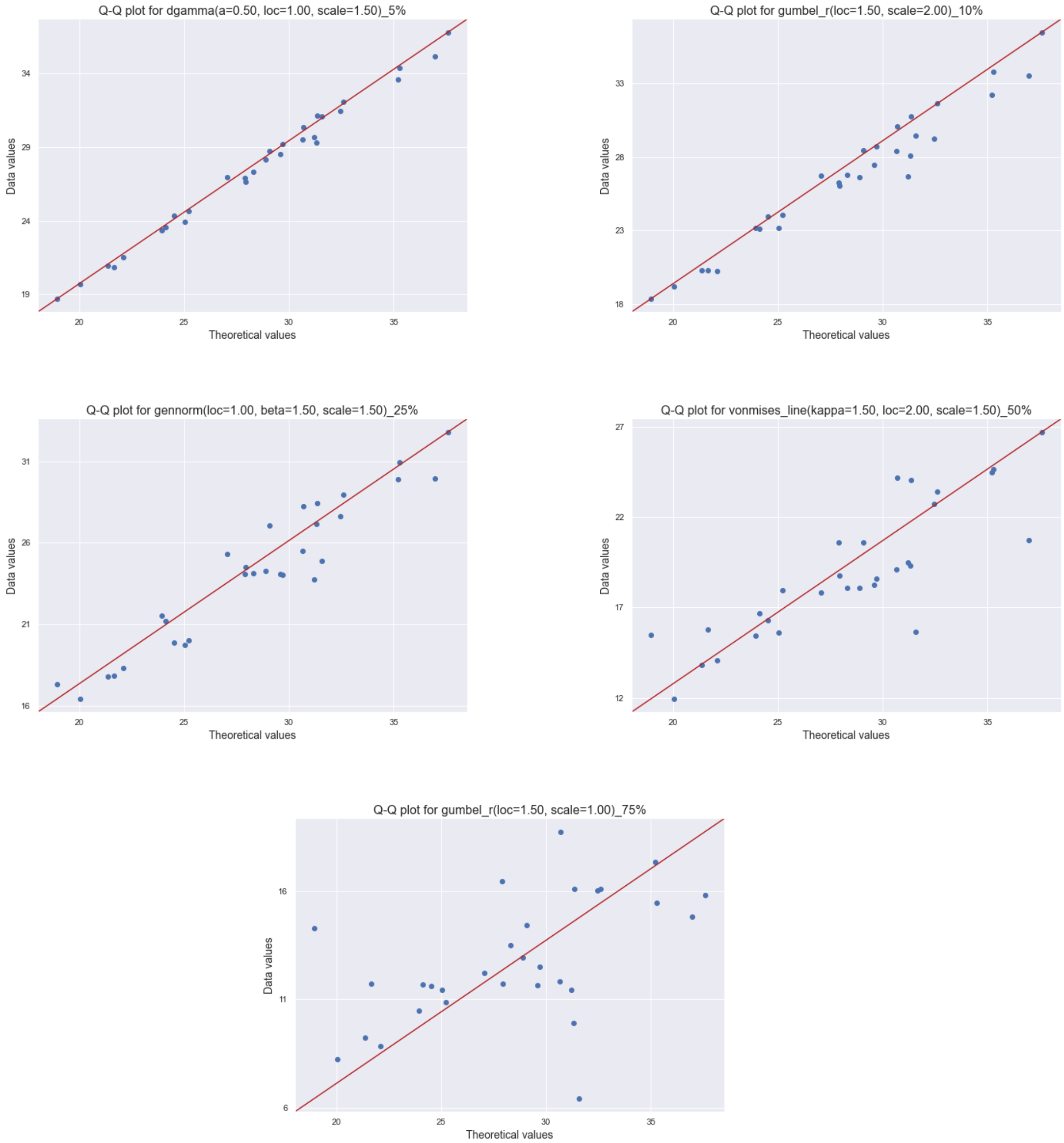

Figure B.16: Q-Q plots for golden-standard rMSSD values (i.e. theoretical values) and fitted rMSSD values (i.e. data values) computed at different missing rates

B.4 Assessment of experiment repeatability

Table B.4: Goodness of fit evaluation of the fitted probability distributions for NaN datasets from the repeated experiment

Distribution type \ Dataset type	**Training**	**Cooking**	**Cleaning**	**Washing the dishes**	**Watering the plants**
Double-Generalized Gamma	0.214	0.228	0.322	-	0.319
Right-skewed Gumbel	0.233	-	-	0.222	-
Von Mises	0.241	0.219	0.288	0.227	0.281
Generalized Normal	0.223	-	-	-	-
Log-Laplace	0.244	-	-	-	-
Logistic	-	0.225	0.284	-	0.268
Generalized Logistic	0.232	0.237	-	0.222	-
Hyperbolic Secant	-	0.235	-	-	-
Student's t	-	0.237	-	-	-
R-distributed	-	-	0.289	0.231	0.269
Normal	-	-	0.289	0.231	0.269
Kolmogorov-Smirnov two-sided test statistic	-	-	-	0.233	-
Generalized Exponential	-	-	0.289	-	-
Log-Gamma	-	-	-	-	0.276

Table B.5: Goodness of fit evaluation of the fitted probability distributions for NOT NaN datasets from the repeated experiment

Distribution type \ Dataset type	**Training**	**Cooking**	**Cleaning**	**Washing the dishes**	**Watering the plants**
Double-Generalized Gamma	0.226	-	0.222	0.298	-
Right-skewed Gumbel	0.251	0.258	0.235	0.293	-
Beta-Prime	-	-	-	-	0.171
Generalized Normal	0.236	0.221	-	-	-
Log-Laplace	0.184	-	-	-	-
Wald	-	-	-	-	0.165
non-central Student's t	0.238	-	-	-	0.163
Generalized Logistic	0.249	0.257	0.234	0.292	-
Double Weibull	-	0.215	0.234	0.237	-
Gilbrat	-	0.258	0.201	0.334	-
Cauchy	-	0.259	-	0.339	-
Alpha	-	-	0.189	-	0.163
non-central F	-	-	-	-	0.161
Inverted Gamma	-	-	-	-	0.171

Appendix C

Reliability assessment of the PPG-based HRV parameters

C.1 Initial experiment results and discussion

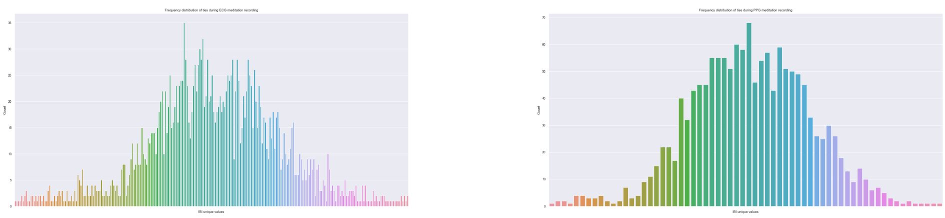

Figure C.1: Distribution of unique inter-beat interval values across ECG and PPG datasets during meditation recording

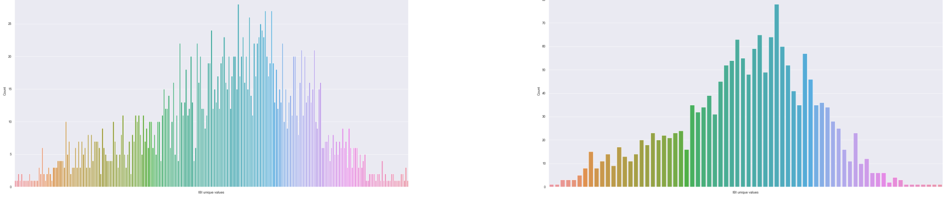

Figure C.2: Distribution of unique inter-beat interval values across ECG and PPG datasets during reading recording

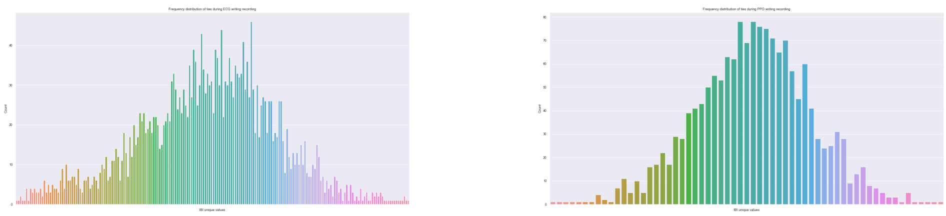

Figure C.3: Distribution of unique inter-beat interval values across ECG and PPG datasets during writing recording

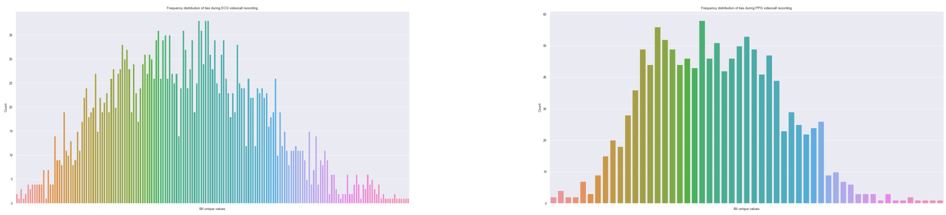

Figure C.4: Distribution of unique inter-beat interval values across ECG and PPG datasets during videocall recording

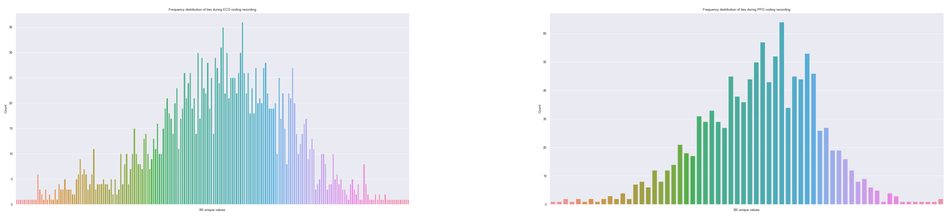

Figure C.5: Distribution of unique inter-beat interval values across ECG and PPG datasets during coding recording

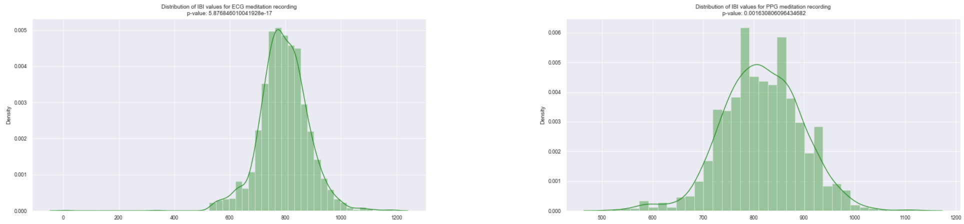

Figure C.6: Anderson-Darling normality test for ECG and PPG datasets during meditation recording

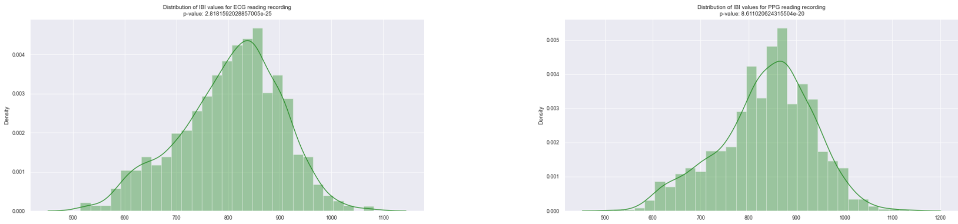

Figure C.7: Anderson-Darling normality test for ECG and PPG datasets during reading recording

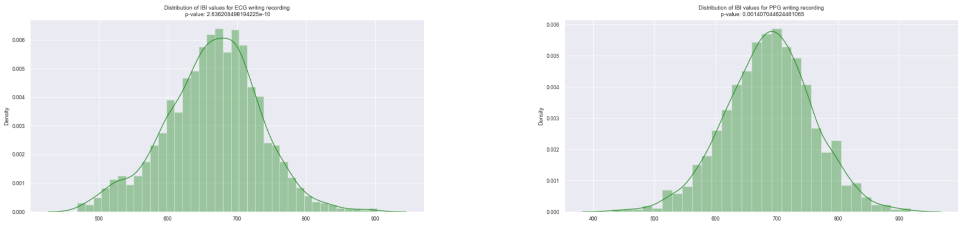

Figure C.8: Anderson-Darling normality test for ECG and PPG datasets during writing recording

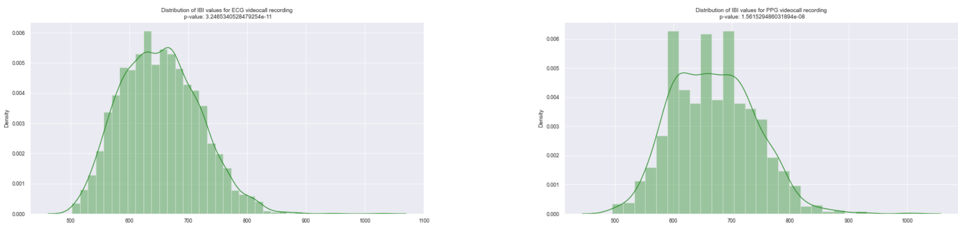

Figure C.9: Anderson-Darling normality test for ECG and PPG datasets during videocall recording

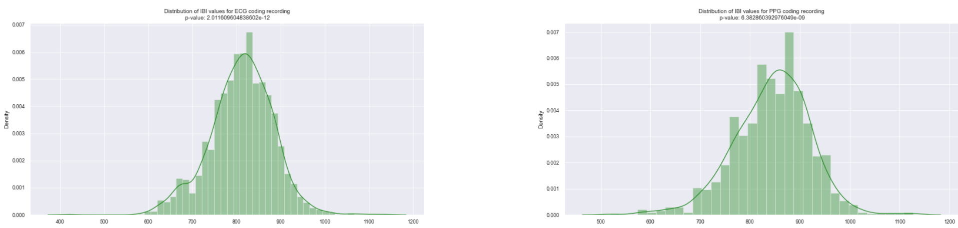

Figure C.10: Anderson-Darling normality test for ECG and PPG datasets during coding recording

Table C.1: Error evaluation between ECG and PPG-derived rMSSD values during initial experiment

Dataset	**Sensor type**	**rMSSD**	**RMSE**	R^2 **score**	**Bias**	**MAE**	**MRE**	ρ
Meditation	ECG	47.740 ± 6.125	17.811	-9.571	-17.400 ± 4.253	17.4	37.6385	0.9
	PPG	65.140 ± 3.066						
Reading	ECG	46.040 ± 3.834	14.838	-17.724	-14.400 ± 4.001	14.4	31.608	0.5
	PPG	60.440 ± 4.717						
Writing	ECG	28.880 ± 8.019	23.544	-9.775	-23.000 ± 5.624	23	89.114	0.4
	PPG	51.880 ± 4.543						
Videocall	ECG	32.180 ± 10.52	26.706	-7.056	-26.440 ± 4.205	26.44	94.749	0.9
	PPG	58.620 ± 9.655						
Coding	ECG	60.460 ± 3.005	18.661	-47.213	-17.720 ± 6.540	17.72	29.628	0.099
	PPG	78.180 ± 5.143						

Table C.2: Error evaluation between ECG and PPG-derived SDNN values during initial experiment

Dataset	**Sensor type**	**SDNN**	**RMSE**	R^2 **score**	**Bias**	**MAE**	**MRE**	ρ
Meditation	ECG	55.500 ± 3.613	7.909	-4.990	-6.960 ± 4.201	6.96	12.937	-0.359
	PPG	62.460 ± 1.843						
Reading	ECG	57.400 ± 8.473	7.588	-0.003	-6.700 ± 3.982	6.7	12.484	0.8
	PPG	64.100 ± 6.463						
Writing	ECG	38.760 ± 5.638	11.455	-4.159	-10.64 ± 6.847	10.64	27.041	-0.3
	PPG	48.440 ± 2.726						
Videocall	ECG	37.180 ± 11.00	12.709	-0.669	-11.760 ± 5.391	11.76	37.178	0.999
	PPG	48.940 ± 7.346						
Coding	ECG	55.060 ± 4.608	11.155	-6.326	-10.140 ± 5.199	10.14	19.0043	0.3
	PPG	65.200 ± 4.020						

Table C.3: Error evaluation between ECG and PPG-derived Mean RRI values during initial experiment

Dataset	**Sensor type**	**Mean RR**	**RMSE**	R^2 **score**	**Bias**	**MAE**	**MRE**	ρ
Meditation	ECG	789.800 ± 23.721	28.348	-0.785	-26.800 ± 10.330	26.8	3.413	0.9
	PPG	816.600 ± 20.206						
Reading	ECG	796.000 ± 27.092	46.757	-2.723	-43.800 ± 18.295	43.8	0.055	0.9
	PPG	839.800 ± 29.380						
Writing	ECG	658.600 ± 48.003	37.995	0.217	-33.200 ± 20.657	33.2	5.205	0.7
	PPG	691.800 ± 35.570						
Videocall	ECG	645.800 ± 47.457	30.381	0.488	-29.400 ± 8.562	29.4	4.644	0.999
	PPG	675.200 ± 40.283						
Coding	ECG	806.800 ± 10.941	37.564	-13.735	-36.200 ± 11.212	36.2	4.495	0.499
	PPG	843.000 ± 11.225						

Table C.4: Error evaluation between ECG and PPG-derived pNN50 values during initial experiment

Dataset	**Sensor type**	**pNN50**	**RMSE**	R^2 **score**	**Bias**	**MAE**	**MRE**	ρ
Meditation	ECG	25.910 ± 5.049	17.022	-13.208	-16.882 ± 2.429	16.882	69.087	0.9
	PPG	42.792 ± 2.804						
Reading	ECG	22.748 ± 2.262	15.116	-54.844	-14.820 ± 3.327	14.82	66.048	-0.2
	PPG	37.568 ± 3.266						
Writing	ECG	10.814 ± 4.063	18.931	-26.137	-18.270 ± 5.542	18.27	199.372	0.1
	PPG	29.084 ± 5.310						
Videocall	ECG	14.612 ± 2.305	20.395	-96.863	-19.492 ± 6.708	19.492	130.149	0.7
	PPG	34.104 ± 8.826						
Coding	ECG	41.198 ± 3.630	15.867	-22.886	-14.948 ± 5.950	14.948	37.428	-0.899
	PPG	56.146 ± 2.929						

C.2 Repeated experiment results evaluation and discussion

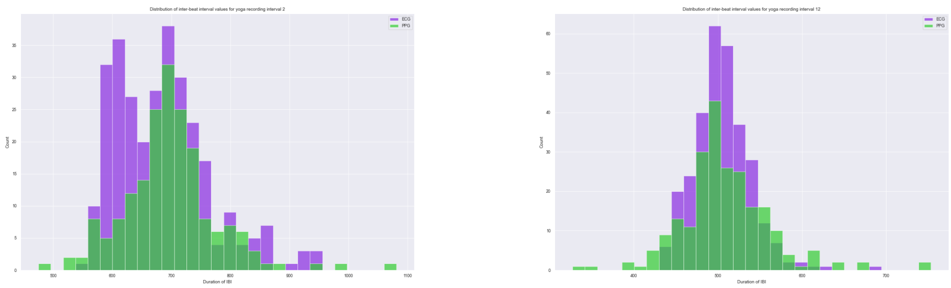

Figure C.11: Frequency distribution of the 5-min sample for ECG and PPG recordings during yoga within repeated experiment

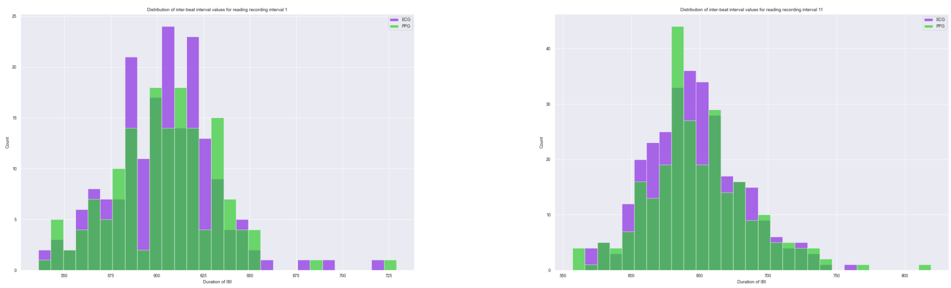

Figure C.12: Frequency distribution of the 5-min sample for ECG and PPG recordings during reading within repeated experiment

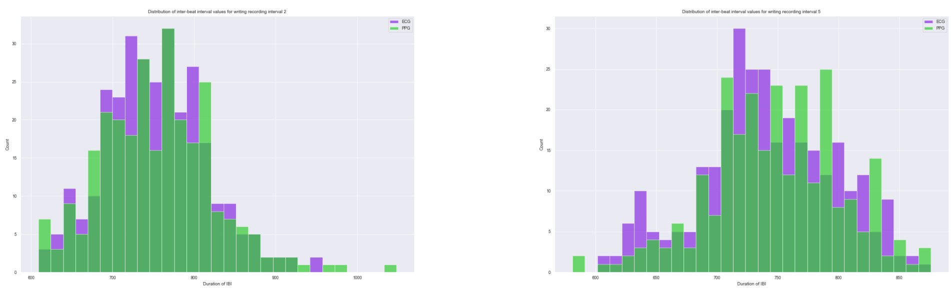

Figure C.13: Frequency distribution of the 5-min sample for ECG and PPG recordings during writing recording within repeated experiment

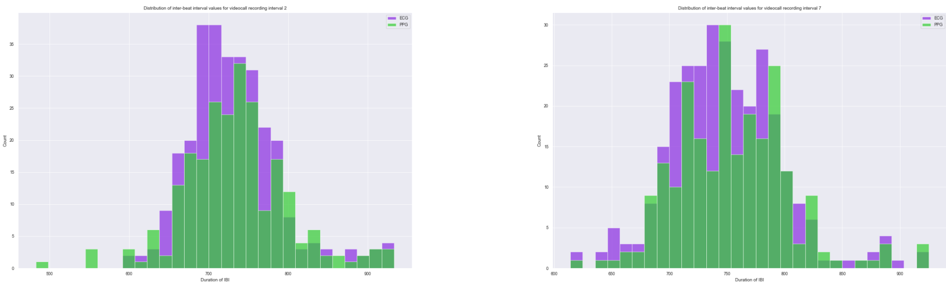

Figure C.14: Frequency distribution of the 5-min sample for ECG and PPG recordings during videocall recording within repeated experiment

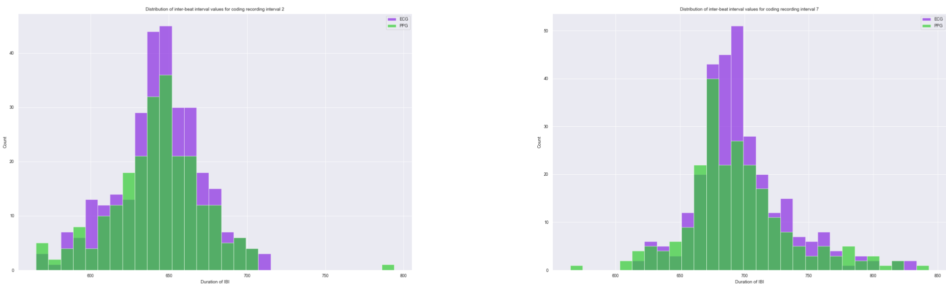

Figure C.15: Frequency distribution of the 5-min sample for ECG and PPG recordings during coding recording within repeated experiment

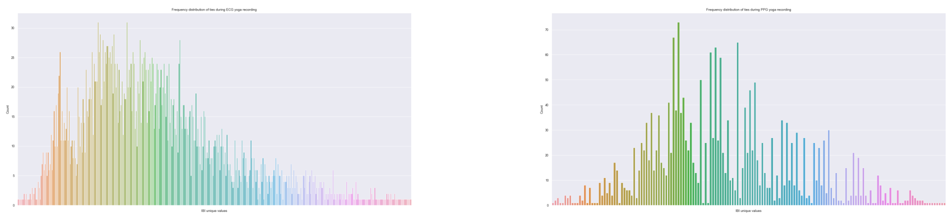

Figure C.16: Distribution of unique inter-beat interval values across ECG and PPG datasets during yoga recording within repeated experiment

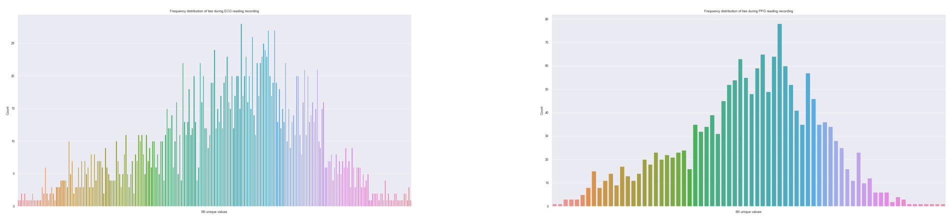

Figure C.17: Distribution of unique inter-beat interval values across ECG and PPG datasets during reading recording within repeated experiment

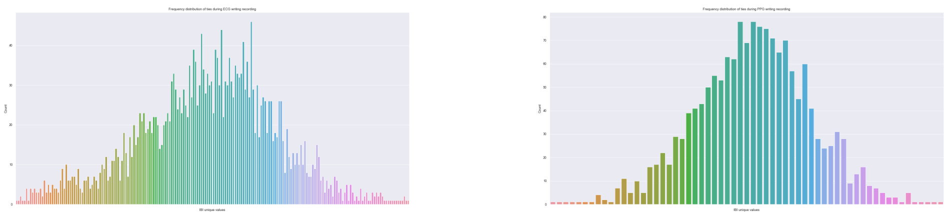

Figure C.18: Distribution of unique inter-beat interval values across ECG and PPG datasets during writing recording within repeated experiment

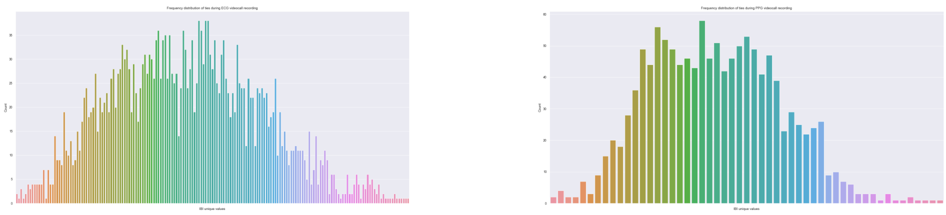

Figure C.19: Distribution of unique inter-beat interval values across ECG and PPG datasets during videocall recording within repeated experiment

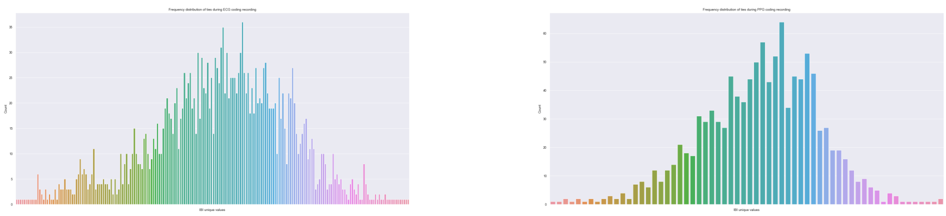

Figure C.20: Distribution of unique inter-beat interval values across ECG and PPG datasets during coding recording within repeated experiment

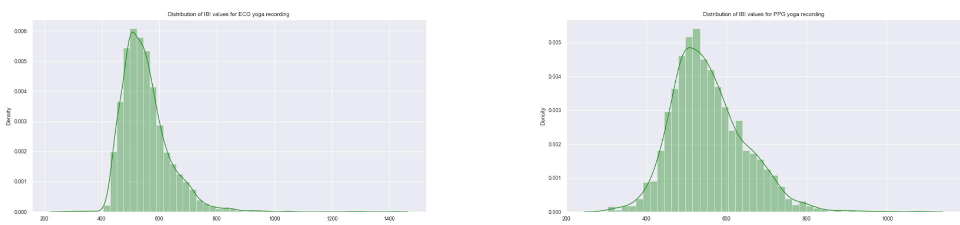

Figure C.21: Anderson-Darling normality test for ECG and PPG datasets during yoga recording within repeated experiment

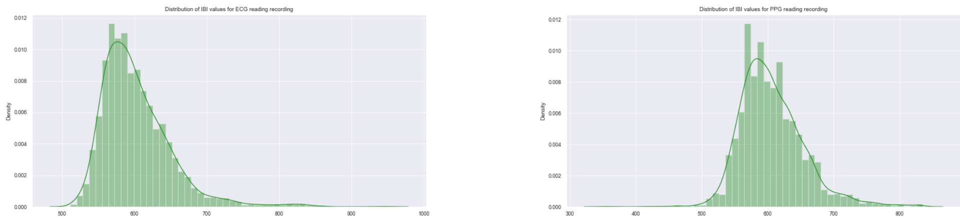

Figure C.22: Anderson-Darling normality test for ECG and PPG datasets during reading recording within repeated experiment

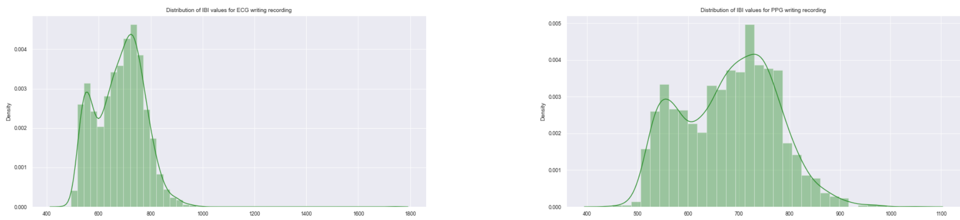

Figure C.23: Anderson-Darling normality test for ECG and PPG datasets during writing recording within repeated experiment

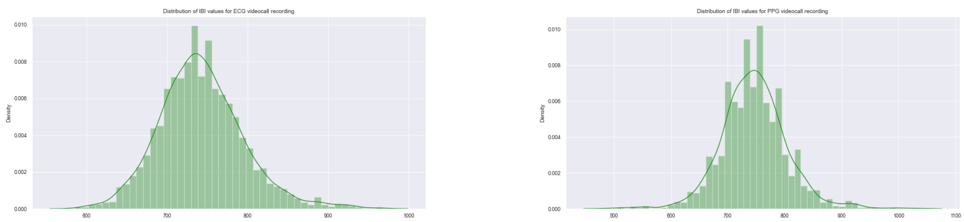

Figure C.24: Anderson-Darling normality test for ECG and PPG datasets during videocall recording within repeated experiment

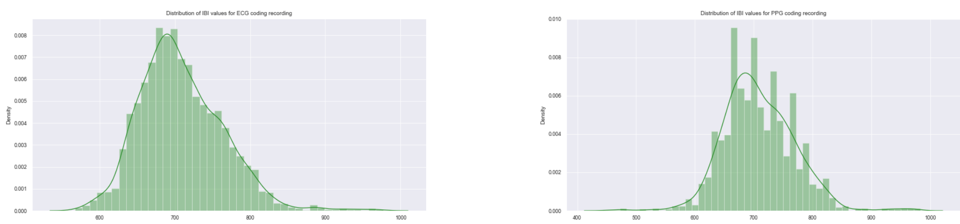

Figure C.25: Anderson-Darling normality test for ECG and PPG datasets during coding recording within repeated experiment

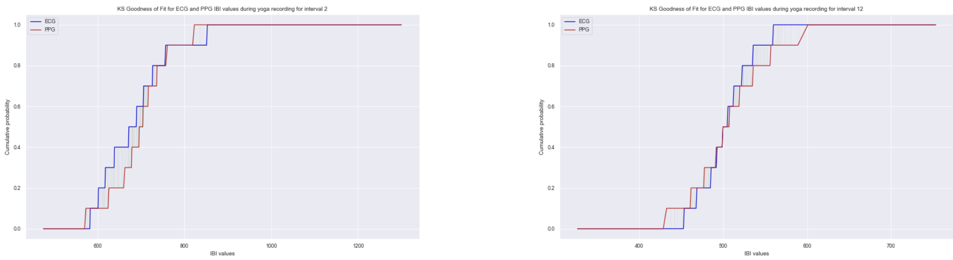

Figure C.26: Cumulative probability plot of the 5-min sample for ECG and PPG datasets during yoga recording within repeated experiment

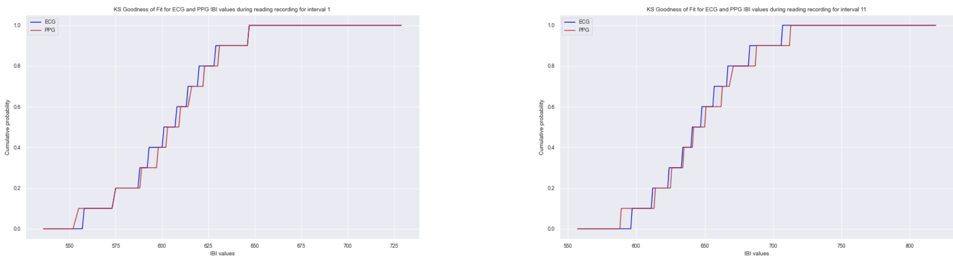

Figure C.27: Cumulative probability plot of the 5-min sample for ECG and PPG datasets during reading recording within repeated experiment

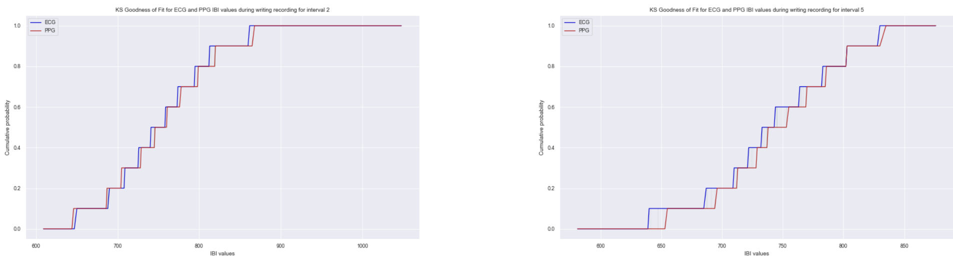

Figure C.28: Cumulative probability plot of the 5-min sample for ECG and PPG datasets during writing recording within repeated experiment

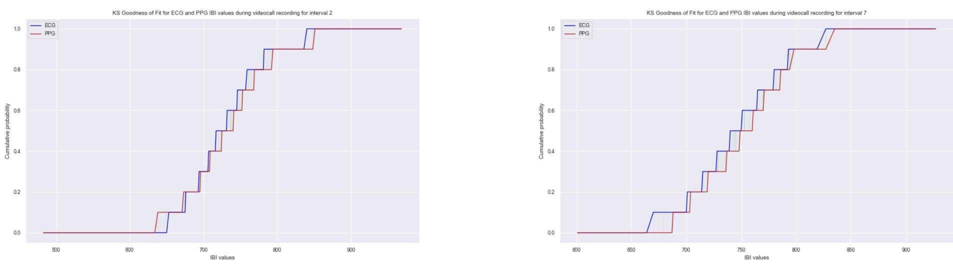

Figure C.29: Cumulative probability plot of the 5-min sample for ECG and PPG datasets during videocall recording within repeated experiment

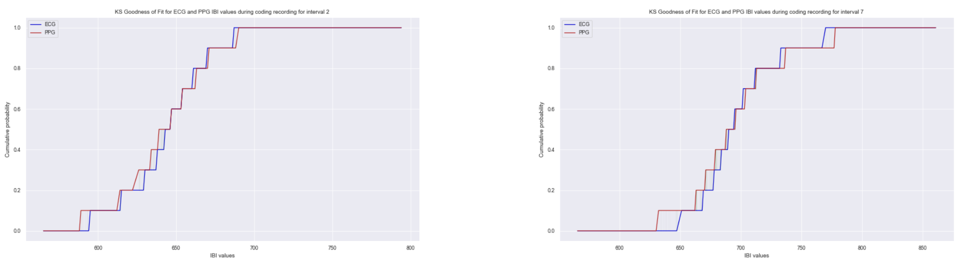

Figure C.30: Cumulative probability plot of the 5-min sample for ECG and PPG datasets during coding recording within repeated experiment

Table C.5: Continuous distribution types of the ECG data 5-min samples for all activity recordings

Dataset / Sample No.	**Yoga**	**Reading**	**Writing**	**Videocall**	**Coding**
1	Generalized Extreme	Logistic	Double Weibull	Johnson SU	Hyperbolic Secant
2	Beta Prime	Johnson SU	Beta Prime	Exponentially Modified Normal	Generalized Normal
3	Johnson SU	Exponentially Modified Normal	Double Gamma	Generalized Normal	Beta
4	Inverted Gamma	Birnbaum-Saunders	Double Weibull	Beta Prime	Double Weibull
5	Chi-squared	Johnson SU	Log-Logistic	Generalized Normal	Generalized Normal
6	Logistic	Johnson SU	Burr (Type XII)	Burr (Type XII)	Double Gamma
7	Generalized Extreme	Johnson SU	Johnson SU	Log-Logistic	Johnson SU
8	Chi	Birnbaum-Saunders	Chi	Beta	Johnson SU
9	Exponentially Modified Normal	Logistic	Exponentiated Weibull	Alpha	Exponentially Modified Normal
10	Johnson SU	Log-Logistic	Double Weibull	Exponentially Modified Normal	Hyperbolic Secant
11	Von Mises	Mielke	Inverted Gamma	Burr (Type XII)	Generalized Logistic
12	Generalized Normal	Exponentially Modified Normal	Tukey-Lamdba	Generalized Normal	Beta Prime
13	Double Weibull	Johnson SU	Generalized Extreme	Maxwell	Generalized Normal

Table C.6: Continuous distribution types of the PPG data 5-min samples for all activity recordings

Dataset / Sample No.	**Yoga**	**Reading**	**Writing**	**Videocall**	**Coding**
1	Hyperbolic Secant	Logarithmic Gamma	Double Weibull	Johnson SU	Johnson SU
2	Johnson SU	Johnson SU	Logistic	Double Gamma	Generalized Normal
3	Generalized Normal	Beta	Logistic	Johnson SU	Hyperbolic Secant
4	Burr (Type XII)	Generalized Logistic	Hyperbolic Secant	Beta	Log-Logistic
5	Log-Logistic	Logistic	Exponentially Modified Normal	Beta	Hyperbolic Secant
6	Generalized Normal	Tukey-Lamdba	Burr (Type XII)	Alpha	Log-Logistic
7	Burr (Type XII)	Exponentially Modified Normal	Gauss Hypergeometric	Logistic	Johnson SU
8	Burr (Type XII)	Log-Logistic	Generalized Normal	Birnbaum-Saunders	Johnson SU
9	Tukey-Lamdba	Double Weibull	Burr (Type XII)	Burr (Type XII)	Generalized Normal
10	scaled Kolmogorov-Smirnov	Tukey-Lamdba	Double Gamma	Log-Logistic	Double Weibull
11	Generalized Normal	Burr (Type XII)	Power Normal	Double Weibull	Johnson SU
12	Johnson SU	Johnson SU	Student's t	Generalized Normal	Double Weibull
13	Power Normal	Johnson SU	Beta	cauchy	Generalized Normal

Table C.7: Error evaluation between ECG and PPG-derived rMSSD values during repeated experiment

Dataset	**Sensor type**	**rMSSD**	**RMSE**	R^2 **score**	**Bias**	**MAE**	**MRE**	ρ
Yoga	ECG	36.517 ± 11.652	13.082	-0.513	-10.43 ± 12.519	10.433	30.646	0.2
	PPG	42.883 ± 5.614						
Reading	ECG	18.467 ± 4.771	13.432	-8.511	-11.700 ± 7.227	11.7	73.041	0.086
	PPG	30.167 ± 6.404						
Writing	ECG	23.933 ± 4.498	13.145	-9.246	-11.483 ± 7.007	11.483	46.644	0.696
	PPG	35.417 ± 10.062						
Videocall	ECG	21.917 ± 2.192	16.539	-67.311	-16.250 ± 3.377	16.25	75.806	-0.086
	PPG	38.167 ± 2.116						
Coding	ECG	17.750 ± 3.098	23.125	-65.871	-22.550 ± 5.606	22.55	134.227	-0.543
	PPG	40.300 ± 2.925						

Table C.8: Error evaluation between ECG and PPG-derived SDNN values during repeated experiment

Dataset	**Sensor type**	**SDNN**	**RMSE**	R^2 **score**	**Bias**	**MAE**	**MRE**	ρ
Yoga	ECG	38.933 ± 10.984	8.442	0.291	-0.100 ± 9.247	0.1	5.003	0.2
	PPG	39.033 ± 5.109						
Reading	ECG	29.300 ± 6.075	6.889	-0.543	-5.399 ± 7.253	5.4	0.189	0.143
	PPG	31.200 ± 7.979						
Writing	ECG	43.000 ± 9.098	3.783	0.793	3.33 ± 4.080	3.333	-3.131	0.928
	PPG	42.333 ± 12.045						
Videocall	ECG	39.850 ± 6.093	6.039	-0.178	-5.016 ± 6.109	5.017	7.739	0.2
	PPG	42.167 ± 2.767						
Coding	ECG	32.633 ± 5.427	5.645	-0.298	-4.683 ± 3.506	4.683	15.228	0.772
	PPG	37.283 ± 4.681						

Table C.9: Error evaluation between ECG and PPG-derived Mean RRI values during repeated experiment

Dataset	**Sensor type**	**Mean RR**	**RMSE**	R^2 **score**	**Bias**	**MAE**	**MRE**	ρ
Yoga	ECG	548.667 ± 57.106	39.491	0.426	34.83 ± 12.729	34.833	-0.603	0.086
	PPG	542.500 ± 31.691						
Reading	ECG	595.833 ± 24.145	28.341	-0.653	-15.500 ± 25.990	15.5	2.646	0.725
	PPG	611.333 ± 29.891						
Writing	ECG	710.833 ± 54.756	21.852	0.809	14.166 ± 20.993	14.167	-1.677	0.986
	PPG	700.333 ± 74.161						
Videocall	ECG	745.500 ± 8.643	6.621	0.296	-1.833 ± 6.969	1.833	0.251	0.6
	PPG	747.333 ± 7.891						
Coding	ECG	722.833 ± 31.708	16.386	0.679	12.166 ± 16.630	12.166	-0.916	1
	PPG	716.667 ± 45.315						

Table C.10: Error evaluation between ECG and PPG-derived pNN50 values during repeated experiment

Dataset	Sensor type	pNN50	RMSE	R^2 score	Bias	MAE	MRE	ρ
Yoga	ECG	3.887 ± 2.448	16.179	-51.431	-15.713 ± 4.224	15.713	603.855	0.029
	PPG	19.600 ± 4.860						
Reading	ECG	1.880 ± 0.923	7.149	-70.927	-6.280 ± 3.744	6.28	377.102	0.314
	PPG	8.160 ± 4.022						
Writing	ECG	4.908 ± 2.340	8.556	-15.045	-7.278 ± 4.926	7.278	149.287	0.657
	PPG	12.187 ± 6.293						
Videocall	ECG	3.250 ± 1.657	11.266	-54.487	-11.072 ± 2.283	11.072	524.181	-0.486
	PPG	14.322 ± 1.041						
Coding	ECG	1.392 ± 0.749	9.538	-193.626	-9.427 ± 1.588	9.427	980.221	0.086
	PPG	10.818 ± 1.460						

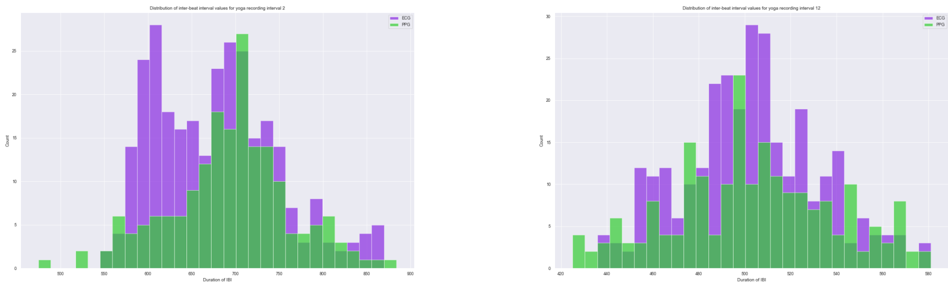

Figure C.31: Frequency distribution of the cleaned 5-min sample for ECG and PPG recordings during yoga within repeated experiment

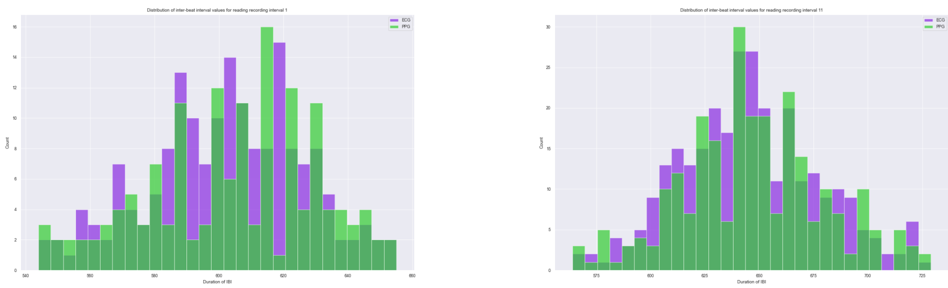

Figure C.32: Frequency distribution of the cleaned 5-min sample for ECG and PPG recordings during reading within repeated experiment

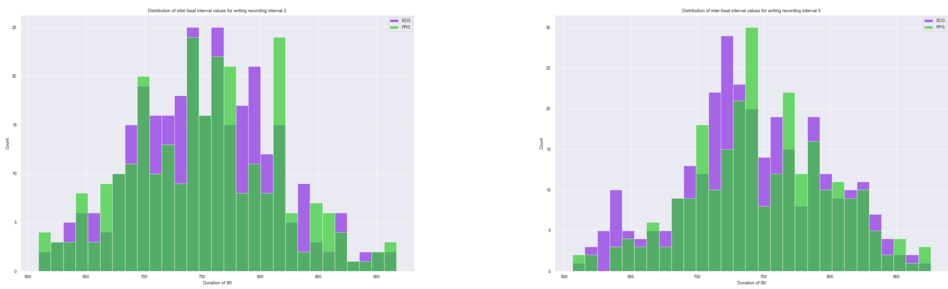

Figure C.33: Frequency distribution of the cleaned 5-min sample for ECG and PPG recordings during writing recording within repeated experiment

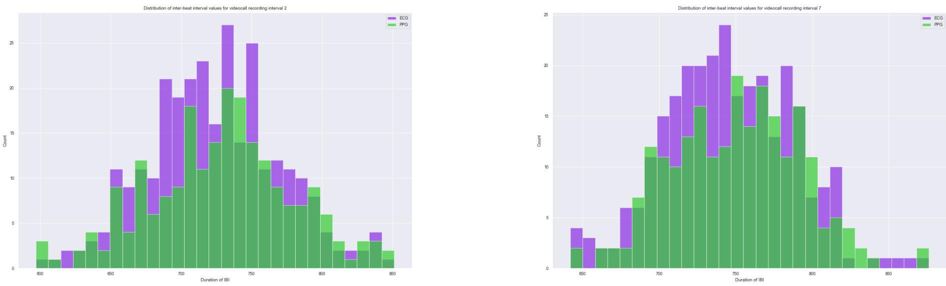

Figure C.34: Frequency distribution of the cleaned 5-min sample for ECG and PPG recordings during videocall recording within repeated experiment

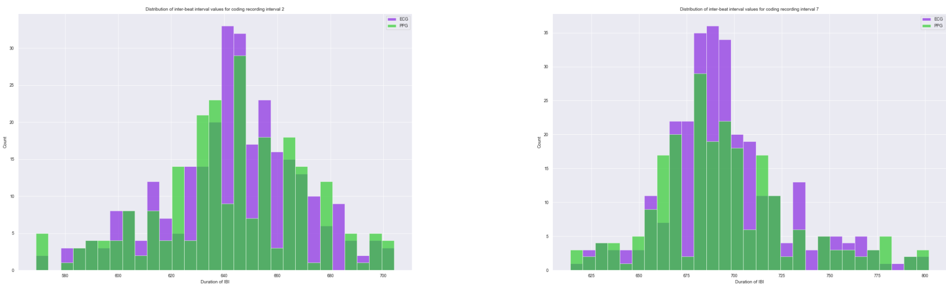

Figure C.35: Frequency distribution of the cleaned 5-min sample for ECG and PPG recordings during coding recording within repeated experiment

Table C.11: Error evaluation between ECG and PPG-derived rMSSD values during repeated experiment after outliers elimination

Dataset	**Sensor type**	**rMSSD**	**RMSE**	R^2 **score**	**Bias**	**MAE**	**MRE**	ρ
Yoga	ECG	36.517 ± 11.652	10.897	-0.049	-7.627 ± 10.352	7.627	26.598	0.543
	PPG	41.944 ± 5.301						
Reading	ECG	18.467 ± 4.771	12.468	-7.195	-11.268 ± 5.846	11.268	69.185	0.543
	PPG	29.735 ± 5.788						
Writing	ECG	23.933 ± 4.498	12.059	-7.625	-11.319 ± 4.557	11.319	47.266	0.638
	PPG	35.253 ± 7.634						
Videocall	ECG	21.917 ± 2.192	16.596	-67.770	-16.115 ± 4.344	16.115	74.869	-0.087
	PPG	38.031 ± 3.941						
Coding	ECG	17.750 ± 3.098	18.600	-42.268	-17.626 ± 6.506	17.626	104.048	0.2
	PPG	35.376 ± 5.471						

Table C.12: Error evaluation between ECG and PPG-derived SDNN values during repeated experiment after outliers elimination

Dataset	**Sensor type**	**SDNN**	**RMSE**	R^2 **score**	**Bias**	**MAE**	**MRE**	ρ
Yoga	ECG	38.933 ± 10.984	11.291	-0.268	-3.794 ± 11.650	3.794	15.667	-0.029
	PPG	42.728 ± 8.395						
Reading	ECG	29.300 ± 6.075	6.475	-0.363	-1.346 ± 6.938	1.346	7.065	0.314
	PPG	30.646 ± 6.823						
Writing	ECG	43.000 ± 9.098	5.351	0.585	-2.462 ± 5.204	2.462	4.663	0.812
	PPG	45.462 ± 12.300						
Videocall	ECG	39.850 ± 6.093	7.371	-0.756	-3.178 ± 7.285	3.178	10.357	-0.429
	PPG	43.028 ± 2.728						
Coding	ECG	32.633 ± 5.427	5.163	-0.087	-4.135 ± 3.842	4.135	12.264	0.657
	PPG	36.422 ± 5.560						

Table C.13: Error evaluation between ECG and PPG-derived Mean RRI values during repeated experiment after outliers elimination

Dataset	**Sensor type**	**Mean RR**	**RMSE**	R^2 **score**	**Bias**	**MAE**	**MRE**	ρ
Yoga	ECG	548.667 ± 57.106	47.431	0.172	-1.296 ± 51.938	1.296	0.768	-0.086
	PPG	549.963 ± 44.894						
Reading	ECG	595.833 ± 24.145	33.033	-1.246	-18.066 ± 30.294	18.066	3.1025	3.103
	PPG	613.900 ± 30.155						
Writing	ECG	710.833 ± 54.756	31.238	0.609	14.868 ± 30.094	14.868	-2.364	0.986
	PPG	695.965 ± 81.976						
Videocall	ECG	745.500 ± 8.643	5.545	0.506	-0.868 ± 5.998	0.868	0.121	0.771
	PPG	746.368 ± 6.770						
Coding	ECG	722.833 ± 31.708	21.516	0.447	8.435 ± 21.683	8.435	-1.267	1.0
	PPG	714.398 ± 52.103						

Table C.14: Error evaluation between ECG and PPG-derived pNN50 values during repeated experiment after outliers elimination

Dataset	Sensor type	pNN50	RMSE	R^2 score	Bias	MAE	MRE	ρ
Yoga	ECG	3.887 ± 2.448	18.204	-65.370	-17.968 ± 3.196	17.968	665.984	0.543
	PPG	21.855 ± 3.876						
Reading	ECG	1.880 ± 0.923	7.534	-78.861	-6.509 ± 4.156	6.509	393.369	0.371
	PPG	8.389 ± 4.405						
Writing	ECG	4.908 ± 2.340	9.639	-19.370	-8.611 ± 4.747	8.611	189.531	0.829
	PPG	13.519 ± 6.058						
Videocall	ECG	3.250 ± 1.657	11.605	-57.878	-11.039 ± 3.924	11.039	541.620	-0.314
	PPG	14.289 ± 3.051						
Coding	ECG	1.392 ± 0.749	10.199	-221.553	-9.958 ± 2.415	9.957	1049.129	-0.2
	PPG	11.349 ± 2.280						