Exzellenzcluster
Cognitive Interaction Technology
Kognitronik und Sensorik
Prof. Dr.-Ing. U. Rückert

Design and Application of Wireless Body Sensors

zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEUR (Dr.-Ing.)

der Technischen Fakultät
der Universität Bielefeld

genehmigte Dissertation

von

Timm Hörmann

Referent: Prof. Dr.-Ing. Ulrich Rückert
Korreferentin: Prof. Dr.-Ing. Britta Wrede

Tag der mündlichen Prüfung: 19.12.2019

Bielefeld / 2019

Printed on acid-free paper according to ISO 9706.

Contents

1 Introduction

Research on *wireless body sensors* is an interdisciplinary field bringing together science on living and technical systems. It is related closely to biomedical engineering but relaxes the constraint to be of medical use. Thus, the field of application for body sensor technology is wide.

The use of body sensors is often intended to satisfy individual needs. In this respect, an example is found with the Quantified-Self community, which represents a mindset of gaining "self-knowledge through numbers" [5]. Furthermore, body sensors can be used in team-sports, e.g. to optimize training [140] or in occupational health, e.g. to maintain health and well-being [219]. Moreover, they are of general interest in public or digital health-related issues [18, 179].

As a consequence, body sensors are of great interest in different scientific disciplines, and their use becomes more and more widespread. This can be seen, for example, in the steadily rising number of publications [41] relating to the keywords "body sensor", "fitness tracker", or "wearable device" (Figure 1.1).

On the one hand, the construction of wireless body sensors is a challenge in itself (e.g. energy optimizations or miniaturization) from an engineering point of view. On the other hand, the application and utilization of the data obtained with body sensors are valuable in the conduct of experiments in sports-, social- or health sciences, and in psychology. Likewise, other scientific disciplines like machine learning, data science, and signal processing have their hands on the data obtained using body sensors.

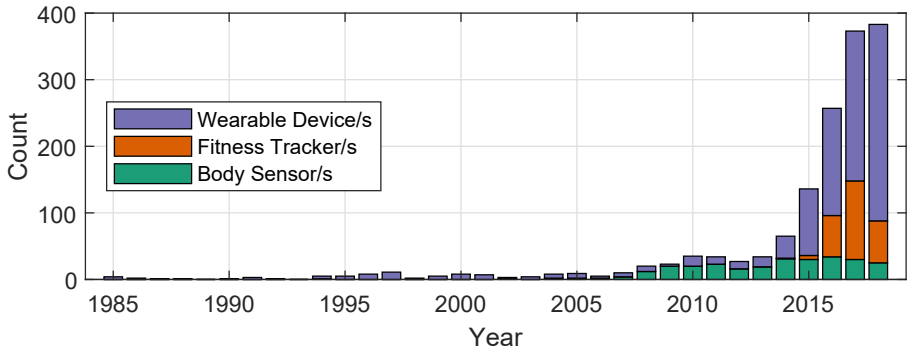


Figure 1.1: Interest in "Body Sensor"-technology (including "Fitness Tracker/s", "Body Sensor/s" and "Wearable/s") is growing. This can be seen, for example, in the increasing number of publications in the PubMed database (publications counts are not normalized per year).

1.1 Contribution

The contributions of this thesis cover both the design and the application of wireless body sensors.

At first, the practical implementation of a compact, highly-integrated, and scalable wireless body sensor is revealed. With the outline of the design's implementation, requirements and matching technological components are identified. In addition, selected applications or algorithms are evaluated in terms of energy consumption. In summary, the results provide insights into inevitable trade-offs affecting the design of wireless body sensors, e.g. concerning the physical dimensions, power consumption.

Moreover, this thesis adds up new results and research data regarding the use of wireless body sensors for occupational health and safety. Therefore, practical experiments are conducted to estimate cognitive workload and physical activity. The obtained results complement the scientific consensus and furthermore clarify application-specific aspects of body sensors, regarding field use. In addition, with a detailed analysis of the results, it is explained when and why wireless body sensors reach their limits and what measures can be taken to achieve valid results.

Parts of this thesis were previously presented at international conferences, and excerpts were published in the corresponding conference proceedings [269, 270, 271, 272, 274, 275], as journal article [273] or book chapter [276]. The main contributions of this thesis can be summarized as follows:

1. The design process of a wireless body sensor (BI-Vital, Bielefeld-Vitalmonitor, version 5.0) is presented:
   - Specific requirements and design aspects are identified and discussed. Also, challenges, issues, and trade-offs are pointed out. As a result, a fully-functional module is realized in practice.
   - The device's functionality is verified. As it follows on from a previous version, it is compared and evaluated against its predecessor. Its degree of maturity corresponds to that of a technical prototype or pre-series product.
   - The device's applicability for advanced embedded-inference methods, here arrhythmia detection, is demonstrated and evaluated in terms of latency and energy consumption.
   - As a by-product of this thesis, a small batch of this module was produced, which will be used for teaching purposes in the future.
2. The use of body sensor technology in the professional context (occupational health), to detect or estimate cognitive workload is presented:
   - An experiment is conducted, which is highlighting the applicability and effectiveness of body sensors for psycho-physiological measures.

- The possibility to realize a fine-grained estimation of cognitive workload on the basis of short-time signals is illustrated.
- Besides, the presented results add to the scientific community, by highlight indications of the uncertainty in self-reported cognitive workload, which was again confirmed in other works.

3. The method of physical activity estimation using wireless body sensors is demonstrated. This is in the context of occupational safety addressing the user group of firefighters.
   - An experiment is conducted, which has its focus on differences in physical activity due to the use of personal protective equipment. The results complement the scientific community with insights on the applicability of wireless body sensors for physical activity estimation under more realistic conditions.
   - The results strengthen the so far ambiguous scientific consensus on the necessary complexity of machine learning models for the estimation of energy expenditure.
   - As a secondary outcome, the experimental results provide insights on the exercise intensity during the physical ability test (G26.6) mandatory for firefighters in Germany who are using respiratory protective equipment.

1.2 Outline

This thesis is structured (Figure 1.2) as follows:

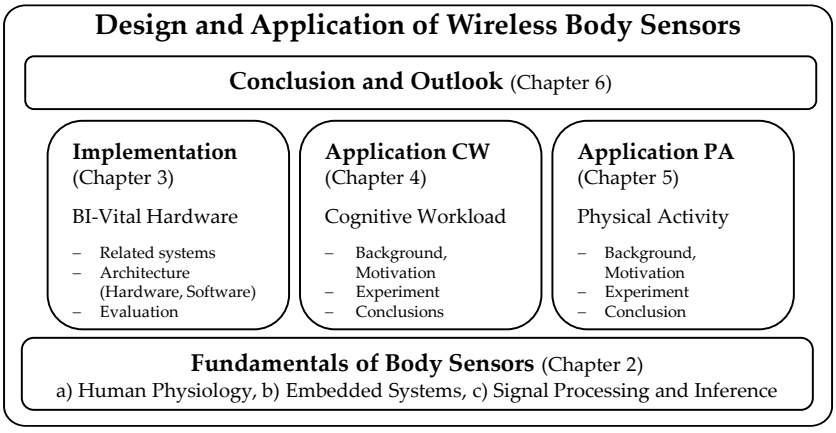


Figure 1.2: Outline of the Thesis

In chapter 2, the fundamentals of body sensors are introduced to provide basic knowledge of the disjunctive scientific sub-areas that flow into this thesis. In summary, state-of-the-art knowledge on human physiology is discussed (sec. 2.1), followed by the origins, design, and technical aspects of body sensors (sec. 2.2). The chapter closes with an introduction to embedded algorithms and standard machine learning methods. Both are extensively addressed in the following chapters and thus form the fundamental basis for the evaluation and analysis of the experimental data presented in this thesis.

The technical design aspects of wireless body sensors are addressed in chapter 3. Conventional devices available on the market, for consumers or the scientific community, are presented (sec. 3.1), and an extensive requirement analysis (sec. 3.2) on the redesign of a predecessor device used at Bielefeld University is given. The implementation of the newly designed wireless body sensor *BI-Vital* is outlined in the subsequent (sec. 3.3). The chapter closes with a prospect on possible future applications and design extensions of the *BI-Vital*.

In chapter 4 and chapter 5, the focus is on the practical application of body sensors. Both chapters address professional use in the field of occupational health and safety.

In chapter 4, firstly, an overview of the concept of cognitive workload and measurement methods is given (sec. 4.2). Afterward, an experimental setup to provoke cognitive workload in participants who are working on a tablet computer, methods to process psycho-psychological measures (based on heart rate, heart rate variability and electrodermal activity), and models to predict cognitive workload by means of machine learning methods are presented (sec. 4.3). The chapter closes with a conclusion and remarks on future work regarding mobile and unobtrusive cognitive workload estimation (sec. 4.4).

In chapter 5, state-of-the-art methods to estimate physical activity are presented and applied to the special user-group of firefighters. The chapter gives a detailed overview of techniques to quantify physical activity (sec. 5.2-5.3). Subsequently, in-depth analysis of the data recorded, steps to process and utilize the information for physical activity estimation are revealed (sec. 5.5). The chapter closes with a conclusion and remarks on future work to the subject of physical activity estimation.

The thesis's results are concluded in chapter 6. Finally, a prospect for future research in the field of wireless body sensors is given.

2 Fundamentals of Body Sensors

Body sensors make up an interdisciplinary field of research that connects science about living and technical systems. It covers the measurement and analysis of data from the human body by means of miniaturized wearable electronic devices. The term *body sensor*, however, is multifaceted and only vaguely defined [57, 212]. It is used in this work to distinguish from the branch of *biomedical engineering* and *wearable computing* (or *wearable technology*). The former has its focus on medical applications, while the latter does not necessarily involve any sensors at all.

In this work, the term *body sensor* is used in order to highlight two things: Firstly, it highlights the use of electronic (and in some circumstance wireless) wearable sensor systems, which are about to be distinguished from medical or diagnostic equipment. Secondly, it emphasizes the broader view of the application, including non-medical use-cases. Following this point of view, the design and application of body sensors remain closely related to biomedical engineering, however, loosen the restriction of its scope to medical applications. In contrast, body sensors are designed to satisfy individual needs [5], can be used in team-sport [140], and are of interest in public health-related issues [179].

This chapter is intended to provide a basic understanding of the topics of living systems (i.e. human physiology, sec. 2.1) and technical systems (i.e. wireless electronic devices or sensor nodes, sec. 2.2). In the following, both branches are brought together with the introduction of wireless body sensors (sec. 2.2.4). At the end of this chapter, the general concept of data processing, and specific algorithms for wireless body sensor (WBS) applications are introduced (sec. 2.3).

2.1 Human Physiology

Physiology is about the characterization of all chemical or physical interactions that make a living system. The concepts are fundamental for understanding and predicting the behavior of living systems. In its broadest definition, behavior includes any interaction with the physical environment. It could be defined as the summary of energetic transduction between a living system (a human) and its physical environment. From a less physical but more physiologic point of view, (human) behavior is all observable muscular and secretory responses. An even narrower definition would summarize behavior as motion. These abstractions set the big picture on the use of body sensors, which is to observe and reflect human physiology (direct or indirect) in order to make predictions towards or simply observe human behavior (abstract or specific).

Starting from the broadest definition, the energy transfers in a living system, it can be seen that total energy expenditure indeed is of particular interest. However, due to the complexity of living systems, and in particular of human behavior, the characterization of energetic transfers alone does not provide an adequate full picture.

As in biology, the use of body sensors includes the perception of motion not only in terms of energy expenditure but also in terms of a characterization or classification of physical activities human activity recognition (HAR). Moreover, in sports science, the analysis of sequences of movements, the kinetics, seen individually or linked to bodily functions, is of interest. Finally, not only motion behavior is of interested, but also psychological effects. An example is the subjectively perceived cognitive workload (CW) of a person. In addition, more complex behavior like the interaction of individuals in a group could be of interest in body sensor applications.

This chapter gives a brief overview of the energy use, transfer, and balance in human from a physiological point of view. The starting point is the top view on energy transfer, beginning with movement and muscle contraction going over to the control mechanism in the nervous system. Thereby the working principle of the regulatory systems is explained, which are involved in controlling observable values such as heart rate, perspiration, or respiration. This is important to understand the contribution of physiological and mental strain. In the following, the link between physical and physiological work is discussed.

2.1.1 Energy Expenditure

Physically speaking, behavior (sports, physical work, or any physical activity (PA) in general), can be seen as heat transfer (Figure 2.1). This is because any change in the internal state of an organism (metabolism) is based on chemical processes that have a certain heat loss. The same applies to external work because the muscle in the human body needs energy to contract (i.e. move). This energy transfer in the human body is called the metabolic rate or energy expenditure (EE). Its observable values are the heat loss, the energy transferred through work, and the energy stored (chemically). According to the first law of thermodynamics, these are in equilibrium with the energy up-taken (trough food and oxygen). [111, pp. 393]

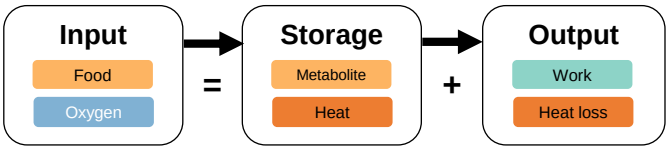


Figure 2.1: Big picture of the energy flow in the human organism. Chemical energy enters the system through food, is transformed (metabolized), partly stored in the organism, and finally transferred by work or dissipated as heat. [111, pp. 393]

The universal physiological energy store is adenosine triphosphate (ATP). Because of its universal nature, ATP is also used as a unit of energy measure in biological systems (eq. 2.3 - 2.4). The sum of consumed ATP makes up the entire physiological energy expenditure (EE). Instead of ATP, EE can also be directly quantified in the physically derived SI unit joule (J) or kilojoule (kJ) for energy, work, and heat (eq. 2.1). In conjunction with the energy content of food, often, the outdated unit calorie (cal) or kilo-calorie (kcal) (eq. 2.2) is found. [111, p. 405]

$$1\,\mathrm{J} = 1\,\mathrm{W/s} \qquad = 0.2388\,\mathrm{cal} \tag{2.1}$$

$$4.1868\,\mathrm{J} = 4.1868\,\mathrm{W/s} \qquad = 1\,\mathrm{cal} \tag{2.2}$$

$$ATP + H_2O \rightarrow ADP + Pi \quad +\Delta G^\circ \quad \Delta G^\circ = -30.5\,\mathrm{kJ/mol} = -7.3\,\mathrm{kcal/mol} \tag{2.3}$$

$$ATP + H_2O \rightarrow AMP + PPi + \Delta G^\circ \quad \Delta G^\circ = -45.6\,\mathrm{kJ/mol} = -10.9\,\mathrm{kcal/mol} \tag{2.4}$$

The daily EE of an adult human is in the order of 8000 kJ to 16 000 kJ (or 1911 kcal to 3822 kcal). This is equivalent to an average power of 93 W to 185 W. As this EE comprises different components, it is also called total energy expenditure (TEE). It includes the basal metabolic rate (BMR), the thermic effect of food (TEF) and the physical activity-related energy expenditure (PAEE) (Figure 2.2, eq. 2.5) [160, pp. 193]. Other components affecting TEE are, for instance, growth or pregnancy, which are usually neglected, since these only apply in special conditions [31, p. 26]. The BMR or resting metabolic rate (RMR), the TEF, and other factors are more or less static. They depend on age, size, and sex. In contrast, PA is a dynamic fraction of TEE depending on behavior only.

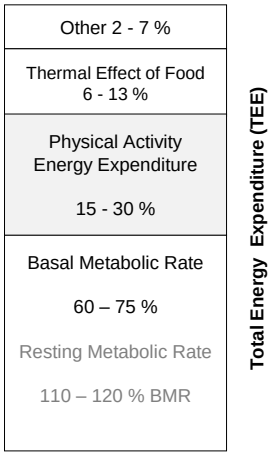


Figure 2.2: Fractions of the TEE [31][p. 26]

$$\mathrm{TEE} = \mathrm{BMR} + \mathrm{TEF} + \mathrm{PAEE} \tag{2.5}$$

As per definition, the BMR is the amount of TEE needed to keep the organism alive. Typically, BMR accounts for 60 % to 75 % of TEE. The BMR depends on anthropomorphic values of a person, like age, weight, sex, or healthiness. It is considered a static value, although it is not a real constant. Instead, external factors, e.g. environmental temperature, affect the BMR. This is why, often, RMR is refereed to in experiments and literature instead of the BMR. The concept of RMR relaxes the strict constraints of BMR measure, which requires a fixed ambient temperature of 28 °C and that the participant is lying and fasting.

The TEF is the fraction needed to convert food into physiological energy by digestion. It makes up 6 % to 10 % of the TEE. The terms food-induced thermogenesis (FIT), nutrient-induced thermogenesis (NIT), or diet-induced thermogenesis (DIT) are used anonymously. In contrast to the RMR and the TEF, PAEE is a highly dynamic component of the TEE. It depends on behavior only. On average, 15 % to 30 % of the TEE can be attributed to PA. However, the percentage can be considerably higher or lower for a particular way of living.

2.1.2 Cellular Respiration

For the human organism to be able to convert energy at all, it must be available in a physiologically usable form. The universal metabolite to store and provide energy is ATP. With ATP, energy is stored in the 3 pyrophosphate linkages. It can be released through hydrolysis, where either only 1 or 2 of the terminal pyrophosphate linkages are separated from the ATP. These processes yield adenosine diphosphate (ADP) or adenosine monophosphate (AMP). Both ADP and AMP are re-synthesized to ATP in different metabolic pathways (Figure 2.3), which are briefly discussed in the following (based on [252, pp. 79] and [160, pp. 143]).

The energy metabolism starts with the intake of food and its digestion. Besides, water is a primary component. The main ingredients of food can be separated into proteins, fats, and carbohydrates. Proteins are further processed into amino acids, fats into fatty acids and glycerol, carbohydrates into glucose. These ingredients are then metabolized either aerobically using oxygen or anaerobically without oxygen. On the cellular level, metabolism is divided into catabolic pathways (catabolism), in which free energy is released by breaking down complex molecules and anabolic pathways (anabolism), in which complex molecules are build up consuming free energy.

The *aerobic metabolism* is a complex combination of multiple (long) metabolic pathways. In summary, glucose (from carbohydrates) and glycerol (from fat) enter a process named glycolysis (or lipolysis), which results in the synthesis of pyruvate. This process has a positive energy balance, which yields 2 mol ATP. The pyruvate itself is further metabolized into acetyl coenzyme A (Acetyl-CoA).

In a separate process, the β-Oxidation, fatty acids (non-glycerol parts of fat) are also cleaved into Acetyl-CoA. This process, however, has a negative energy balance and requires 2 mol ATP. Amino acids (from proteins) are involved in this process in multiple ways, adding up to pyruvate or Acetyl-CoA synthesis and other steps of the *Krebs-cycle*.

In the Krebs-cycle, the Acetyl-CoA complex is broken down step-wise into carbon dioxide, water, and free hydrogen-ions using oxygen. This process alone yields 2 ATP. The hydrogen-ions are finally processed by oxidative phosphorylation, which adds to ATP synthesis mostly. In total, the oxidative metabolism of carbohydrates yields 38 mol ATP (glucose). In direct comparison, the oxidative metabolism of fats provides several times more ATP. It, however, depends on the specific fatty acid (i.e. 1 mol palmitic acid yields 129 mol ATP).

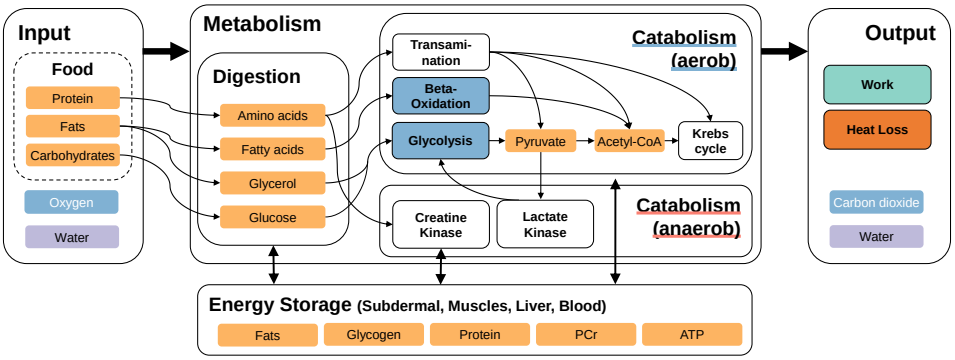


Figure 2.3: Energy metabolism is about energy transfer in the human body. It roughly consists of energy intake through food, its digestion and metabolic processing, and final its observable output, which is work, heat loss, and excretion of metabolic end-products. The metabolic pathways can be characterized as either aerobic or anaerobic. They differ in energy capacity and speed. The processes depicted are only excerpts of a much more complex system, which includes more details and further interactions.

Regarding *anaerobic metabolism*, phosphorylated creatine (PCr) and lactate are of interest. PCr is used to phosphorylate ADP back to ATP directly. It is a very fast energy source, comparable to ATP itself (1 mol PCr yields 1 mol ATP). In the absence (or lack of sufficient amounts) of oxygen, pyruvate is metabolized into lactate. The metabolism of lactate (anaerobic glycolysis) yields 2 mol ATP per 1 mol pyruvate. From an energetic point of view, lactate metabolism is not favorable, but it is advantageous because it is faster compared to oxidative metabolism.

While in the past, lactate was mainly seen as an end product, today, it is also understood as a buffer useful during recovery or less exerting exercises. Lactate also acts as a shuttling metabolite transferring energy from one cell to another.

The metabolites are not always directly converted into ATP or used for other metabolic pathways. If the ATP and PCr stores in blood and muscles are filled, the remaining metabolites are stored. Smaller energy stores for carbohydrates are present in the liver (2.4 MJ) and muscles (5 MJ). Most of the energy is stored in fats, either directly in the muscles (8.4 MJ) or subcutaneously (427 MJ).

The pathways and energy balance depicted are only excerpts of a much more complex metabolism. However, it becomes clear that the availability and storage capacities of specific metabolites and the complexity (e.g., set-up time) of the corresponding pathways affect physiological energy transfers. Pre-stored ATP and anaerobic metabolism of PCr are the quickest and thus firstly depleted storages in case of heavy, short-term exercise. Metabolism of fats, in contrast, is a slowly progressing process, which in turn could offer energy for up to several days.

2.1.3 Cardio-respiratory System

During exercise, high amounts of energy are converted continuously. The short-term peak performance thereby depends on the ability of the muscles to store energy, namely in the form of ATP and PCr. The maximum endurance capacity, in turn, depends on the efficiency of the aerobic metabolisms (glycolysis and β-Oxidation). Both pathways require oxygen to be transported into the muscles. During exercise, a shift in energy supply mechanism can be observed from *fast* to *slow* energy sources (Figure 2.4).

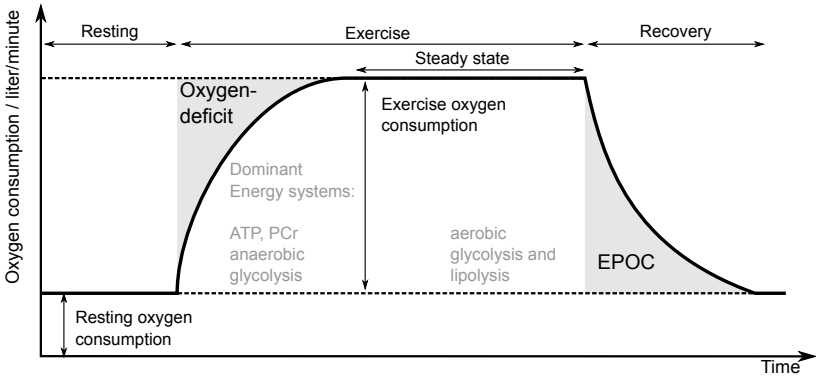
In order to provide oxygen to the muscles in the first place, oxygen needs to be taken up through the lungs into the blood. Afterward, oxygen-saturated blood is transported directly to the (muscle-) cells. Consequently, it can be observed that, during exercise, the cardiac output and respiratory minute volume increases. Both are the result of heart or respiration rate multiplied by the stroke volume or tidal volume, respectively. In turn, the total amount of blood, and within it, the total amount of oxygen pumped through the body rises. [160, p. 296]

Regarding cardiac output, firstly, the stroke volume increases. After that, heart rate rises virtually linear in accordance with the intensity of PA. Just like cardiac output, the respiratory minute volume is firstly increased by raising the tidal volume. While more air enters the lungs, also the amount of oxygen up-taken from the air is increased (ventilation equivalent). In addition, the respiration rate accelerates.

Simultaneously to the increased activity of the cardiac and ventilatory system, body temperature rises. This is partly explained, due to the aerobic metabolic pathways (glycolysis and β-Oxidation), e.g. the cellular respiration and the muscle activity, which are excited. As a consequence, the body temperature rises, which in turn causes an increase in perspiration. Perspiration and blood circulation within the skin is adapted to establish an equilibrium between produced heat and heat loss, also to prevent a heat stroke. Additionally, due to vasoconstriction in specific organs and vasodilation in the muscles used, the blood flow is adapted. In this way, blood flow through the skin and the muscles is increased. This allows the energy transfer to the muscles to be facilitated and at the same time, heat loss via the skin to be regulated.

If the intensity of an exercise is not rapidly changed, an equilibrium between internal stress (physiological energy supply) and external stress (physically performed work) establishes from a certain point in time (Figure 2.4a). This is known as *homeostasis* or *steady-state*. Thereby, the maximum performance is primarily determined by the maximal oxygen uptake rate (OUR), which limits the aerobic metabolic pathways and is also known as aerobic capacity (VO_2max).

Before and after the steady-state is reached, a certain time delay between physical work and OUR can be observed. This is due to the fact that with the onset of a PA, the different metabolic pathways must be activated first, which takes a specific time. After exercising, the equilibrium of metabolisms found during resting-conditions needs to be re-established. This includes re-synthesis of AMP and ADP to ATP or lactate into glycogen. Among other factors, this causes oxygen uptake to be high after exercise, in



(a) Excess Post-Exercise Oxygen Consumption. Polarlys, CC BY 2.5, translated and desaturated, https://commons.wikimedia.org/wiki/File:Excess_post-exercise_oxygen_consumption.png

Metabolite, ***Pathway***	ATP and PCr	Carbohydrates, *anaerobic glycolysis*	carbohydrates, *aerobic glycolysis*		fat, *aerobic lipolysis*
Duration, ***Type***	immediate (< 45 s) *anaerobe*	short (120 s) mostly *anaerobe*	medium (2 – 8 min.) *aerobe* and *anaerobe*	long (8 – 60 min.) mostly *aerobe*	long (> 60 min.) *aerobe*

(b) Temporal dynamics in energy contribution. Adopted from [93, p. 95].

Figure 2.4: The relative contribution of the energy metabolism pathways shows a temporal dynamic (b). During exercise, high amounts of energy need to be converted continuously. Due to the temporal dynamics, PA and oxygen uptake (aerobic energy contribution), however, is not perfectly aligned. This is explained by the fact that the metabolism needs a specific time to adopt. This effect is known as excess post-exercise oxygen consumption (EPOC) (a).

resting conditions. This effect is known as excess post-exercise oxygen consumption (EPOC) [160, p. 171-173]. Likewise, oxygen uptake, cardiac output, and respiratory minute volume are lagging behind abrupt changes in energy demand.

Despite this lag, a shift from aerobic to anaerobic metabolism can be observed. Especially in long-lasting PA, e.g. endurance training, the concentration of the anaerobic metabolite lactate rises. An accumulation of above 4 mmol/L in atrial blood is often defined as the point where exercise intensity exceeds the aerobic capacity. It is denoted as the *anaerobic threshold*. Similarly, this point can be found with ventilation and heart rate. For instance, it can be observed that the otherwise linear relation between the ventilation volume (or heart rate) and OUR is broken at a certain point. This point is known as the *ventilatory anaerobic threshold* or the heart rate's *deflection point* [37]. Both loosely coincide with the anaerobic threshold and are used to characterize the transition from aerobe to anaerobe endurance capacity [94].

2.1.4 Control and Regulation

The body reacts to intense PA in a chain of multiple auto-regulations in order to maintain *homeostasis* (sec. 2.1.3). To these auto-regulations, specific reflexes and corresponding receptors are known (Figure 2.5). Regarding exercise response reactions, chemo-, mechano- and baroreceptors are of interest, which belong to the metabo-, mechano-, or baroreflex, respectively [39, 77]. The metaboreflex describes the direct reactions to concentrations changes of metabolites like ATP or PCr. The mechanoreceptors react to muscle contractions. The baroreflex (triggered by baro-receptors) reduces blood pressure by decreasing heart rate. All those mechanisms attribute to keep or re-establish a metabolite equilibrium (i.e. *homeostasis*) within the (muscle-) cells and to regulate blood pressure or flow. Moreover, they affect each-others sensitivity. [77]

The homeostatic reflexes are part of the autonomic nervous system (ANS), which coordinates the control of bodily function. In addition, they provide information to the *central control* in the brain or central nervous system (CNS). The ANS itself is not part of the CNS but the peripheral nervous system. Bodily functions are thus indirectly controlled through the ANS because it transmits signals from the CNS to the peripheral organs (as being part of the peripheral nervous system). Because of its mediating function, it can be understood as a mainly efferent system (transmitting messages towards the organs and not vice versa). The ANS itself is divided into the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). [51]

The PNS is often summarized as the *rest and digest* system because it stimulates salivation and digestion, while it decreases heart rate and respiration The SNS, in contrast, is typically denoted as the *fight or flight* system since it inhibits digestion but stimulates energy metabolism in the liver, enhances respiration and increases heart rate. Additionally, the SNS excites sweat glands and thus increases perspiration. The PNS and SNS are sometimes seen as antagonists. However, this is a simplified view. While some bodily functions are exclusively excited through either the SNS or PNS for other systems, e.g. the cardiovascular system, this does not apply. The cardiovascular system is not simply turned on or off by the SNS or PNS. Instead, interactions of both systems contribute to the control of heart rate. In this respect, it can be found that during exercise, heart rate is firstly increased solely due to the missing inhibition of the PNS. Later, as PA continues, heart rate is further increased due to additional activation through the SNS. [51, 164]

In addition to the described auto-regulations, higher brain regions affect the responses of the ANS. This control is often addressed as the central command, triggering the adaptation of bodily functions through the ANS. However, the exact working mechanisms (due to the complexity of the brain), are unknown [255]. What is known is that signals from the various receptors are combined in the brain or CNS[1]. Thus, regulatory responses can be observed even in the absence of PA. Taking, i.e. stressful situations,

[1] Mechano-, metabo- and baroreceptors were mentioned here as they affect heart rate the most. Besides, other receptors (e.g. thermo or pulmonary) and mechanism (e.g. pulmonary activity) do have influence.

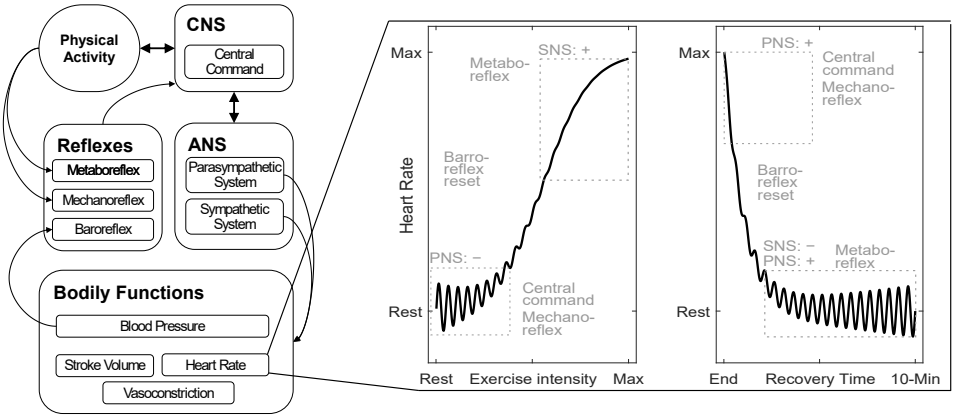


Figure 2.5: Excerpt on the regulation of heart rate during rest, exercise, and recovery. Higher brain centers activate muscles and ANS through central command. This causes anticipatory adaptations of bodily functions: Changes in heart rate, stroke volume, and vasoconstriction are causing blood pressure variations, which affects the baroreflex. Together with the metabo-, and mechanoreflex (caused by activation of the muscles), this attributes to the compensating adaption of bodily functions. During exercise, the anticipatory effect firstly decreases the activity of the PNS, which in turn raises the heart rate. Later with increasing intensity, SNS predominates control on heart rate. The reverse pattern is found during recovery. Adopted from [164]

which require increased arousal the metabolism is adapted beforehand. These mechanisms are beneficial in order to prepare the body to excessive PA. A well-known example is the fight-or-flight reflex, initiated by central command and controlled by the SNS.

Also, in athletes, an anticipatory effect on heart rate and blood pressure prior to PA can be observed [160, p. 331]. It can be summarized that bodily functions are regulated in a compensatory and anticipatory way. The central command from higher brain centers (CNS) to the locomotive system, e.g. the muscles, that induce movements (or to plan them) do also activate the ANS, which in turn adapts metabolism. This is the anticipatory effect on homeostasis. With increased or prolonged PA, the compensatory mechanisms (auto-regulation) predominate control. This is the compensatory regulation of bodily functions.

In a first approximation, the observable bodily reactions, however, remain the same: The heart beats faster, respiration increases, and most often also perspiration occurs. However, due to the different chains of action, variations can be observed, especially in heart rate (which is controlled by the PNS and the SNS) and perspiration (controlled exclusively by the SNS). One the one hand, these reactions allow, at least partially, to make conclusions on the activation of the SNS and PNS. On the other hand, when estimating energy expenditure, the non-activity related responses of the cardiovascular system must be taken into account, in order to avoid overestimation.

2.2 Wearable Sensor Systems

In this work, the focus is on sensor systems known as wireless body sensors (WBSs). These devices are specialized entities of *embedded systems* worn closely or directly on the human body. They are used to recording physiological, kinematic, or environmental data. Research on the design and application of such devices has a long history, reaching from simple and more sophisticated mechanical devices, over first, yet unwieldy, electronic variants, up to the miniaturized devices that are available today.

The history of body sensors is resumed in short as part of this chapter (sec. 2.2.1). This is done in order to classify the different device categories, their fields of application, and the research paradigms associated with WBS. Therefore, for instance, the concepts of *Wearables*, *Handhelds*, or *Mobile Devices* are distinguished from each other (sec. 2.2.2). From a technological perspective, however, WBS can be seen as a subgroup of *wireless sensor nodes* (sec. 2.2.3). At the end of this section, related categories, paradigms, and technological aspects are brought together in order to define the domain of WBSs (sec. 2.2.4).

2.2.1 Origin and History

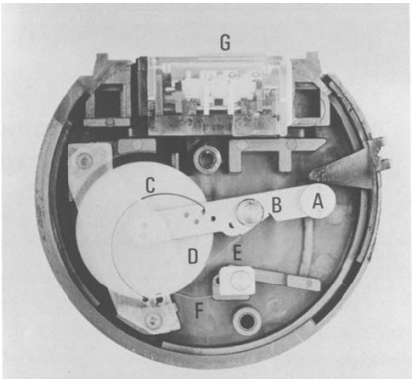
Today, many WBS exist, some of which are sold as commercial products. Most devices are designed for private use or research and development purpose. However, having devices that keep track of bodily functions is not a new idea. The first examples of devices that can be described as (non-wireless) wearable body sensors can already be found as early as in the 15th century. One often found example is the sketch of a mechanical pedometer shaped like a pendulum by Leonardo da Vinci. It was intended to be used as a military device to track the daily distance soldiers had traveled and built up maps. Most likely, however, it was never built.

Later, in the 18th century, Abraham-Louis Perrelet invented the first automatic pocket watch. It was self-winding in the user's pocket while it is walking. Shortly after his invention, Perrelet re-used this principle to build a pedometer in the year 1780. Maybe this was the first truly wearable body sensor. However, the invention of the pedometer is also attributed to other well-known historical figures like Thomas Jefferson or Robert Hooke. Presumably, an exact inventor cannot be named. [150]

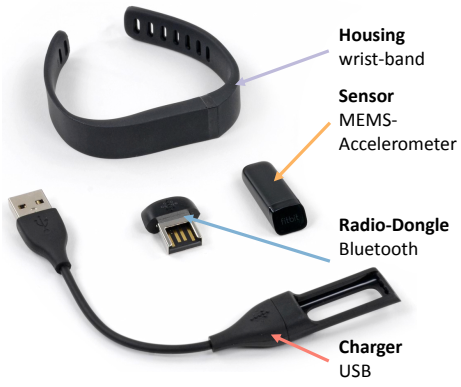
What is certainly known, is that in 1965, with the release of the *manpo-kei* mechanical pedometer (developed and produced by Yamasa Tokei Keiki Co., Ltd.; today YAMAX [260], Figure 2.6a) the idea of pedometry was reinvented as an instrument for public health use [238]. Back then, it was motivated by a scientific study exposing the number of 10 000 steps per day to be the ideal value for a healthy or active way of life. The name *manpo-kei* literally translates to *10,000 steps meter*. Still today, the guideline of walking 10,000 steps a day to promote a healthy way of living is prevailing [238].

This physical viewpoint, quantifying external work or behavior of a person, is only one aspect of body sensors. Also, the underlying physiological processes can be of

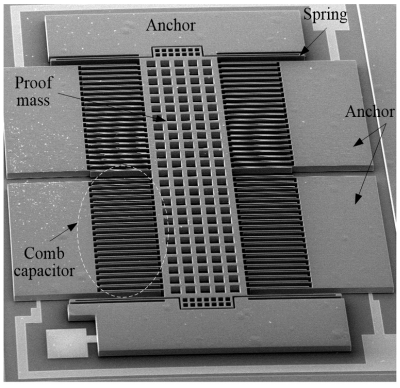
(a) Manpo-Kei (1965)



(b) Mechanical Pedometer (1987)



(c) FitBit Flex (2013)



(d) SEM MEMS accelerometer

Figure 2.6: Evolution of the Pedometer (step counting or motion tracking). The pocket-watch sized, mechanical *manpo-kei* was brought to marked in 1965 (a). Alongside an image of the insides of a successor device (*Digiwalker*) illustrating the mechanical working principle (b). In comparison, a micro-electromechanical system (MEMS) based, wrist-band sized electronic Pedometer (FitBit Flex) (c). Alongside a scanning electron microscopy (SEM) image of a MEMS-accelerometer design (d).

(a) Yoshida1338, CC BY-SA 4.0,
`https://commons.wikimedia.org/w/index.php?curid=65782173`
(b) Reprinted with permission, ©1987, Springer [27]
(c) Sam Lionheart, CC BY-NC-SA 3.0, cropped and annotations added to original,
`https://www.ifixit.com/Guide/Image/meta/dfmmiupZ2B3fWLnd`
(d) He et al., CC BY 4.0, [107]

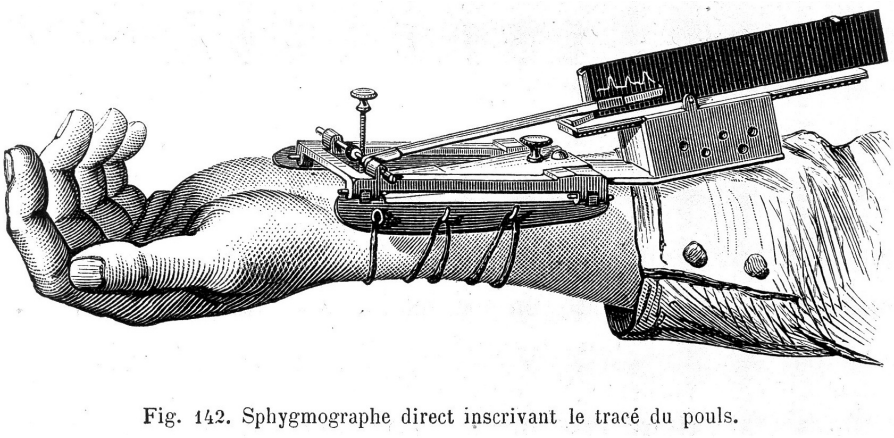


Fig. 142. Sphygmographe direct inscrivant le tracé du pouls.

Figure 2.7: Sphygmograph by Étienne-Jules Marey (1863). The mobile, mechanical device captured pulse waves from blood pressure changes and allowed for heart rate measurement Wellcome Collection, CC BY 4.0, `https://wellcomecollection.org/works/r89k5uqm`

direct interest. Besides body temperature, sweat rate, or respiration, heart rate is one of the most interesting vital signs.

Humans probably ever reasoned the influence of emotions and physical activity on the heartbeat intuitively and through observation. As with the pedometer, blood flow and heart rate have been scientifically studied several times in the past centuries, partly independently of each other.

A look at history reaches back to the ancient Greeks or Chinese, who described frequency and rhythm of the heart, the 13th century naming Ibn al-Nafis, who foretold the existence of capillaries and William Harvey, who explained and experimentally examined blood flow in 1518 [160, p. xvii]. The first documented use of simple mechanical devices to monitor heart rate (from blood pressure) dates back to the 16th or early 17th century [160, p. xvii]. More robust mechanical devices, like the *Sphygmograph* (Figure 2.7) were developed in the mid-19th century.

By the time the Sphygmograph was developed, cardiac activity was already verified as an electrical phenomenon. Carlo Matteucci experimentally showed the electrical activity of the heart muscle in 1842. Therefore, he used a frog's leg and demonstrated that its muscles would contract when connected to the heart nerve. Also, the first instruments to visualize such electrical phenomena (*galvanometers*) were existent, however not sensitive enough. Both areas of research only came together with the invention of Bernstein's *Differential Rheotome* in 1873. [45, pp. 22]

After this synergy of physiology and technology findings, the development of devices to record the electrical activity of the heart rapidly advanced. Starting from Lippmann's

invention of the capillary electrometer in 1873, over the first commercial string galvanometer presented in 1911 (Figure 2.8a), towards the first mobile electrocardiogram (ECG) monitor presented in 1949 by Holter (Figure 2.8b). [130]

Being able to measure external and internal processes of and on the human body influenced physical exercise experiments. For instance, in 1919, Magne [152] already investigated the relationship between steps and energy expenditure. Similarly, [68] examined a change in heart rate shortly after exercise in 1935. However, at that time, measuring steps, body acceleration, or heart rate was a difficult task. It was merely possible to obtained data during exercise and, if so, under laboratory conditions only. Developments of smaller, unobtrusive devices like the *manpo-kei* paved the path for new scientific experiments even for non-physicians. It was now more feasible to measure and research the effects of bodily functions outside the laboratory and without costly equipment under free-living conditions.

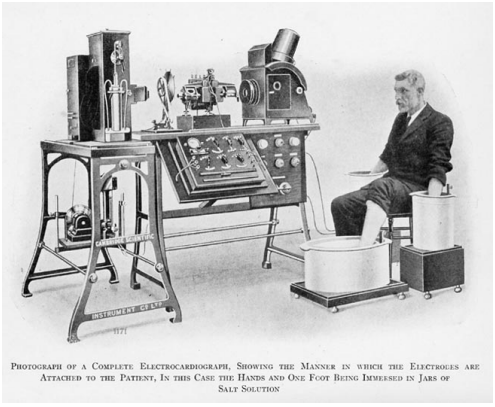
It takes another 20 years after the *manpo-kei* release until advantages in the area of micro-electronic development from the mid-'60s [114] impacts the design of WBS. In the year 1987, YAMAX, which initially put the *manpo-kei* onto the market, introduced its first electronic version of a pedometer. About the same time, in 1982, the *Sport Tester PE2000* was introduced to the market by Polar Electro [80]. It was the first consumer-grade wireless heart rate monitor (HRM), which followed a first fingertip device from 1977.

Another 10 years later, the first integrated acceleration sensor, based on MEMS technology (ADXL50), was brought to market by Analog Devices Inc. in the year 1992 [32, 216]. The device was originally designed as a sensor for automotive airbag systems. Finally, 13 years after the first fully integrated MEMS accelerometer, MEMS-technology based pedometers, nowadays named *fitness trackers*, successfully emerged on the market (Figure 2.6c). One example can be found with the *FitBit Flex* released in 2011 by, identically named, FitBit Incorporation. These developments mark the beginnings of WBSs.

Today, numerous WBSs are present. Their use goes beyond counting steps and accessing heart rate but is extended to classify activities, estimate energy expenditure, or track one's quality of sleep. Devices primarily designed for personal use are often called *fitness trackers* or *health trackers* (Quantified Self). Evermore, *smartphones* and *smartwatches* also emerge as digital-health tools for daily use [18]. An excellent example of this is the Apple Watch ECG-app[2], which allows to spot-check for cardiac arrhythmia. Moreover, it was successfully certified as a Class II device by the FDA at the end of 2018[3].

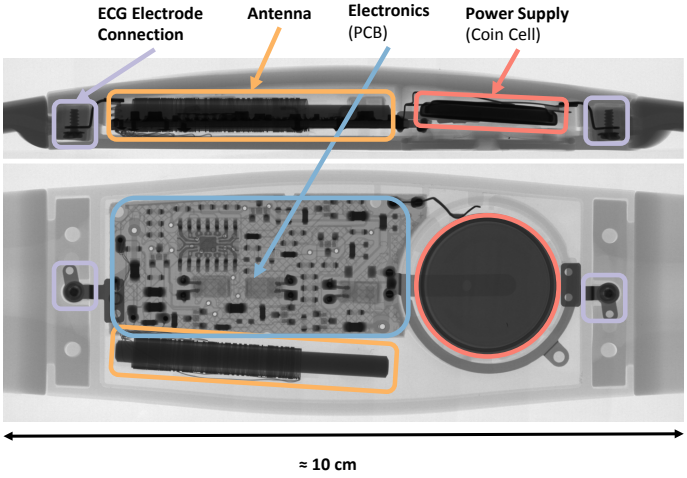
[2] Apple Inc., `https://www.apple.com/healthcare/apple-watch/`

[3] The FDA clearance says that a detection of atrial fibrillation is possible. Albeit, it is underlined that the device cannot replace medical consultation and clinical findings. The FDA Device Classification can be found under De Novo Number DEN180044, "Electrocardiograph Software For Over-The-Counter Use", `https://www.accessdata.fda.gov/cdrh_docs/pdf18/DEN180044.pdf`



(a) String Galvanometer (1911)

(b) Mobile ECG (1949)



(c) HRM radiograph (2010). Side- (top) and frontal-view (bottom)

Figure 2.8: Evolution of heart rate monitors. The first commercial device to measure ECG was based on Einthoven's String Galvanometer and presented in 1911 (a). A first mobile ECG was developed by Norman J. Holter in 1949 (b). It was the size of a back-pack and weighed about 40 kg. A miniaturized, consumer-grade electronic HRM was presented in 1982 by Polar Electro. It consisted of a chest strap transmitter (c), which recorded ECG. The heart rate was calculated online and transmitted to a wrist-watch-like receiver to display it to the wearer.
(a) Public Domain,
https://commons.wikimedia.org/wiki/File:Willem_Einthoven_ECG.jpg
(b) CardioNetworks 2012, CC BY-SA 3.0, https://commons.wikimedia.org/wiki/File:De-Norman_Holter_(CardioNetworks_ECGpedia).jpg
(c) SecretDisc 2010, CC BY-SA 3.0, annotations added to original, https://commons.wikimedia.org/wiki/File:X-ray_heart_rate_monitor_sensor_belt.jpg,

Besides personal use, such devices have also become accepted and validated research tools for application, e.g. in epidemiological or intervention research [76, 179]. Here, simple pedometers, accelerometers, or HRMs are used to support subjective measures. Moreover, they are also used as a cost-effective alternative to otherwise elaborate laboratory methods.

2.2.2 Paradigms and Ontology

For electronic devices that have the property of being worn on or close to the body, various terms were coined. Also, different paradigms exist that lead to conceptual partitions of (wearable sensors) systems, e.g. in respect to their application or used-case.

A frequently used term is *Wearables Computing*. This term is widely used in the scientific context and beyond. Devices belonging to this category are also briefly called *Wearables*. Following this vivid description, it covers all objects, devices, and helpers of everyday life that are worn on the body, such as glasses, watches, or even a bunch of keys. However, the term *Wearables* explicitly refers to the digitized (smart) variants of the respective predecessors, i.e. to smart-glasses or smart-watches. Some of these devices are already commercialized, e.g. smartwatches. The *Wearables* family also includes accessories extended by sensors such as bracelets, ear-studs, or belts up to complete garments (e.g. jackets, trousers), and even tattoos or sticking plaster [19] are considered. However, the latter is still a prospect of future wearable applications.

The connecting element of all *Wearables* is the personal relationship and the immediate proximity to the wearer's body. Indeed, they can be classified by their location, which is either inside, on, or next to the body. In this respect, *Wearables* share properties of garments. Indeed, parallel to the work on *Wearables*, research is also being carried out on garments, namely *Smart Cloth*, which integrates electronics. Both share common use-cases.

Wearables and *Smart Cloth* must be clearly distinguished from *Mobile Devices*. This is mainly because those devices are not worn but only held in hands during use. Such devices are typically referred to as *Handhelds*. Handhelds such as the smartphone or devices that are designed for mobile use (being portable), such as the laptop, are not part of the *Wearable* family. In the field of *Mobile Devices*, the technological perspective is addressed primarily. The term *Mobile Computing* [86] is also used as a substitute. Most often, *Mobile Computing* is about concrete technological concepts, especially concerning wireless communication.

The *Wearable Computing* paradigm, however, merely defines the topic of wireless communication and portability as a necessary design criterion. Essentially, it is about designing the devices in such a way that it is suitable for ubiquitous use and integrates seamlessly (unobtrusively or inconspicuously for the user) in their daily routine. In this manner, *Wearable Computing* has its roots in the paradigm of *Ubiquitous Computing* and other relatives, like *Pervasive Computing*, *Ambient Intelligence*, *Internet of Things*.

The term *Ubiquitous Computing* was first used by Weiser [251]. In his essay, he describes the idea of ubiquitous technology (computers), which integrates itself unobtrusively (i.e. in the ideal case completely transparently) into everyday life. He predicted that the ongoing miniaturization of digital circuits would lead to their integration into everyday objects, e.g. into a pen or a mirror, which would then increase their original usefulness. As a consequence, he proclaims that ubiquitous systems would replace classic computers in the 21st century. Laptops or similar devices, according to Weiser, are only precursors of this development. The same applies to modern technologies such as smartphones and tablet computers. He furthermore emphasizes the potentials and challenges of interlinking ubiquitous systems of his time.

Building on this aspect of *Ubiquitous Computing*, in the 1990s, the industry formulated a more pragmatic, primarily technology-driven variant of the *Ubiquitous Computing* paradigm under the keyword *Pervasive Computing* [100]. With reference to the available technologies of the time, the focus was initially on the conception of new business models for web-based and mobile applications but also the miniaturization of devices.

In contrast, research in the field of *Ambient Intelligence* focuses neither directly on hardware or software, but on social implications, i.e. the role of the participant interacting with technology [249]. It is about research and development of technological systems that respond to and support people. Ambient Intelligence is based on the technological concepts of Ubiquitous or Pervasive Computing.

With the field of home automation or *smart home*, Ambient Intelligence examples are found. Here, intensive research is carried out on scenarios and applications, e.g. concerning the design of furniture or other everyday objects, in such a way that they adapt to the user. Research is not limited to the digitalization of the environment, but also takes up ideas and developments from the field of *Wearable Computing* (e.g. integrate *wearables* into a *smart home*-environments).

At the turn of the millennium, a partial return to the scenario of intelligent and interlinked everyday objects, initially forecast by Weiser, took place. The keyword coined in 1999 by Ashton [25] is “Internet of Things”. Against the background of the massive increase in data and computer capacity, Ashton notes that the full potential of the Internet is not yet exhausted. This would only be achieved by closing the information gap between the real and virtual worlds. This idea of having information-processing systems globally interlinked (i.e. *Internet of Things*) is the consequent continuation of the development of wireless sensor systems, including *Wearables*.

2.2.3 Technological Aspects

In the following, technological aspects of body sensors are outlined. These comprise the use of embedded systems (sec. 2.2.3.1), wireless communication (sec. 2.2.3.2), and the application of sensors or electrodes (sec. 2.2.3.3). This section closes with general remarks on the architecture of wireless sensor nodes (sec. 2.2.3.4) and more specific aspects regarding wireless body sensors (sec. 2.2.4).

2.2.3.1 Embedded Systems

First body sensors were mechanical or, if based on electronic components, stationary, i.e. non-portable apparatus. In comparison, if WBSs are referred to today, it is about miniaturized electronic devices. In this regard, they are understood as an embedded system (ES). More generally, ESs make up the information processing part of a WBS.

Commonly, ESs are defined as electronic computing devices build for a specific application. This is in contrast to general-purpose computing, e.g. personal computers, which are used to virtually run *any* kind of software application. The development of ES thereby includes both hardware and software components of the system. Also, input (sensors) and output (actors) components are considered as part of ESs. Multiple ESs that are connected to each other and measure or manipulate their physical environment are also known as cyber-physical systems (CPSs). All CPSs are ESs, but the term is used to highlight the connection of an ES and its physical environment.

Today, ESs are found in everyday objects like household devices (e.g. washing-machines), consumer-electronics, in factory automation, cars, robots, and many more devices. While in general computing, performance is the often dominating requirement, with ESs other (conflicting) design criteria can be identified. Most of them are non-functional, like size, weight, power consumption (run-time or *autonomy*), or usability [156, p. 31]. Usually, these requirements can be met given that the use-case of an ES is defined before its implementation. Besides, a co-design of software and hardware, and an (at least basic) understanding of the application's domain is helpful in this respect.

2.2.3.2 Wireless Communication

Any WBS needs a wireless communication interface in order to interchange signals with other devices or with the wearer. In the context of ESs, mainly radio waves in the frequency range between 30 Hz–3 GHz are used.

The use of particular frequency-bands (e.g. the maximal power emitted) is regulated by the International Telecommunication Union (ITU). Free and globally usable frequency-bands are collected as industrial, scientific, and medical radio bands (ISMs). The most used frequency-bands are in the Sub-GHz band (433 MHz, 868 MHz, 915 MHz), the 2.4 GHz-, and the 5 GHz-band. Different technical implementations are available, which can be categorized by the criteria range, bandwidth, and power consumption.

In terms of range, the groups of near-field communication (NFC), wireless body-area (WBAN), personal-area networks (WPAN), and furthermore local- (WLAN), and wide-area networks (WWAN) are distinguished. The range of these reaches from $<$1 m (NFC), over 1 m to 10 m (WBAN and WPAN), up to about 100 m (WLAN) or even 100 km (WWAN).

Regarding data throughput (bandwidth) and power consumption, the different frequency-bands and the range need to be considered: Firstly, that is because, according

to the Shannon–Hartley theorem[4], the bandwidth is proportional to the frequency. Secondly, the power consumption in wireless communication is a result of range and data throughput (bandwidth). In general, lowering both will result in low power consumption and vice-versa. Theoretical bounds for this rule of thumb are found in the Inverse-square law[5].

Wireless-communications protocols for WPANs are standardized by the IEEE 802.15 working group. These focus but are not limited to radio waves. Besides, light-based communication (e.g. Infra-red) is part of the WPAN-standard. Also, in contrast to narrow-band protocols, ultra-wideband (UWB)-technology is used alternatively in WPANs. Due to the very-wide frequency spectrum used (3 GHz to 10 GHz), the mean signal power is low, while in turn, the range is limited.

Most wide-spread wireless-communications protocols for WBSs operate in the 2.4 GHz-band. They are not limited to the standards of the IEEE 802.15 but include other standards as well. Technologies typically found are, for example, Bluetooth Low Energy (BLE), ZigBee, Z-Wave, ANT, and EnOcean [171].

2.2.3.3 Sensors and Electrodes

Sensors and electrodes connect ESs and the surrounding environment. Therefore, a physical effect (which is to be measured) needs to be translated or transduced into an electrical quantity. In fact, this is already a possible definition of the term sensor. In this respect, also electrodes can be understood as sensors. This is although no *translation* takes place since the physical effect is already an electrical quantity.

A variety of stimuli can be of interest in a sensor system, including but not limited to mechanical, thermal, biological, chemical, or optical stimuli. Likewise, a multitude of effects exist, that might be utilized to construct a sensor and translate the stimuli into either a voltage, current a change of resistance or capacitance. These effects are, for example, thermoelectric, photoelectric, or piezoelectric effects.

From the engineering perspective, an ideal sensor is expected to be sensitive to the measured quantity only, is not affecting the measured property itself, and to have no measurement uncertainty. In reality, however, this is not achievable.

Regarding measurement uncertainty, a sensor's measurement is always affected by random errors, e.g. noise and systematic errors. Random errors are also described as *precision*, while systematic errors are typically denoted as the sensor's *accuracy* or *trueness*. In addition to these limitations, the sensor's measurement resolution might be bounded (e.g. in the upper range due to a saturation effect), or the repeatability cannot be guaranteed (e.g. because of a hysteresis effect).

[4] The Shannon–Hartley theorem states, that the channel capacity $C/\mathrm{bit/s}$ (data throughput) depends on the bandwidth B/Hz, and the signal-to-noise ratio. Thus, assuming a fixed noise level N/W, with increasing the signal power S/W the channel capacity is increased as well: $C = B \cdot \log_2\left(1 + \frac{S}{N}\right)$

[5] The Inverse-square law states, that for any physical quantity its intensity is proportional to the inverse of the squared distance between transmitter and receiver: intensity $\propto \frac{1}{\text{distance}^2}$

2.2.3.4 Wireless Sensor Nodes and Networks

The topics of wireless communication, sensors, and embedded systems come together in the research area on wireless sensor nodes. Wireless sensor nodes form a subset of ESs that, per definition, include one or more sensors and a wireless communication interface. They are of particular interest in a variety of application scenarios in the industry (e.g. predictive maintenance, structural health monitoring or smart grids), logistics (e.g. asset tracking), medicine (e.g. e-health, fitness) or private living (e.g. smart home, gaming, and entertainment).

Initial sketches for the architecture of sensor nodes and their application scenarios were described in 1993 by Wise [257]. A few years later, Mason et al. [158] introduced a first functional wireless sensor node (WSN) for monitoring environmental parameters (temperature, air pressure, humidity, vibration). This prototype had an integration density of 25 cm^3 and an operating time of up to 330 days. Following on from this, in 2001, Warneke et al. [248], sketched the vision of "intelligent dust". It outlines the idea of ubiquitous and miniaturized WSN for measuring environmental parameters such as temperature, humidity, or radiation in the size of dust particles (1 mm^3). As an important result of their work, they specify the key requirements for WSNs (or smart dust) still existing today: miniaturization and energy efficiency.

It is the ubiquitous and autonomous character demanded by WSNs, which dictates the requirements for wireless communication and miniaturization of all components. Only these fundamental criteria enable unrestricted and free placement in the room (portability) and guarantee a maximum of operating times (trough energy efficiency) and thus the highest possible autonomy (low maintenance intensity). In extreme cases, WSNs are entirely based on energy harvesting methods and are thus enabled to sense their environment autonomously.

In general, the architecture of WSNs includes a microcontroller unit (MCU), a wireless interface, an energy store, and one or more sensor elements (Figure 2.9) [261]. Optionally, memory can be added to the architecture.

The design space and optimization criteria of sensor nodes comprise "space requirements" (miniaturization), "operating time" (autonomy), and "cost efficiency" [198]. As discussed by Romer et al. [198], these criteria affect each other. Optimizing a particular dimension often contradicts other functional or non-functional requirements. For instance, it could be aimed towards making a WSN smaller. However, optimizing for space might demand more expensive components. Also, this limits the available space for energy storage. Other requirements mentioned by Romer et al. [198] cover the communication modalities. These include topology, coverage, and network size, among other factors. Similar requirements are mentioned by Yang [261]. They also mention, among other factors, "robustness" and "security" requirements.

Current application scenarios, which are associated with the headlines *Digitization*, internet of things (IoT), or *Advanced Manufacturing* (also known as *Industry 4.0*), pick up the *Smart Dust*-vision utilizing WSN. In the concrete scenarios, sensor nodes

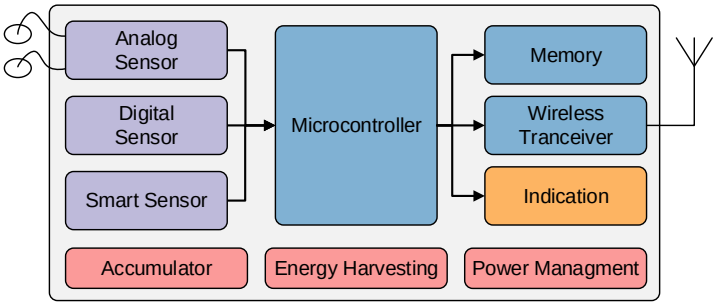


Figure 2.9: An exemplary architecture of a wireless sensor node. It includes 3 sensors (analog and digital), a microcontroller unit, a wireless interface, memory, power management, energy storage, and energy harvesting solution.

(individually or in a network) store or transmit measured values and thus provide real-time and long-term data that were previously not available.

Often, however, not the recorded raw signals themselves are of interest but derived information, which can be obtained from the aggregation of the various heterogeneous sensor signals. Examples include sensor networks that are predicting environmental disasters (probabilities of forest fires, storms, or tidal waves from) or sensor nodes that are used for monitoring transported goods (estimating food quality). Other scenarios describe, for example, the energy optimization of private households (context-specific switching on or off of domestic appliances). An overview of current developments is given by Rawat et al. [190].

Many modern sensors available at the market today can already be understood as *smart sensors*. These smart sensors not only record signals but also process them. For instance, modern inertial measurement units free the system designer from the need to implement fusion or other evaluation algorithms. Instead, these advanced signal processing steps are already an integral part of the sensor itself. An example is knocking detection in acceleration sensors, which can be found in mobile phones to switch on the device when touched.

Besides, a smart sensor can even be fully, freely programmable. In this respect, the sensor itself could already be considered an ES or system-on-chip (SoC). In this case, it is furthermore possible to avoid adding additional MCU to the system's architecture of a sensor node (Figure 2.9). Instead, the smart sensor is connected directly to the memory or communications component.

2.2.4 Wireless Body Sensors

Wireless body sensors are a sub-group of WSNs. They are distinguished by the fact that they are worn directly on the wearer's body or skin. In addition, devices exist

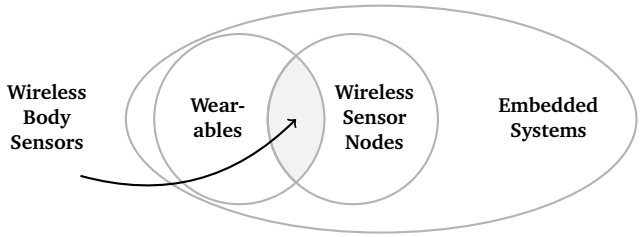


Figure 2.10: Relation of WBSs to the domains of Wearables, WSNs, and ESs.

that are implanted. The International Electrotechnical Commission (IEC) strategic group 10 defined dividing lines of WBS, which can either be *near-body* (no contact to the organism), *on-body* (contact to the organism), or *in-body* (implanted). They furthermore identify the separate category of *electronic textiles* and summarize all with the umbrella term *Wearable Smart Device*.

WBS can also be found as part of a network, called wireless body sensor network (WBSN) or wireless body area network (WBAN) (or in short BAN or BSN without the prefix *wireless*). In this respect, WSN can be categorized by the degree of distribution. Thus, a WBS can either be a stand-alone device (e.g. smartwatch), used combined in a closed system (e.g. as an additional sensor in a smart textile), or be part of a fully distributed system, interlinked with other external devices (e.g. furniture in a smart home). In contrast to a wearable computing device, which could be a smartwatch with a display only, a WBS necessarily embeds one or more sensors, which capture data from its wearer.

Regardless of being near, on, or in the body, aspired use of a WBS is to allow for continuous monitoring of vital, kinematic, and environmental parameters. The domain of WBS is strongly influenced by the ideas of *Wearable*, *Pervasive*, and *Ubiquitous Computing*. At the same time, it is technologically grounded on the developments found in the areas of *Mobile Computing* and IoT. Thus, WBSs can be seen as specific ESs, making up the cut set between *wearables* and WSNs (Figure 2.10). Their purpose is to 1. measure, 2. evaluate, and 3. provide physiological, kinematic, or environmental data to (and from) the person (wearer) using the WBS.

Due to the miniaturized design, WBSs offer advantages compared to stationary solutions. What makes WBSs such an interesting technology is their unobtrusiveness and the fact that they can be used for objective long-term measurements. This opens up the way towards new findings in medical diagnostics, public health, or sports science. A list of sensors used in WBSs includes:

- Acceleration, angular velocity or magnetic field sensors
- electrode sensors
- force sensors

- temperature sensors
- humidity sensors

A list of possible raw parameters that could be accessible by WBSs includes:

- Motion, altitude, force of impact
- Electrical activity of the heart or other muscles
- Thoracic movement
- Body and skin temperature
- Sweat level or rate

Based on these raw parameters, extended measures can be calculated, for instance:

- Step count (and cadence), stride length, posture
- Heart rate, its variability, and recovery
- Respiration rate and breathing volume
- Risk of heat stroke
- Dehydration

Some of these measures are derived in multiple-steps. For instance, the electrical activity of the heart is used to calculate heart rate, which in turn is used to calculate its variability. Finally, multiple raw or derived measures can be combined in order to calculate more abstract entities like:

- Absolute orientation of the body
- Running speed and or distance traveled
- Physical activity or energy expenditure
- Cognitive workload or fatigue

Multiple terms exist to address consumer-grade WBSs [184], e.g. *activity tracker* or *fitness tracker*. There is, however, no sharp separation between these terms. In the simplest case, using one of these terms could describe a pedometer or HRM (sec. 2.2.1). In this understanding, an *activity* or *fitness tracker* is a WBS, embedding an accelerometer used to count steps or, embedding electrodes used to calculate heart rate. Also, a more sophisticated device could be covered by these terms, e.g. a WBS capturing both motion and heart rate in parallel.

2.3 Wearable Algorithms

The final step in body sensor applications is the analysis of the (physiological) data recorded (Figure 2.11). The goal thereof is to extract information out of the raw data. This general concept of WBSs computing is also referred to as *wearable algorithms* [205, p. 353][52]. In the following, an overview of bio-signal processing in general (sec. 2.3.1-2.3.2), and selected machine learning methods (sec. 2.3.3) to extract information from the latter, in particular, is given.

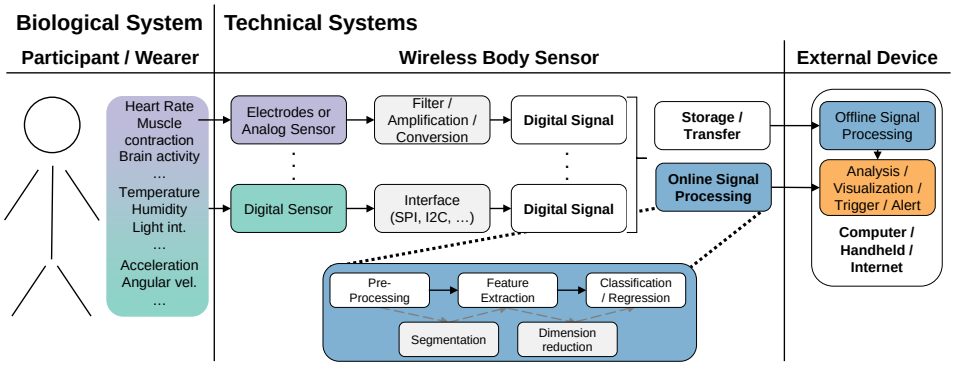


Figure 2.11: Conceptual flow of data in WBS applications. Data from the biological system (wearer) enters the technical system (WBS) through sensors (analog or digital). After digitization, data is either transferred to an external system or processed locally. The data processing typically covers, pre-processing, feature extraction, and classification or regression steps (not all steps are needed, and other steps are possible as well, e.g. segmentation or dimension reduction). The data is finally used for further analysis, visualization, or to trigger other mechanisms (e.g. alert the user). Based on [205, p. 354]

2.3.1 Digital signal processing

In general, digital signal processing is about processing, storing, and transmitting digital data. Here data is referred to as a single value or symbol. In this context, a signal is a series of data indexed in time, that is a time series. Information is the result of applying specific algorithms to a signal or multiple signals. However, information might serve as new data for another algorithm in order to extract other information as well. An example is the body temperature (data), which is used to detect fever (information), which in turn could be used to calculate a health index (also information). A discussion of these definitions can be found in [200]. Processing can be executed *online*, i.e. locally on the WBS, or *offline* on an external system, to which the raw data was transmitted. Nevertheless, the processing steps in both cases are similar.

Firstly, a physical effect has to be chosen that allows observing the quantity of interest. The effect is considered an analog signal. This means it is a continuous time-varying quantity. This analog signal (or effect) is then measured using a sensor, which translates it into an electrical quantity (e.g. current, charge, voltage, or resistance). In case, the physical effect is already an electrical signal, instead of a sensor, electrodes are used. Typically, the intended final quantity is a voltage.

For further processing, the voltage (or other electrical quantity) often needs to be amplified in order to match the voltage domain of the digital system (e.g. 1.8 V to 5 V). Also, filters are used to remove noise or unwanted frequencies. Furthermore, a sensor could also comprise calibration or justification elements, e.g. to compensate

for temperature drifts. Amplifiers, filters, and other subsidiary components can come as part of a sensor or as independent components. The pre-processed signal is then digitized by means of an analog-digital converter and ready for further digital signal processing (DSP) operations. This digitization typically comprises two steps, namely discretization and quantization.

Discretization is the process of sampling a continuous signal over time. Typically, single instantaneous values of a signal are measured (sampled) at equidistant time intervals. The result is a time-discrete signal. This process furthermore limits the bandwidth of the signal, because information on signal changes, i.e. frequencies higher than the sampling rate, are not preserved. This relationship is expressed in the Whittaker-Kotelnikow-Shannon or Nyquist-Shannon sampling theorem. It states that the minimum sample frequency (f_s) must be at least twice as large as the maximal frequency (f_{max}) within the signal under observation (given equidistant samples, (eq. 2.6)). In real technical systems, however, a higher sampling rate is often chosen (oversampling), and additional mechanisms are introduced to remove unwanted signal components at an early stage (alias filter).

The quantization of a signal limits its resolution in the value range. Limitation depends on the chosen maximum value range and the quantization level. Both determine the smallest value that can be represented by the least significant bit (LSB) (also known as quantum). Since all values are mapped to multiples of the LSB, a rounding error in the order of 1 LSB results (±0.5 LSB in a first approximation). This is also called quantization-noise or -error (eq. 2.7). In binary digital systems, a doubling of the resolution (which corresponds to a reduction of the quantization noise by half) can be achieved by increasing the resolution by 1 bit. In real technical realizations, the quantization-error, e.g. due to noise, is often higher than 1 LSB.

$$f_{max} = \frac{1}{2} \cdot f_s \qquad (2.6)$$

$$Q = \frac{q}{2^b} \qquad (2.7)$$

f_s - sample frequency / Hz
f_m - maximal frequency / Hz

Q - least significant bit / -
q - full scale range / -
b - number of bits / bit

2.3.2 Bio-signal processing

The ECG serves as a good example and starting-point for bio-signal processing because of its central importance and manifold applications in medicine, psychology, and sports. Heart rate and the electrical activity of the heart have already been mentioned briefly (sec. 2.1 and sec. 2.2.1). In medicine, the ECG is used to detect pathological changes in the heart muscle. In psychology, the variability between individual heartbeats is of interest as a physio-psychological parameter, which allows concluding the activation of the ANS. Similarly, training overload is studied in sport science, and of course, the relationship between performance and heart rate is studied in detail.

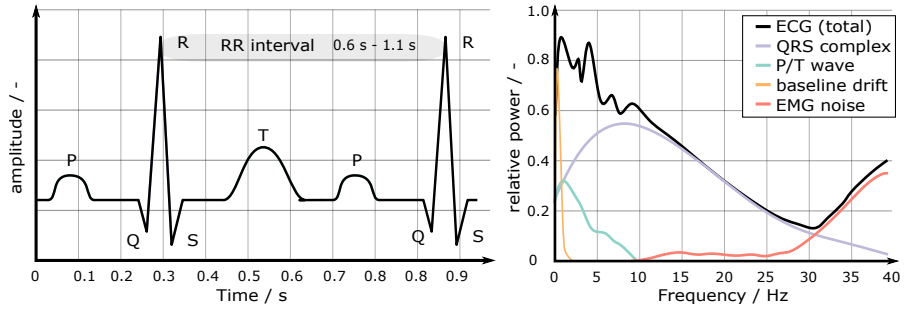


Figure 2.12: Normal morphology of the ECG reflecting the de- and re-polarization of the different parts of the heart muscle (left). Alongside, a qualitative example of the ECG frequency spectrum along with selected, isolated components (right). ECG is based on Hank van Helvete, CC BY-SA 2.0, `https://upload.wikimedia.org/wikipedia/commons/0/00/EKG_Komplex.svg`. Spectrum is adopted from [81]; original [235]

The ECG depicts the (summarized) electrical activity of the heart muscle. Therefore, electrodes are placed on the participant's chest. Between those electrodes, electrical changes due to heart muscle contraction become visible. These are reflecting the de-polarizing and re-polarization of different regions in the heart, namely the atrium and ventricle muscles (left and right). Although the pause in between to contractions may change, the transition of depolarization and thus the order of contractions of the heart muscle takes place according to the ever same scheme.

The prominent waves and peaks are labeled with the letters P, Q, R, S, T in sequence [130] (Figure 2.12). The origin of depolarization is the sinoatrial or sinus node (in the right atrium). The depolarization spreads, passing the atrial muscles going into the atrioventricular node. This leads to contraction of the atria and can be seen as the P-wave in the ECG. Starting at the atrioventricular node, the depolarization spreads further into the bundle of His along the ventricles into the Purkinje fibers. The ventricles contract and the characteristic part of the ECG waveform (QRS)-complex becomes visible. The re-polarization of the chamber muscles, in turn, becomes visible as the T-wave in the ECG.

Due to the different waves, the ECG consists of different frequencies. QRS-complex contributes higher (10 Hz to 17 Hz), while P- and T-wave (0.1 Hz to 7 Hz) contribute lower frequencies to the complete spectrum. Heart rate emerges in the range of 0.67 Hz to 3.67 Hz (given heart rate limits of 40 bpm to 220 bpm). [242]

What makes the processing of the ECG challenging is that a multitude of signals is overlapping the ECG spectrum. For example, other muscles in the human body produce electromyography (EMG)-noise (5 Hz to 50 Hz), respiration affects the ECG baseline (0.12 Hz to 0.5 Hz). Also, power line frequency (50 Hz or 60 Hz) interferes. Besides, motion artifacts introduce high-frequency noise.

Changes in the morphology of the ECG indicate pathological changes in the heart muscle[6]. A distinction is made between unipolar configurations of ECG-electrodes. In such the voltage of each individual electrode is measured against a common reference electrode. Each electrode is referred to as a lead. Alternatively, electrical changes can be measured directly between two electrodes. This is referred to as a bipolar lead. In textbooks, most often, the so-called Einthoven II lead is depicted. In this configuration, electrodes are placed between the right hand and left foot. In mobile systems, which only aim to detect heart rate, electrodes are often placed on the left and right side of the chest. This is also a bipolar lead. It is similar to the Einthoven I lead, in which electrodes are placed on the left and right hand.

To obtain heart rate or heart rate variability from electrocardiogram (ECG), the so-called QRS-complex must be detected within the ECG-signal. The mean distance between successive QRS-complexes, also known as RR-distance, usually averaged over 5 to 10 complexes, is the heart rate (HR). The calculated variance of these distances, in turn, is the heart rate variability (HRV)[7]. The de facto standard for QRS-complex detection in WSN is the algorithm proposed by Pan et al. [180] in 1985. A revised open-source implementation was published by *Patrick S. Hamilton*[8].

The algorithm consists of typical DSP operations (Figure 2.13). At first, unwanted high- and low-frequency components are filtered. Next, a differentiator is used to emphasize the sign changes of the QRS-complex (compared to T- and P-waves). These changes are then smoothed by applying a windowed moving-average filter. The resulting signal is a summary of the QRS-complex, which can now be detected by a simple peak detection. As an additional step, different detection rules (thresholds) are applied in order to mark physiologically reasonable peaks only. For instance, these rules take into account the minimal (physiological) de- and repolarization times of the heart muscles. Concerning this example, all peaks 200 ms before and after a larger peak are ignored.

2.3.3 Machine learning

WBSs provide physiological and bio-kinematic data, including heart rate or acceleration. Using these raw data combined with a suitable model allows obtaining abstract information, e.g. the physical activity or the running speed of a wearer. For this purpose, at first, such a model has to be found. This model is intended to provide a mapping between the sensory data and a given target value. Machine learning (ML) methods have proven their effectiveness in building such models [97, 189, 205].

[6] Analyzing the morphology of the ECG, however, requires a standardized positioning of the electrodes since the observable morphology depends on the relative position of the ECG electrodes to the heart. Also, in order to allow conclusions about the angel of the heart, at least 3 leads are necessary.

[7] Variance is only one metric to present heart rate variability (HRV). This topic will be addressed in chapter 4

[8] Open Source ECG Analysis Software, 2002, licensed under GNU Library General Public License `http://www.eplimited.com/confirmation.htm`

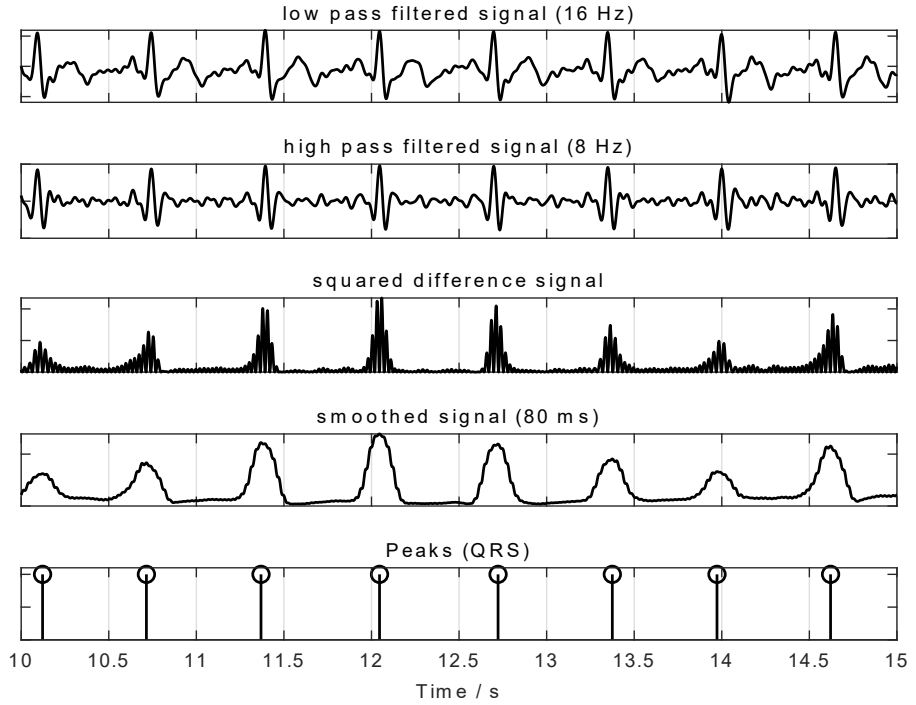


Figure 2.13: Visualization of the different signal processing steps of the Pan-Tompkins [180] algorithm for QRS-detection. A low- and high-pass filter form the first two steps of the algorithm before the signal is squared and smoothed by means of a moving average filter. Finally, the peaks of the smoothed signal are identified by an adaptive threshold mechanism.

In the classic ML scenario of sensor data processing, sensors are used to sample raw data on the basis of which *features* are extracted. A feature could be the mean value or variance of the raw data. In this way, from the measuring data (or time series), one or more features are derived. An example is the ECG, from which the mean heart rate could be calculated or acceleration data from which peaks could be counted. Often various pre-processing steps are applied to the raw data beforehand.

A prediction (classification or regression) is then made based on these features. This procedure is known as supervised learning. It requires that for all input patterns (features) corresponding output patterns (target values or ground truth) are known. In this way, based on the provided random sample of features, a generalized model or mapping M is to be found (eq. 2.8) [205].

In this respect, the idea of using ML techniques is to come up with a model that can be later used for inference based on new unseen data (eq. 2.9). Therefore, a function

m that maps the features vector $\mathbf{x^i}$ (independent variable/s) to a corresponding target value y (depended variable) is needed. Here, each feature vector $\mathbf{x^i}$ is drawn from the input space $\mathbf{X}$, while each target y_i belongs to the output space $\mathbf{y}$.

In practical application, such a function often cannot be found, e.g. because of noisy data. The output of the function m (prediction) is therefore affected by certain deviations ϵ. Hence, the output of a ML model m is $\hat{y} = y + \epsilon$. [11, 174]

$$M : \mathbf{X} \rightarrow \mathbf{y} \; : \; \mathbf{X} \in \mathbb{R}^{o \times n}, \mathbf{y} \in \mathbb{R}^n \tag{2.8}$$

$$m : \mathbf{x^i} \rightarrow \hat{y}_i \; : \; \mathbf{x^i} \in \mathbf{X}, \hat{y}_i \in \mathbf{y} + \epsilon \tag{2.9}$$

M	-	ML task	$\mathbf{y}$	-	target vector	$\mathbf{X}$	-	feature matrix
m	-	ML model	y_i	-	target scalar	$\mathbf{x^i}$	-	feature vector
ϵ	-	error or noise	$\hat{y}_i$	-	prediction scalar			

2.3.3.1 Supervised Classification and Regression

Supervised ML-tasks can be divided into regression and classification problems. In regression, continuous target values (interval scaled) are to be mapped by a function. The desired target function could, for example, be a simple linear function. This linear function should now accurately represent the exact shape of the target values. It could, for example, be chosen such that the mean absolute distance to the data points is minimized (the linear function is placed in the center of the data points).

Indeed, the deviation or error of a regression is one aspect of evaluating its quality. Therefore, different metrics, e.g. mean absolute error (MAE), median absolute deviation (MAD), mean squared error (MSE), root mean squared error (RMSE), coefficient of determination (R^2), or mean absolute percentage error (MAPE) can be used (eq. 2.10 - 2.15).

$$\mathrm{MAE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^{n} (|\hat{y}_i - y_i|) \tag{2.10}$$

$$\mathrm{MSE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{2.11}$$

$$\mathrm{MAPE}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{|\hat{y}_i - y_i|}{|y_i|} \right) \tag{2.12}$$

$$\mathrm{MAD}(\mathbf{y}, \hat{\mathbf{y}}) = \mathrm{median}(|\hat{\mathbf{y}} - \mathbf{y}|) \tag{2.13}$$

$$\mathrm{RMSE}(\mathbf{y}, \hat{\mathbf{y}}) = \sqrt{\mathrm{MSE}(\mathbf{y}, \hat{\mathbf{y}})} \tag{2.14}$$

$$\mathrm{R}^2(\mathbf{y}, \hat{\mathbf{y}}) = 1 - \frac{\mathrm{MSE}(\mathbf{y}, \hat{\mathbf{y}})}{\mathrm{variance}(\mathbf{y})} \tag{2.15}$$

In contrast to regression, in classification problems, a function is to be found that separates discrete target values (nominal or ordinal scaled) from each other. The target function could again be a simple linear function. However, now the function is chosen in such a way that it separates certain classes from each other. Therefore, it is also called the *decision boundary*.

Often, the linear function is chosen in such a way that the mean (squared) distance to the respective classes is minimal (the linear function would then be centered between the classes). However, in classification, the target values, also known as *classes*, are usually categorical (not metric) and therefore do not necessarily have to keep a certain distance (interval) or have an order at all. Therefore, different loss functions and measures of quality are applied compared to regression problems.

With regard to classification problems, a common error metric is the *accuracy* (or *hit rate*, eq. 2.16). It is defined as the percentage of matches between the true and the predicted class label. Those matches are referred to as true positive (TP) ($\hat{y}_i = a \; \forall \; y_i = a$) or true negative (TN) ($\hat{y}_i \neq a \; \forall \; y_i = a$), while mismatches are referred to as false positive (FP) ($\hat{y}_i = a \; \forall \; y_i = b$) or false negative (FN) ($\hat{y}_i = b \; \forall \; y_i = a$). Given TP, TN, FP, and FN, the error type can be specified in more detail. For instance, *sensitivity* (true-positive rate (TPR), also known as *recall*, eq. 2.17), *specificity* (true-negative rate (TNR), eq. 2.18) or precision (positive-predictive value (PPV), eq. 2.19) can be calculated.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (2.16) \qquad \text{Specificity} = \frac{TN}{TN + FP} \quad (2.18)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2.17) \qquad \text{Precision} = \frac{TP}{TP + FP} \quad (2.19)$$

Moreover, classification and regression problems can be converted into each other. Given a regression problem, the input values can be transformed into (discrete) class labels by rounding floats to integers. Likewise, class labels can be modeled by regression techniques, e.g. by rounding the model's output to the closest integer value (class label). The latter, however, is done under the assumption that the class labels are interval or ordinal scaled. Using regression techniques for classification on nominal values is likely to produce biased results as there is no natural continuous relationship among different labels.

Finally, it should be noted that, based on a specific ML algorithm, it may be necessary to pre-process the input features. The rationale for this is that the input data is sometimes scaled unevenly, or their distributions differ. Two conventional approaches for normalization or standardization are min-max normalization (eq. 2.20, scaling to the value range between 0 and 1) or the z-standardization (eq. 2.21, fix a mean value of 0 and a standard deviation of 1).

$$X_n = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (2.20) \qquad X_z = \frac{X - \mu(X)}{\sigma(X)} \quad (2.21)$$

μ - mean value μ - standard deviation

In the former, a direct representation of a decision boundary or regression function was described. These approaches are known as discriminative models. In contrast,

the probability distribution could be modeled as well. Such approaches are known as generative models.

2.3.3.2 Model Selection: Bias and Variance

Usually, ML algorithms are used to construct models, which can then be used for inference on new unseen data. The objective is that the model has no, or only a very small error (i.e. high accuracy). This is not limited to the training of the model but should apply to new unseen data in particular. Here, the goal is to find a generalization that fits the entire statistical population using the test samples only.

However, ML algorithms and models differ widely in terms of applicability, flexibility, and complexity. In general, there is a link between the model's complexity and its ability to store information. This is theoretically expressed in the theory of *VC-Dimension* or *entropic capacity*. In simplified terms, the model's complexity reflects the variance that it can *store*.

The trade-off in limiting or extending a model's complexity is known as the bias-variance dilemma. It explains the fact that simplistic models (less complex, e.g. linear regression) tend not to describe the data appropriately. In this manner, not all inter-relationships and dynamics among different features can be mapped. This is known as high bias or under-fitting. On the other hand, more complex models tend to be sensitive to data fluctuation. Hence, they are mapping noise rather than meaningful relations. Such models are said to have high variance or to be over-fitted.

The result of fitting a model with insufficient or overmuch degree of freedom is a less adequate mapping of the real underlying relations in the data. In general, a so-called *triple trade-off* exists between the complexity of a model, the amount of training data, and the model's ability to generalize to new unseen data ([11], Figure 2.14).

To avoid over-fitting, the data sample can be split into a training-, a validation-, and a test-set. The straight forward approach is to use a *hold-out validation*. Therefore, a fixed partition of data for each set, e.g. 60 % for training, 20 % for validation, and 20 % for testing. Training and validation data are used while fitting the model. As a consequence, the hyperparameters are tuned to match the samples from both sets. By this means, the effect of over-fitting is regularized. The test set is finally used to estimate the generalization error (off-training-set error).

To further limit any generalization error, cross-validation (CV) strategies can be used. In this way, multiple partitions of the training and validation sets are created. E.g., with a k-fold strategy, k partitions are randomly drawn from the data set (without test set). During the hyperparameter optimization, k-different models are trained. Finally, the hyperparameter set is chosen, which is minimizing the mean deviation among all splits. Often, class labels are not equally distributed in the data set. Thus stratified folds should be used to avoid possible over-fitting for a particular class.

Another CV strategy is the leave-one-group-out (LOGO) partition. With LOGO, data from one specific group, e.g. different measuring days or machines are validated against

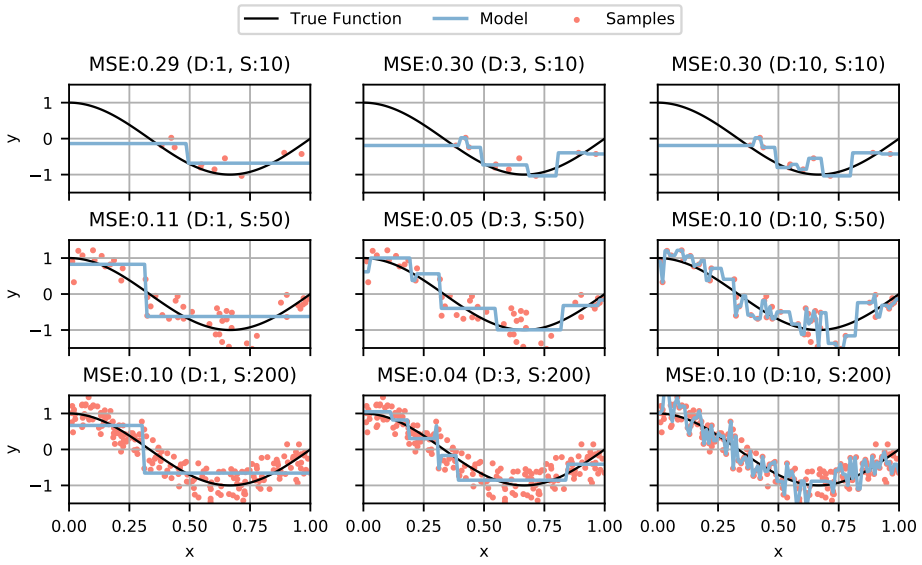


Figure 2.14: Example of the bias-variance dilemma. Decision trees (DTs) models are trained on random samples of a sinusoidal signal superimposed by random noise. The maximum number of splits of the DT is limited to 1, 3, or 10 splits. The number of samples provided for training is limited to 10, 50, or 200 samples. The lowest MSE is found when a maximum of training samples is provided, and the model's complexity is tuned just right.

each other. This is particularly important if the data set contains data from different persons. To compensate for inter-personal differences (e.g. age, weight), group-wise validation is needed. This special case is also known as leave-one-subject-out (LOSO).

Also, combinations of validation schemes can be used, e.g. a k-fold-group-validation. Still, validation is only an instrument to estimate generalizability on the sample under consideration. There is no insight on possible out-of-test errors, which result from violating the assumption of having identically and independently distributed training and test data.

2.3.3.3 Algorithms

Several ML algorithms exist to deal with linear and non-linear problems. They can be distinguished from each other by their properties of being parametric or non-parametric or having a generative or discriminative (conditional) character.

Regarding parametric models, the degree of freedom is fixed. An example is the degree of a polynomial function that could be used to fit a regression problem. In

contrast, with non-parametric models, no such limitation exists. Hence, no prior assumption on the data is needed. These algorithms, however, require larger amounts of training data, and over-fitting is more likely because no unique solution exists (in theory, an infinite number of parameters has to be determined).

Generative and discriminative models differ in their *perspective* on the feature space. While a generative classifier searches for a decision boundary separating distinct classes, a generative approach models the feature space's distribution. As a practical consequence, with generative models, not only class label can be determined but also a probability for the membership to a particular class.

In fact, no single method or algorithm can be considered optimal or superior against another[9]. Instead, they differ in terms of computational complexity, memory demands, and by the amount of training data necessary or the effort needed to tune hyperparameters. In the following, a brief description of selected ML algorithms used in this work is given. An in-depth discussion of these and other algorithms can be found in [11, 104].

Naive Bayes The naive Bayes (NB) classifier provides a generative model of the feature space. It is grounded on Bayes' Theorem, exploiting the conditional probability to see a label given a particular feature. In this way, it is used to estimate the probability distribution of the feature space given a specific class label. Thereby, the estimate is based on the (naive) assumption that, given a specific class label, the corresponding predictors are conditionally independent of each other. Thus, if features interact with each other (e.g. they are correlated), the performance of naive Bayes (NB) most likely degrades. The inference is inexpensive concerning computational and memory demands.

Support vector machine The support vector machine (SVM) is a kernel-based discriminative classifier or regression method. Utilizing the kernel trick, a support vector machine (SVM) constructs a hyperplane that allows non-linear separation of the feature space. Often polynomial, Gaussian, or radial-basis functions are used as kernel functions. The hyperplane is chosen such that it maximizes the distance between two separated classes (or minimizes it in case of a regression problem). Therefore, out of the feature space, candidates are selected that serve as *support vectors*. In general, SVM can be considered robust against over-fitting and capable of fitting highly non-linear problems. Because support vectors need to be stored, if many are needed for an SVM model, it is also a memory and computation-intensive. Furthermore, choosing an appropriate kernel can be difficult.

[9] In fact, it can be shown that in general given any 2 methods non of them can be superior to the other in all regards. This is known as the *no free lunch theorem* [258].

Gaussian process A Gaussian process (GP) (also known as Kriging), is a generative non-parametric kernel-based regression and classification method. With Gaussian process (GP), the observations of the training set are seen as random samples from a multivariate Gaussian distribution. That is a set of multiple functions, which is defined by a mean and a covariance matrix. Modeling the distribution function is similar to the NB idea. GP-based models are typically robust against noise, and inference is fast. Also, memory requirements are low. Training and hyperparameter tuning, in contrast, is difficult compared to other methods.

K-nearest neighbor The k-nearest neighbor (KNN) method belongs to the group of lazy or instance-based learners. Classification or regression is based on a query of the similarity (or distance) of a new observation and the known samples from the training set. Typically, a Euclidean distance measure is used. Each new observation is then classified by a majority vote with respect to its k nearest neighbors. k-nearest neighbor (KNN) is well suited to solve high-dimensional non-linear problems. However, the entire feature space needs to be stored, and in case of an exhaustive implementation, all distances need to be calculated to generate a prediction. That makes KNN both computational and memory intensive. Furthermore, if the neighborhood k is small, it is prone to over-fitting.

Decision tree The decision tree (DT) is a discriminative method, which follows a divide-and-conquer approach, meaning that multiple decision rules are created and arranged in a tree-like structure. Different metrics to decide on the quality of a split (decision rule) exist. DTs allow non-parametric modeling, for regression or classification even on categorical data. Because, only decision rules are stored, and only one path of a tree is evaluated for a given prediction, they are lightweight in both computationally and memory demands. Un-constrained trees (unlimited depth, splits with few samples only) are, however, prone to over-fitting.

Multivariate adaptive regression splines The method of multivariate adaptive regression splines (MARS) is a discriminative regression technique. It can be written as a weighted sum of multiple linear regressions, which are called basis functions. Each basis function is build up as a linear function. It is chosen such that it describes a local fraction of one specific independent variable from the feature vector. Through interactions between various basis functions, complex and non-linear relationships can be modeled. The inference is fast, and memory demands are low.

Artificial neural network Another discriminative method is the artificial neural network (ANN). Here, *neurons* are interconnected in a graph-like structure. These neurons are grouped in layers. The internal layers are named hidden layers. An ANN with only one hidden layer is referred to as a shallow network, whereas ANNs with several hidden

layers are also referred to as deep neural networks. In the classical fully-connected (dense) ANN, each neuron is a weighted sum of all its inputs (from the previous layer) and a bias value. The sum is then used in an activation function (linear, radial basis, logistic, and others). In regression, often a single neuron is found in the last layer, which equals the response. In classification, typically one neuron exists for each class. The class label is e.g. chosen by a majority vote. With ANN, non-linear functions can be modeled[10]. Depending on the architecture (e.g. network size), lightweight but also complex models can be built. That is for both computational and memory demands. Tuning hyperparameters or forecasting a suitable architecture is difficult.

Ensembles Instead of using a single model, multiple models can be used in combination, which is then called an ensemble. For instance, with the *random forest* algorithm, multiple DTs or *decision stumps* (that is a tree with a single decision rule) are used in a majority-vote fashion. The idea is to combine multiple independent weak-learner or base-learners to construct a stronger model. Ensembles are sub-dived into *bagging* and *boosting* approaches. In the case of *bagging*, training is on different subsets of the training data. Either all features are used, but only sub-sections of the data set (Pasting), or vice versa, i.e. the entire data set but only some features (*subspaces*). Also, mixtures of those partitions are possible (*random patches*). *Boosting* instead is an iterative process. Again, training is on different partitions, but errors or misclassification rate is used to weight new estimates. In this way, with each iteration, the training concentrates on *hard* data points. Because the base-learners can be very simple, memory and computational demands are typically acceptable for small ensembles. For the same reason, the ensemble is often less sensitive to over-fitting.

[10] In this respect, the universal approximation theorem exists, which points that neural networks can approximate any continuous function in the euclidean space.

3 Wireless Body Sensor: BI-Vital

The very beginning of wireless body sensor (WBS) applications is the sensor node's hardware itself. Regarding the design and implementation of WBS different, partly contradictory, requirements need to be fulfilled. Some essential properties of WBSs are apparent and can be deduced from their intended use as unobtrusive and portable devices, which are to be worn directly on or near the wearer's body.

These properties include miniaturization of the overall system in terms of size and weight, the presence of wireless communication capabilities and the integration of an independent power supply, through a (rechargeable) battery or application of an energy-harvesting method. Therefore, and besides application-specific requirements, the designer of a WSN has to optimize size, cost, robustness, storage, and computation needs as well as the total energy demands concurrently.

In the subsequent, the design of the WBS *BI-Vital* (BI-V5.0) is presented. It is a follow-up device whose development has already been discussed in [270]. In summary, the BG-V5 is a highly-integrated, yet scalable WBS for educational use and research, which supports low-power or high-performance applications.

In this chapter, firstly, related systems are presented (sec. 3.1), and in the following fundamental requirements are deduced (sec. 3.2). In the following, its implementation details are outlined (sec. 3.3). This includes the selection of components and other hardware-specific design decisions like the choice of an antenna, or the enclosure's construction are explained (sec. 3.3.2). Moreover, it comprises an outline of the firmware's architecture (sec. 3.3.3).

To conclude, the BG-V5 is empirically characterized (in terms of performance and power consumption) and compared to its predecessor (sec. 3.4). This chapter closes with an application example of the BG-V5 for the use as an on-line ECG classification system to detect certain arrhythmias utilizing a convolutional ANN (sec. 3.5). Finally, the results are summarized, and a prospect on the future development of WBSs is given (sec. 3.6).

3.1 Related Systems

In the following, a brief overview of modern WBSs' hardware architecture is given. Single and multi-sensor architectures are refereed. The explanations focus on specific design considerations and application or use-case examples.

3.1.1 Consumer Devices

Today, numerous consumer graded WBS are available and sold using terms like fitness or activity trackers, health-bands, or heart rate monitors. These devices come in different shapes and are intended to be worn on the wrist, chest, or hip. In the following, the design of typical devices is described.

3.1.1.1 Fitness or Activity Trackers

Fitness or activity trackers are brought to marked by different companies already. Among them are Fitbit Inc., Withing SA, Misfit Inc., or Xiaomi Corporation, to name a few. In general, the manufacturers do not reveal information on the internal design, components chosen, nor the algorithms implemented.

However, for the *FitBit Flex*, which was released in 2013, images of the internals exist (Figure 3.1). Also, reverse engineering results are available and were briefly discussed in a presentation at *Hack.lu* conference in 2015 [23]. Therefore, the device serves as an example at this point. A comparison of the Fitbit Flex and competing products is given in [121]. The Fitbit Flex is furthermore an interesting device because it was one of the first commercially successful WBS.

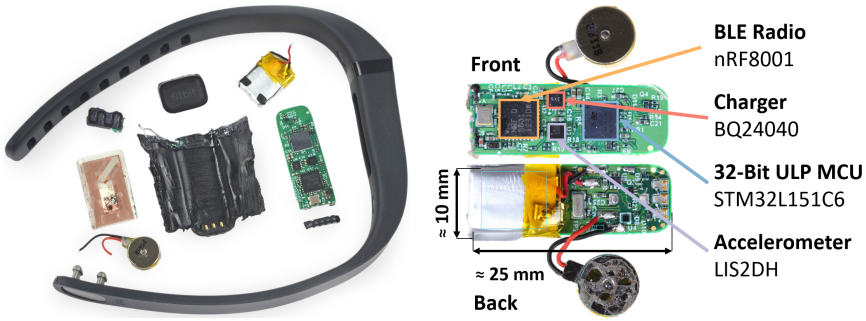
The devices can be seen as an example of a minimal WBS system (sec. 2.2.4). It contains a single sensor only. This is a 3-axial MEMS accelerometer (LIS2DE). The system furthermore contains a 32-bit MCU (STM32L151C6), a BLE radio for communication (nRF8001), and a battery charger (BQ24040). Along with the lithium polymer battery (Li-Pol) and vibration motor, the dimension of the overall is around 25 mm x 10 mm (width x height).

All information tracked and provided by the Fitbit Flex is derived from the accelerometer signals. On the one hand, acceleration is used to count steps. On the other hand, acceleration intensity and frequency are taken as sleep quality indicator. The device works hand-in-hand with a smartphone application, which is used to visualize and keep a diary of the information. Regarding step counting, validity was verified by [181].

Similar architectures can be found along with other manufacturers of wrist-band like fitness trackers. Nearly all of them embed an acceleration sensor (86 %) or heart rate sensor (33 %) [24]. Some forgo the vibration motor or light-emitting diodes (LEDs) and make use of tiny displays. Some integrate flash-memory or make use of SoC-architectures that combine MCU and radio in one chip instead of using independent solutions.

3.1.1.2 Heart Rate Monitors

As with fitness trackers, consumer graded HRMs are widely available on the market. They are typically clipped on to a chest strap. These devices then make use of a 1-lead ECG to capture heart rate. Alternatives can be found with wristbands that use



(a) Dismantled device **(b)** printed circuit board (PCB) details

Figure 3.1: A modern pedometer or fitness-tracker, the *Fitbit Flex*. On the left, the dismantled device (a). All the physical components of the wristwatch-sized device can be seen: Wristband, opened enclosure, rechargeable battery, vibration motor, NFC and BLE-antenna, and PCB. Next to the photograph (b), detail of the internal hardware or PCB. iFixit, CC BY-NC-SA 3.0, `https://de.ifixit.com/Teardown/Fitbit+Flex+Teardown/16050`

optical methods, e.g. photoplethysmography (PPG). Well-known manufacturers are Garmin Ltd., Suunto Oy, Wahoo Fitness LLC, or Polar Electro Oy. The validity of mobile HRMs was examined by [250]. They found close agreement of RR-Intervals for 2 consumer-grade HRMs compared to a medical 5-lead ECG system.

With early HRMs, the electronic components were placed in an enclosure together with the electrodes. Today's models are often mounted on top of a chest strap using snap fasteners. The shape of these devices is very similar, typically realized as round or oval nodes. This design has the advantage that the electrode belt is washable without the electronics.

Regarding chest straps, textile or polymer-based electrode material is used. This also often causes differences in signal quality as electrode polarization, impedance, and noise are worse compared to medical electrodes[1]. Nevertheless, heart rate detection is typically in close agreement, as most distortions can be handled by filtering the raw signal accordingly [54].

The device itself functions almost exclusively as a data sink for the heart rate. The 1-lead ECG is amplified, converted by an analog-digital converter (ADC), and then further processed by the MCU. The calculated heart rate (HR) information is then transmitted wirelessly. Standards used are based on the 5.3 kHz band (gym equipment) or 2.4 GHz ISM-band (Protocols used, e.g. BLE or ANT to allow for operation with a smartphone). [191]

In addition, some models also have integrated GPS, motion sensors, or memory units. An example of a modern HRM can be found with the *Wahoo Fitness TICKRx*. The

[1] Medical-grade electrochemical electrodes use silver-silver chloride pairs (short Ag-AgCl).

TICKRx outer dimensions are 70 mm x 38 mm x 13 mm (width x height x depth). It weighs 8.5 g and is powered by a coin cell. Its enclosure is waterproof (IPX7). Details on the internals are not provided[2].

3.1.2 Research Devices

Besides commercial products, WBSs that are designed for research purposes are available on the market as well. Those systems can also be referred to as multi-modal WBSs because they often combine multiple sensor elements. In the following, the design of WBSs marketed for application in research is outlined.

3.1.2.1 ActiGraph

The ActiGraph is a WBS widely used in research. Looking up the search term "ActiGraph" in the NCBI PubMed Central database alone, results in 2439[3] publications from sociology, epidemiology, or sports science referring to the device.

The most recent device is the wrist-watch like *wGT3X-BT*[4] (2018). It can be worn on the wrist, waist, ankle, or thigh. The enclosure is water-proof, has a physical dimension of 33 mm x 46 mm x 15 mm (width x height x depth) and weighs 19 g. It has a 4 GB built-in memory storing data for up to 180 days (1-Minute samples, approximately 256 B/min). It offers a USB-interface and a wireless interface via BLE.

The wGT3X-BT uses 3-axis MEMS accelerometer (±8 g) to track physical activity and sleep. An ambient light sensor is integrated as well to support the distinction between activity and sleep. Additionally, a heart rate monitor (sec. 3.1.1.2) can be connected wirelessly to the device via BLE. Its data is then additionally recorded. To reliably detect if the device is worn, a wear-sensor is integrated into the device, which is based on a capacitive coupling to the wearer's skin. Moreover, the device can detect and log the presence of other nearby ActiGraph devices.

3.1.2.2 Shimmer 3

Another manufacturer of WBSs for research purposes is Shimmer. Its development started around 2008 continues to this day.

The *Shimmer 3 ECG*[5] is a WBS that combines 9-DoF inertial measurement unit (IMU) and 5-lead ECG analog front end. The analog front-end can also be adapted to 3, or 2 lead use, as well as to record EMG signals. The IMU consists of 4 different sensors: The LSM303DLHC from STMicroelectronics serves as a wide-range accelerometer (±2 g to

[2] Some internal details can be deduced from images published by the Federal Communications Commission (FCC) in report PADWF109 available to the public domain `https://fccid.io/PADWF109`.
[3] September 2019, https://www.ncbi.nlm.nih.gov/pubmed/?term=ActiGraph
[4] ActiGraph LLC, `https://www.theactigraph.com/actigraph-wgt3x-bt/`
[5] Shimmer, `http://www.shimmersensing.com/products/ecg-development-kit`

±16 g) and magnetometer (±130 µT to ±810 µT). Besides, a low-noise accelerometer, the Kionix KXRB5-2042, is build-in. The gyroscope and the pressure sensor are the Invensense MPU9150 (±250 DPS to ±2000 DPS) and Bosh BMP180 (300 hPa to 1100 hPa), respectively.

It is powered by a Li-Pol (450 mA h, 3.7 V, 1.7 W h). The device's physical dimension is 65 mm x 32 mm x 12 mm (width x height x depth) and weighs 31 g. Typically, the device is worn with a necklace or is fixated by a strap on to the origin of interest (e.g., leg or arm). Processing in the Shimmer is done by a 16-bit MSP430 MCU (24 MHz, 16 kB RAM, 256 kB ROM). For communication, the RN42 from Microchip[6]is used, which implements a Bluetooth 2.1 protocol stack.

The Shimmer is an excellent example of a WBS acting as a data-logger or sink node. It is available as a commercial product yet was designed with research and development purposes in mind. Its design can be taken as a reference example of optimizing size, cost, storage capabilities, robustness, and flexibility.

3.1.2.3 MoviSense

The development of the MoviSense sensor started as a research project and was commercialized as a spin-off from KIT Karlsruhe University in 2009. The newest version of the MoviSense is the *EcgMove 4*[7].

Embedded sensors include a 1-lead ECG analog front end, 6-DoF inertial measurement (accelerometer ±16 g and gyroscope ±2000 DPS), air pressure (300 hPa to 1100 hPa) and temperature. The internal memory of the device offers 4 GB of storage. The wireless communication is realized utilizing BLE. The device's enclosure is water-proof (IP67) and has a physical dimension of 62 mm x 39 mm x 12 mm (width x height x depth). It weighs 26 g and is powered by a Li-Pol. The device is worn at the chest, clipped onto an electrode-belt.

The MoviSense is designed to fit different research applications. It is mainly used as a data-logger, storing all sensor data for subsequent off-line analysis. It is thus a good example of a WBS acting as a data sink. Just like the Shimmer 3 (sec. 3.1.2.2), it is a commercial product designed for research and development use. In comparison to Shimmer 3 it offers a water-proof enclosure and extended storage capabilities.

3.1.3 Previous version: BG-V4.2

The *BI-Vital* version 4.2 (BG-V4.2) is the predecessor of the new version 5 that is developed as part of this work. Its development is originally described in [64, 254]. It is a WBS that combines 1-lead ECG analog front end, a wide-range accelerometer (*LIS331HH*[8], ±6 g to ±24 g), and additional analog circuits to connect a temperature

[6]Microchip RN42, `https://www.microchip.com/wwwproducts/en/RN42`
[7]movisens GmbH, `https://www.movisens.com/en/products/ecg-sensor`
[8]STMicroelectronics N.V., `www.st.com/en/mems-and-sensors/lis331hh.html`

sensor (NTC), electrodermal activity (EDA) electrodes or a respiration sensor [269]. It is powered by a non-rechargeable coin-cell (CR2025, approximately 140 mA h to 170 mA h, 3 V, 0.4 W h to 0.5 W h). The enclosure was adopted from a commercially available product with the dimensions of 63 mm x 47 mm x 12 mm (width x height x depth). The module weighs 31 g and is worn on top of an ECG chest strap (electrode belt) or compatible clothing.

As with the Shimmer, the processing is done by a 16-bit MSP430FG4618 MCU (16 MHz, 8 kB RAM, 116 kB ROM). Data can be stored into the embedded S25FL064K NOR-flash (64 Mbit) or transmitted wirelessly with the nRF24L01 based on a proprietary protocol in the 2.4 GHz ISM-band.

3.2 Requirements

Due to the close relationship between WBSs and WSNs, they share a common design space. Operating time (or power consumption) and space requirements are as important for WBS as they are for WSN. Developing a WBS for research or educational use also implies covering a broad spectrum of functionalities and applications. However, as the device is not necessarily intended to be produced in high quantities, cost efficiency is less critical.

In order to cover a wide range of use-cases, it is intended to integrate more than a single sensing element. Also, components should be selected or designed in a way that they offer the highest accuracy. These requirements dominate cost, yet energy and size constraints must not be violated.

Likewise, communication remains essential, although with a different focus than in WSN. In contrast to WSNs, WBSs do not primarily interface with other devices. Instead, they need to communicate with the wearer[9]. Communication, therefore, needs to be thought of as part of the user interface because it is usually triggered by the user or intended to notify him or her.

All these requirements affect each other. For instance, specifying the device's dimension limits size for components and energy storage. Therefore, and besides application-specific requirements, the designer of a WSN has to concurrently optimize size, cost, robustness, storage- and computation needs while keeping the total energy demands low. The considerations taken into account during the design phase of the BG-V5 are given in the following.

3.2.1 Shape and Enclosure

Before components can be selected or the shape of the enclosure can be discussed, the measuring position must be determined. Different positioning for a WBS can be thought

[9] Communication might happen through a mediating device, e.g. a wrist-watch.

off. Following the use case, many body parts for wearing a WBS are conceivable. These include the wearer's head, shoulders, chest, waist, arms, hands, legs, or feet. In some applications, the position is of lesser importance, e.g. if only environmental data (e.g. air quality) is to be measured.

Other examples of WBS (e.g. presented in sec. 3.1), need to be attached to the chest, in order to enable the recording of an ECG or to the hip to allow detection and counting of steps. In order to analyze not only steps but record gait parameters, a wearable is most probably best positioned on the wearer's feet or legs. If, however, an analysis of, e.g. throwing movements is intended, a device must be placed on the shoulder, arm, hand, or on all these positions. Apart from that, other measures could determine distinctive positioning, e.g. the wearers' ear is a good choice in case that the body's core temperature estimation is intended.

The chest is an appropriate choice because it allows to include a wide range of parameters. From own experience [269] and other work [14], it is already known that accelerometer signals recorded from the chest (i.e. body's center of mass) are well suited to perform activity recognition and estimate EE. If it is intended to obtain heart rate, and ECG is assumed to be the most precise method, this also determines the positioning at the chest (for the complete device or at least electrodes). The use of an ordinary chest strap equipped with ECG-electrodes is suitable. Also, other parameters can be obtained from the chest, like respiration [269] or skin temperature. The possibility to place the module at other body parts, e.g. through hook and loop fasteners, remains unaffected.

Placing the sensor on the chest, i.e. clipping it onto a chest strap, limits the overall dimensions of the device. The examples of other body sensors, like the BI-V4.2 or Shimmer (sec. 3.1), can serve as orientation. Also, it is needed to keep the electrodes' distance of around 45 mm, which is found in commonly available chest straps. It is furthermore intended to keep the sensor's dimensions in the same order as it is found with comparable devices. This is setting a limit for the absolute maximum dimensions of about 70 mm x 40 mm x 15 mm (width x height x depth). In this way, it should be possible to hold the device in hand and to attach it not only to a chest strap but as well to other parts of the body (like foot or arm) without disturbing the wearer.

3.2.2 Energy supply

Regarding the energy supply of a WSN, having a fully self-sustainable solution would be ideal. It would maximize mobility and prevent the need to replace depleted batteries (primary cells) or the need to recharge a secondary cell. Such solutions, known as *energy-harvesting*-systems, have recently been presented for WSN [129]. In their experiment Kim et al. [129] yield around 1 mW from a 2 W-RF source (0 dB m). This was for a wearable device positioned at the chest. However, this required a radio device (i.e. a smartphone) permanently radiating energy nearby to the wearable device.

Typically, the amount of *harvestable* energy from other sources (except solar energy), e.g. vibration or thermal differences is even lower. For vibration or thermal differences on the human body, it is about $4\,\mu W/cm^2$ to $30\,\mu W/cm^2$ [74, p. 56]. Facing these limitations, it can be concluded that energy-harvesting methods are applicable for specific use-cases only. Due to their unpredictable and thus unreliable nature, they are mostly applicable in adaptive or event-driven applications, where sensing or transmitting information is rarely needed.

For other applications, in general, the power consumption of a WSN will exceed the *harvestable* amount of energy by orders of magnitude. For instance, power consumption for continuously transmitting raw data with a Shimmer 3 ECG sensor (sec. 3.1.2.2) is reported to be around 30 mW [229][10].

The intended application of the BG-V5 as a data-logger pinpoints the use of a battery. Taking the assumption of a maximum volume of 42 mL for the overall system (outer dimensions) a volume of 1 mL to 10 mL can be reasoned for the battery. This would result in a maximal capacity of 0.55 W h to 5.46 W h given a lithium-ion manganese-oxide (Li-MnO_2) based cell or 1.35 W h to 13.5 W h assuming the use of zinc-air primary cells, which have a typical energy density of 546 W h/L or 1350 W h/L), respectively.

Referring to the Shimmer 3 streaming application, this would allow a run-time of 18 h to 450 h. In reality, however, different challenges reaching these theoretical run-times using primary cells are faced. That is explained by the fact that, although the energy density of primary cells (such as Li-MnO_2 or zink-air) is high, their power density is limited in comparison to secondary cells (e.g. Li-Pol). For embedded applications, even if the mean power consumption is low, the peak consumption thus might exceed the maximum tolerable discharge current (power) of a primary cell. Secondary cells, solve the issue of low power density as they are more tolerable to high peak discharge currents. However, their energy density is typically lower.

It can be concluded that the optimal choice of an energy source strongly depends on the application. Energy harvesting is best for autonomous applications but is applicable for a limited set of use cases only. Primary cells are advantageous in scenarios where there is low power consumption, but only if no peak loads exceeding the maximal discharge current appear. Also, due to the low self-discharge and high energy density of primary cells, the run-time of a device can be maximized. However, given an application with higher power consumption or peak loads, the self-discharge rate is negligible, and the higher energy density is canceled due to the lower power density.

Another problem with primary cells is that they need to be replaced once they are depleted. Also, it can be challenging to determine the exact state of charge because of the flat discharge curve. In general, secondary cells provide more flexibility. Their power density is high, and they can be recharged and thus reused many times. Their use also reduces maintenance costs by avoiding the need to open the device and replace a depleted primary cell.

[10] The exact reported value is 9.46 mA with no battery voltage specified, but assumed to be 3 V.

3.2.3 Wireless Communication

Wireless communication is needed to transmit data from one device to another. In WBS, often light-weight information is transferred, e.g. HR or temperature to a displaying device. These can be compactly represented (8 bit) and are transmitted at a relatively low sampling rate (1 Hz). The data-rate is thus low (8 bit/s). Raw data, however, such as the ECG, are larger both in the representation (16 bit), but above all, the typical sampling rate is higher (100 Hz). Correspondingly, the data-rate increases (1600 bit/s).

Simple broadcasting methods are suitable for data transmission when low latency and high data-rate is needed. The data packets consist of nothing more than a preamble, the payload, and an optional checksum. In consequence, the overhead is low, and the net rate (or goodput) is often close to the maximum theoretical throughput. The advantages of broadcasting go hand in hand with the lack of advanced or comfort functions, such as addressing, whitening, re-transmission, channel-hopping, encryption, and others on the link-layer level. Furthermore, there is a lack of interoperability regarding the application-layer level.

The selection of a wireless-communication technology in WBS is constrained by the fact that the sensor is applied to the human body. This is because the tissue causes shadowing or absorption of the electromagnetic waves emitted. Here the frequency is the deciding factor. For instance, in [262], it is shown that with 820 MHz damping is typically lower (42.05 dB to 54.60 dB) compared to the application of radio waves in the 2.4 GHz domain (44.46 dB to 64.72 dB)[11]. For the sake of interoperability, existing standards like Bluetooth Low Energy (BLE) somewhat enforce to adapt to the 2.4 GHz band.

BLE has become a wide-spread wireless protocol standard for IoT, WSN, or WBS [98] on the 2.4 GHz ISM-band. The first specification for BLE has been available since 2010. It was specified with low computing power and low data-rate applications in mind. Therefore, it is primarily suitable for communication with devices such as sensors or beacons.

The theoretical maximum transfer rate of BLE (version 4) is 1 Mbit/s. Due to the protocol overhead, as mentioned before, the actual maximum transmission rate is significantly lower. Communication between two BLE devices is organized in *connection events*, which happen at fixed intervals. In each event, a limited number of packages is allowed (typically 6) with a maximum payload of 20 B. Having set, the lowest possible connection interval of 7.5 ms (approximately 133 packages per second), the maximal throughput under ideal conditions is limited to 15 960 B/s or 0.12 Mbit/s.

Modern MCUs for wireless transmission purposes are designed as multi-protocol SoCs that offer software-defined protocol stacks. Thus, it is possible to use existing standards, but keep full flexibility concerning the use of the transceiver for proprietary

[11] Presented values are minimum and maximum of the measurements from experiment “CM4”, comparing path loss in line-of-sight conditions, where the receiver is located 1 m to 4 m away from the sender attached to the human body [262].

protocol development. This typically relaxes constraints on the choice for a specific solution because this offers the possibility for adaptation to future protocols (of course, the restriction remains that these must possibly be developed by oneself.). Moreover, multi-band solutions are available as well.

3.2.4 Non-Functional requirements

Interfacing with the wearer is the key non-functional requirements in WBS. Thinking of the user interface of a WBS, wireless communication can be sufficient as a single interface solution to the user. With WBS tactile interfaces (e.g. vibration motors), LEDs or even smaller displays and devices for acoustic feedback (e.g. a piezo buzzer) are also commonly found.

While the latter only allow for indication, gesture recognition, or tap-detection are other user interface elements, which also provide an input channel. Besides, speech recognition is a way of communicating with the wearer. As with all interfaces, they must be designed with the user in mind. However, even simple solutions, like colored Morse code like interfaces can transmit valuable information to the wearer with only a short learning phase [49]

Besides wearer (user) interaction, other non-functional requirements exist in WBS design, including:

- a mechanism to uniquely identify the device.
- data security, e.g. prevent unauthorized access.
- wearer safety, e.g. the device must not endanger the wearer.
- mechanical robustness, e.g. device must withstand sweat, heat, and minor crashes.
- an internal timer to synchronize measurements.

3.3 Implementation

In the following, the implementation details of the BG-V5 are given. This includes the selection of hardware components and sensors (sec. 3.3.1), the design aspects of the enclosure (sec. 3.3.2), and a summary of the firmware architecture (sec. 3.3.3).

From the precluding requirements discussion above, it is inferred that the device:

- must be attachable onto a chest strap (ECG-electrodes, body's center of mass).
- must not exceed dimension of 70 mm x 40 mm x 15 mm (width x height x depth).
- must be powered by a secondary cell.

Moreover, because different use-cases exist, it must provide a scalable set of sensors, computational resources, and communication protocols in order to fit multiple research applications.

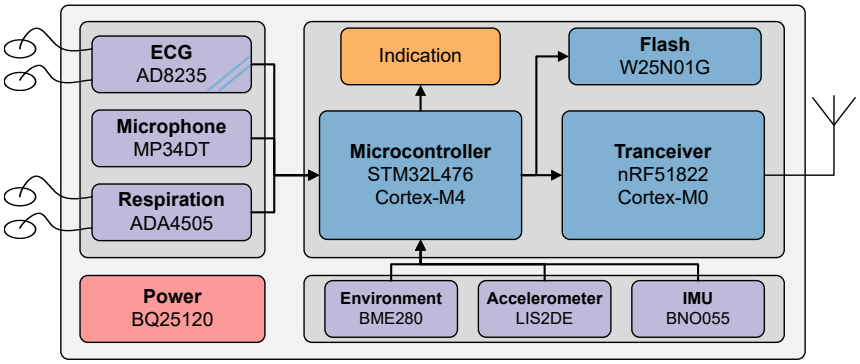


Figure 3.2: Hardware architecture of the BG-V5. All Sensors, the power management, and indication components are connected and controlled by the STM32 MCU. Data is either stored in the flash or transferred to the nRF51 transceiver for wireless transmission.

3.3.1 Hardware Design

The hardware design and the selection of components become apparent, having the requirements identified (sec. 3.2). The BG-V5's design follows one of a classical sensor node. As such it combines a central MCU (sec. 3.3.1.1), sensors (sec. 3.3.1.2 - 3.3.1.4), storage (sec. 3.3.1.5), wireless connectivity (sec. 3.3.1.6) and a power management (sec. 3.3.1.7). Subsequently, the selection of components and other design decisions are discussed (Figure 3.2).

3.3.1.1 Microcontroller

The fields of application of WBS are very different. In the simplest case, they act as a data-logger, but complex systems with embedded inference are also possible. Thus, also the requirements regarding the MCU vary, which is why a scalable MCU is favorable. Thereby, it must be possible to balance performance and power consumption. In this sense, the MCU must be capable of operating in a low-power state but also offer adequate performance. That is to allow for high autonomy (low-power) and high performance (embedded algorithms, e.g. inference). Moreover, it must be available in a small package (wearable design) and offer required interfaces (e.g. digital and analog).

A priori comparison of performance and power consumption trade-offs are difficult and task-dependent. It is often not sufficient to rely solely on benchmark results. This is because benchmarks typically fail to attribute specific characteristics of the later application. For example, on the computational level, a benchmark could lack a comparison of architectural differences like the presence or absence of a floating-point unit (FPU) or single instruction multiple data (SIMD) operations [192].

Table 3.1: Comparison of different MCUs based on benchmark results (ULPMark). The best models available at design time (2016) are listed below. In addition, more recent benchmarks are given that also take into account the efficiency of the peripherals (Per.). Vendors are ON Semiconductor (ONS), Ambiq Micro (AM), Microchip Technology (MT), STMicroelectronics (STM), Texas Instruments (TI)

Device Vendor	Family	Chip	ULPMark		Year
			Core	Per.	
ONS	RSL10 Rev 1.0	C.-M3	1090		2018
AM	APOLLO512-KBR Rev.A3	C.-M4	395	33	2017
MT	ATSAML11E16A rev B	C.-M23	282		2018
STM	STM32L552 Rev1	C.-M33	267	34	2018
STM	STM32L433RC-P	C.-M4	264	107	2017
STM	STM32L476ZG-P	C.-M4	227	81	2017
TI	MSP432P401R Rev. C	C.-M4	164	7	2018
STM	STM32L476	C.-M4	152	63	2018
AM	APOLLO512-KBR Rev.A3	C.-M4	378		2015
STM	STM32L476RG	C.-M4	188		2015
MT	SAML21J18A-UES Rev.A-DC	C.-M0+	186		2015
TI	MSP430FR6972 Rev.A	MSP430	124		2015

Also, in ESs, the efficient use of peripheral components is an issue. Here, efficiency depends on whether a specific peripheral can be used without other components, especially the central processing unit (CPU), being active. Furthermore, it is preferable to have a wide range of peripheral components embedded, to reduce the total count of components. Moreover, this keeps the PCB's floor plan compact, reduces the costs, and avoids communication overhead between additional components.

New benchmarks like the *ULPMark*[12] address these issues. The ULPMark is a family of 3 different benchmarks. It is designed to compare the energy efficiency of MCUs in battery-powered applications. The ULPMark-CoreProfile targets the core's power consumption only (active and low-power mode). With ULPMark-PeripheralProfile ADC, pulse-width modulation (PWM), serial peripheral interface (SPI), and real-time clock (RTC) peripherals are included in the benchmark as well. To limit the list of possible MCU-candidates, the ULPMark results were taken into consideration (Table 3.1).

The STM32L476 [223] combines high performance and power-efficiency. Moreover, it offers a multitude of peripherals, among them, quad serial peripheral interface (QSPI) flash-memory interface, pulse-density modulation (PDM) microphone interface, USB-connectivity (2.0), 12-bit resolution ADC and digital-analog converter (DAC).

[12] ULPMark (Ultra-Low-Power Benchmark) provided by EEMBC (Embedded Microprocessor Benchmark Consortium, `www.eembc.org/ulpmark/`)

Its core clock can be scaled from 0.1 MHz to 80 MHz, which allows for low-power or high-performance applications. The internal flash-memory is up to 1 MB with 128 kB static random-access memory (SRAM). It is available in a 72-ball WLCS-package (wafer level chip scale), with dimensions of 4.4 mm x 3.6 mm (width x height).

Another possible candidate is the Apollo 512 [22], which is furthermore superior in terms of core efficiency. However, the internal ADC is less accurate (10-bit vs. 12-bit resolution), nor does it embed USB connectivity. Another argument in favor of the STM32L4 is its efficiency in terms of peripheral usage (Table 3.1). This is because the STM32L4 core and peripherals are clocked independently. Thus, the CPU can run in low power mode or be turned off, while peripherals, e.g. the ADC, are sampled. [194]

It should be noted that as of today, more promising candidates can be identified. Among them, the RSL10 is an interesting alternative. This is because it combines MCU and radio transceiver in one chip. However, it was not available when the BG-V5 was designed. Based on the benchmark results, size, and peripheral available, the STM32L476 is chosen as the main or central MCU for the BG-V5.

3.3.1.2 Motion Sensors

Motion is a crucial entity of interest in wearable applications. It is either used to detect user-interaction (e.g. the user picks up the device) or to measure the wearer's motion (e.g. step counting). Modern IMUs typically combine 3 different sensors, namely an acceleration sensor, a gyroscope, and a magnetometer:

- Acceleration sensors have a build-in proof mass. Its displacement is proportional to the acceleration in the given axis. The sensing elements are passive. The measuring principle is typically capacity based. On earth, measurements are always affected by the earth's gravitational field, which allows calculating inclination (pitch and roll). [213]
- Gyroscopes are used to measure angular velocity. They are often realized as Coriolis vibratory gyroscopes, utilizing tuning forks configurations in which two proof masses vibrate in opposed direction. Rotation of this configuration results in an orthogonal vibration due to Coriolis force that is proportional to the angular velocity. The sensing elements are active. The measuring principle is typically capacity based. Measurements can be used to calculate angular displacement around the sensitive axis. [213]
- Magnetometers measure magnetic fields. Internally, they are often realized as hall sensors. Thus, a current that flows perpendicular to a magnetic field will result in a proportional voltage, measurable across the magnetic field's axis (Hall effect). Within an IMU, a magnetometer can be used to find the north pole position, thus enable absolute orientation. Its measurement is affected by auxiliary electromagnetic field or metal. [234]

Today's accelerometer, gyroscope, or magnetometer sensors are sensitive on all 3 room axes. Thus combining them leads to 9-dimensions of freedom. For simple applications, a single accelerometer can be sufficient. If the orientation is of interest, the combined use of gyroscope and accelerometer is preferable. If the absolute orientation is of interest, an additional magnetometer can provide north pole direction, thus give absolute orientation.

The power consumption in MEMS accelerometers is a function of output data-rate, noise level, and measurement resolution. For the BG-V5, an acceleration sensor is selected out of 18 different devices from 5 manufacturers (Figure 3.3)[13].

The LIS2DW12 was found to be most suitable as it has the lowest power consumption in the intended data output range ($\leq$100 Hz). A possible alternative is found with the mCube MC3672. However, no information on low-sampling power consumption is available from the datasheet. The LIS2DW12 is thus chosen in favor. In the BG-V5, a variant, the LIS2DE12 [143] was chosen, which provides 8-bit resolution instead of 16-bit resolution.

The physical dimension of the LIS2DE12 is 2.0 mm x 2.0 mm (width x height). Its active power consumption is specified with 15 µW given an output data-rate of 50 Hz. Standby current is 1 µW. Its full-scale range covers ±2 g to ±16 g.

To fit for other applications, which require absolute orientation, an additional IMU was integrated into the design. It is intended for applications that require low-error and high-resolution, but in return, accept higher demands regarding power consumption. For such applications, the BNO055 [36] IMU was selected. The decision is based on practical experience and comparison.

The BNO055 offers the benefit of having a fully integrated motion processor, which offloads the main MCU from having to calculate fused motion and orientation information, e.g. as quaternions or Euler angles. Therefore, it can be considered a SoC. Its core is based on an ARM Cortex-M0 design.

The physical dimension of the BNO055 is 5.2 mm x 3.8 mm (width x height). Its active power consumption with all sensors active is specified with 37 mW given an output data-rate of 100 Hz. Standby current is 120 µW. The full-scale range of the accelerometer, gyroscope, and magnetometer covers ±2 g to ±16 g, ±125 °/s to ±2000 °/s or ±1300 µT to ±2500 µT, respectively.

3.3.1.3 Environmental Sensors

Environmental data, such as temperature, humidity, or atmospheric pressure, provide auxiliary information of the wearer's environment. For instance, pressure sensors can be used to add a dimension of freedom to IMU measurements (That is, combining the high-frequency components of changes in atmospheric pressure measurements

[13] Devices under consideration are: Analog ADXL312, ADXL362; Bosch BMA253, BMA423, BMA456, BMA280, ST LIS3DH, LIS3DSH, LIS2DW12, KIONIX KX022, KX126, KXCNL-1010, KXTJ3; NXP MMA8652FC, MMA8453Q, FXLS8471Q; mCube MC3672; TDK IAM20381

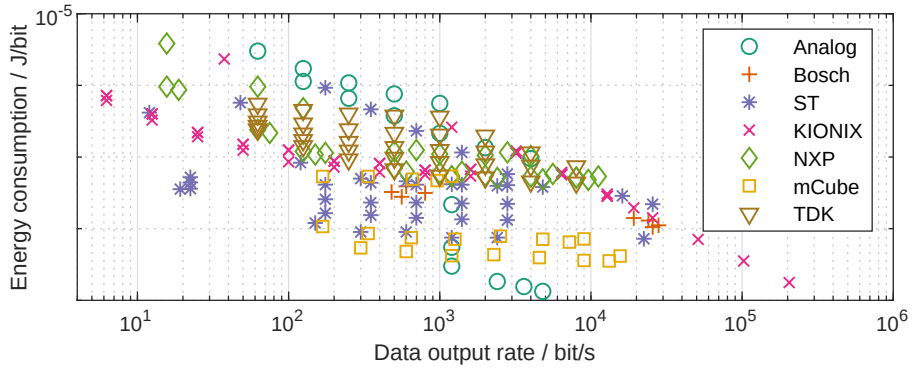


Figure 3.3: Comparison of energy consumption and data output rate of 18 different accelerometers from 5 different manufacturers.

to calculate altitude). Other environmental factors of interest could be luminosity, acoustic noise, or the presence of hazardous or harmful substances e.g. volatile organic compounds or carbon monoxide or carbon dioxide, to name a few.

For the BG-V5, environmental MEMS sensors to measure temperature, relative humidity, atmospheric pressure, and acoustic noise (microphone) were selected and integrated. Due to size constraints, only MEMS components are considered. As with motion sensors (sec. 3.3.1.2), many MEMS sensors to measure environmental information are available on the market. For instance, regarding air pressure sensors only, 34 different sensors from 18 different manufacturers were identified in a market analysis by [162]. Likewise, a broadly spread market landscape exists for temperature and humidity sensors as well as for MEMS microphones. A comparative analysis is thus impractical.

Regarding air pressure, relative humidity, and temperature, the BME280 [35] was selected because it combines all these sensors in a single package. No other device could be identified. Moreover, a pin compatible-variant (BME680) exists, which additionally allows measuring air pollutants, i.e. volatile organic compounds. A coarse comparison of the BME280 and other sensors revealed that it is comparable or outperforms them in terms of accuracy or power consumption (Table 3.2).

The physical dimension of the BME280 is 2.5 mm x 2.5 mm (width x height). Its active power consumption with all sensors active is specified with 15 µW given an output data-rate of 1 Hz. Standby current is <1 µW. The temperature, relative humidity, and air-pressure full-scale range cover −40 °C to 85 °C, 0 % to 100 %, or 300 hPa to 1100 hPa, respectively.

Regarding, the acoustic sensor, the digital MEMS microphone MP34DT04 [168] was selected. Because it is primarily intended to be used as a SPL instrument only, no extensive comparison in terms of noise or sensitivity was carried out. The selected

Table 3.2: Comparison of MEMS environmental sensors (temperature "T", relative humidity "H" and absolute air pressure "P") in terms of precision, maximal, and typical (sampled at 1 Hz) power consumption. Information is taken from the corresponding devices' datasheet.

Device (Manuf.)	Type	Precision	Power con. /μW max.	typ.
Bosch BME280	T	±1.0 °C	630	2
	H	±3 %	612	3
	P	±1.0 hPa	1285	5
Sensirion SHT85	T	±0.1 °C	1980	6
	H	±1.5 %		
Sensirion SHTC3	T	±0.2 °C	1419	20
	H	±2.0 %		
TE-Connectivity HTU21D	T	±0.3 °C	1350	-
	H	±2.0 %		
STMicroelectronics LPS25H	P	±0.1 hPa		63
TE-Connectivity MS5637	P	±0.1 %	3750	3

MP34DT04 has an acoustic overload point of 120 dB SPL and an equivalent input noise of 30 dB SPL [20]. The physical dimension of the MP34DT04 is 3.0 mm x 4.0 mm (width x height). Its power consumption is 1.08 mW if sampled at 2.4 MHz.

3.3.1.4 Vital-sign Sensors

Vital sign monitoring, e.g. capturing bodily functions of the wearer, is a key-function with respect to health- or sports-related applications. In BG-V5 analog circuits are implemented to record respiration rate from an external force sensor [269] and heart rate from a 1-lead (bipolar) ECG.

The components used for analog signal processing are the ADA4505-4 [3], which combines 4 operational amplifiers (OpAmps) in a single component. Additionally, the AD8235 [2] instrumentation amplifier (InAmp) is used. The components physical dimensions are 1.5 mm x 3 mm (width x height) or 1.6 mm x 2.1 mm (width x height) for the ADA4505-4 or AD8235, respectively.

The functional principles of the circuits are outlined in the following (the details of the circuit can be found in appendix A).

Respiration Rate The BG-V5 provides connectors for the respiration sensor described in [269]. It is based on a force-sensitive resistor, which is attached to a chest strap.

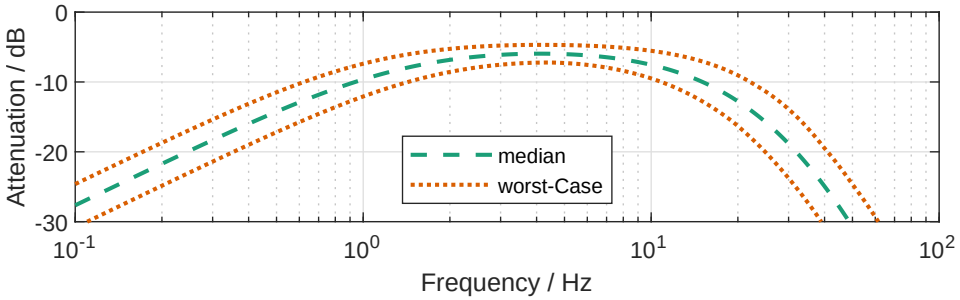


Figure 3.4: Attenuation of the ECG filter used in BG-V5. Here the median response from Monte-Carlo analysis is depicted together with results from a worst-case analysis. The circuit's reference design is found in [254], while here, the components used are different (Appendix A, Figure A.3).

It is based on the measurement of force changes between the chest and chest strap, due to thoracic movement. The resistance of the sensor is low during inhalation or high during expiration. To capture these changes in resistance, the sensor is used in a voltage divider configuration combined with an active low-pass filter (AD4505-4, Sallen-Key, cutoff-frequency: 0.80 Hz, gain: 1).

Heart Rate The heart rate measurement is based on ECG recording with a resolution of 12-bit. The analog pre-processing aims at amplifying the electric potential of the ECG, which is around 1 ± 5 mV. Moreover, only relevant frequencies in the range of 0.05 Hz to 100 Hz should be passed. For the practical application of a mobile ECG monitor, in order to reduce high-frequency noise and motion artifacts, bandwidth is typically limited to 1 Hz to 35 Hz. In addition, due to the electrode offset, high common-mode rejection is required.

The analog pre-processing used in BG-V5 is based on the previous design used in BG-V4.2. It is described in detail in [254]. The BG-V5's design differs in terms of the chosen components. Also, further adaptations in the discrete filter stages exist. This is due to the lower system voltage (1.8 V instead of 3 V) of the BG-V5 and general improvements of the overall power consumption. Also, a reduction of the total number of different components used (bill of material) is intended. The filters architecture and frequency response (Figure 3.4), however, remains unchanged:

In its input stage, passive RC-filters (common mode and differential mode) are used to remove the DC-offset and radio frequency interference. In this way, a high common-mode rejection ratio is maintained at the InAmp's input stage [115]. In order to avoid a dual supply solution, the signal is shifted towards an adjustable reference voltage provided by the DAC from the host MCU (sec. 3.3.1).

Following the passive RC-filter stage, the signal is amplified using the AD8235 InAmp. The InAmp's amplification factor is set to 5. It can be adjusted through an external resistor. Its maximal amplification gain is 200.

The pre-amplified differential ECG signal is then further filtered and amplified. Therefore, a 3-staged active filter configuration is used. This makes use of the *ADA4505-4*. At first, an integrator is added to the InAmp's reference input, which acts as an active high-pass (cutoff-frequency: 1.33 Hz, gain: -1). The 2nd stage is implemented as a low-pass multiple feedback filter (cutoff-frequency: 33.84 Hz, gain: -100). Finally, the low-pass filter in the 3rd stage is based on a Sallen-Key architecture (cutoff-frequency: 33.86 Hz, gain: 1).

The system's expected analog current consumption can be calculated as the sum of the static quiescent current of the active filters, attenuation stages, and power supply (AD8235 [2], AD4505-4 [3], BQ25120 [42, p. 11]). Furthermore, the current consumption of DAC and ADC also adds to this sum (STM32L476, DAC: [223, p. 178], ADC: [223, p. 178]). In total, the power consumption is expected to be 1.4 mW (worst-case scenario, eq. 3.1). Here, the power loss induced by the low-dropout (LDO) regulator used for the analog power supply also needs to be considered (eq. 3.2).

$$\begin{aligned} P_{ECG} &= U_{LiPo} \cdot \left(I_{AD8235} + I_{AD8235} I_{ADC} + I_{DAC} + I_{LDO}\right) \\ &= 3.7\,\mathrm{V} \cdot \left(30\,\mu\mathrm{A} + 21\,\mu\mathrm{A} + 16.6\,\mu\mathrm{A} + 315\,\mu\mathrm{A} + 0.9\,\mu\mathrm{A}\right) \\ &= 3.7\,\mathrm{V} \cdot 383.5\,\mu\mathrm{A} \\ &\approx 1.4\,\mathrm{mW} \end{aligned} \tag{3.1}$$

$$\nu_{LDO} = \frac{U_{LiPo}}{U_{Analog}} = \frac{3.7V}{1.8V} = 0.49 \tag{3.2}$$

The final design is configured to fulfill requirements, as stated in [254] and has the following characteristics:

- gain factor 500 to 1000 (approximately 250 mV to 1500 mV)
- signal filter to the range of 1.3 Hz to 33.8 Hz
- adjustable reference voltage
- high common-mode rejection (94 dB).
- low power (1.4 mW)

For future re-design, a fully integrated analog ECG front-end could be used. Such components are already available on the market. As an example, the *AD8232*[14] or *ADS1291I*[15] can be cited. Both offer advantages since they are smaller and offer extended features, e.g. adaptable filter or advanced noise rejection techniques. Their use was considered, and both were evaluated. However, re-using the BG-V4.2 solution was favored to build upon existing knowledge for the BG-V5 design.

[14] Analog Semiconductors, AD8232, `https://www.analog.com/en/products/ad8232.html`
[15] Texas Instruments, ADS1291I, `www.ti.com/product/ADS1291`

3.3.1.5 Flash memory

Regarding the design of ESs, flash-memory can either be integrated as a chip (part of the PCB) or a removable component. With the micro-SD standard removable flash-memory is available in a package as small as 15 mm x 11 mm (width x height). Matching sockets typically add 1 mm to 2 mm, which must be considered in PCB design. These devices are *managed* variants of bare NOR- or NAND-flashes. These facilitate the access to the underlying memory cells or blocks because they allow an abstract file-system like access to the memory.

In contrast, with un-managed memory, the pages and blocks (groups of pages) must be accessed manually[16]. However, this often allows using the memory more power efficiently [136] due to the flexible memory access. Furthermore, un-managed memory is typically available in smaller packages (BGA, ball grid array), which is important to match size-constraints.

Regarding the different architecture of NAND and NOR-flash, the choice for one or the other depends on the application. NOR-flash offers advantages in data retention (up to 20 years) and higher read speed (due to random access) compared to NAND (sequential read). It is thus preferably used to add additional memory for code storage and execution.

In contrast to NOR-flash, the use of NAND-flash is advantageous for general data storage. That is because typically, the capacity is high, and the power consumption is low. Yet, write and erase operations are slow. Another disadvantage is that the memory access has to happen block-wise. That effects each read and write operation. Moreover, memory can only be erased in full pages, which affects multiple blocks at once. This requires caching data and thus comes with the additional need for temporal memory on the host MCU. [232]

To store sensory data, a 1 Gbit flash-memory, the Winbond W25N01GWBIG [244] is integrated into the BG-V5 design. This is because of the higher memory density and advantages in power consumption of the NAND architecture compared to NOR-flash.

The W25N01 is an un-manged NAND-flash, organized in pages of 2 kB, which are grouped into blocks of 64 pages (128 kB). The component has an additional page buffer for read or write operations. Therefore, the data of a single page does not need to be temporarily stored by the host MCU. Instead, the buffer can be used to simulates random access to a page.

The physical dimensions of the W25N01 are 8 mm x 6 mm (width x height). Its active power consumption is specified with 45 mW for read, write, or erase operations. The standby power-consumption is 18 µW. No comparable flash-memory was identified during the design process.

[16] This means that an additional flash-transaction layer (FTL) must be added to the hosts firmware, in order to manage memory access.

3.3.1.6 Wireless Connectivity

To realize wireless connectivity, a transceiver and antenna need to be selected.

Regarding the transceiver, it is intended to keep compatible with the legacy hardware (receiver) and software used with BG-V4.2. This sets the use of a 2.4 GHz transceiver, which is furthermore compatible with the former used *ShockBurst* (SB) protocol. This limits the pool of potential devices to the *Nordic Semiconductor*[17] family. As it is also intended to support inter-operable wireless protocols such as BLE, ANT, or ZigBee, a multi-protocol transceiver is favorable.

The Nordic nRF51822 [176] was identified to fit those requirements. It offers flexibility in terms of available protocols, which are realized in software (BLE and SB). Furthermore, it is available in a small (weaver-level chip-scale) package and optimized for low-power battery-operated devices.

The nRF51822 is a SoC. Its core is based on an ARM Cortex-M0 design. Its physical dimensions are 3.8 mm x 3.8 mm (width x height). The average active power consumption is 19 mW or 23 mW if the device is configured to transmit (output power 1 mW) or receive data (1 Mbit/s), respectively. The power consumption when the dive is in off-mode is $<3\,\mu$W.

In the BG-V4.2 a *chip antenna*[18] was used, offering a maximum antenna gain of 0.5 dB with physical dimensions of 6.5 mm x 2.2 mm x 1.0 mm (width x height x depth). In the BG-V5, instead, a PCB-antenna[19] was realized, which has a physical dimension of 25.7 mm x 7.5 mm (width x height) and offers a maximum gain of 3.3 dB.

The antenna's length and height are near to the maximum, which is limited by the largest metal-free area of the final device. This, in turn, is determined by the ECG electrodes distance of 45 mm and the batteries dimensions, which is about to be placed beneath the PCB. Given the diameter of the snap-fasteners used (10 mm), this leaves an absolute maximum of 35 mm in length. The final height of the device is determined by the sum of the battery's and antenna's height.

3.3.1.7 Power supply

To power the BG-V5, a rechargeable Li-Pol is used. This decision is based on previous experience with the BG-V4.2, where primary cells (coin cells) were used. In contrast to primary cells, Li-Pol batteries offer a compromise between power and energy density. For this reason, the maximum discharge rate with a Li-Pol is higher. Thus, the energy stored can be used more efficiently.

[17] Nordic Semiconductor, https://www.nordicsemi.com

[18] Linx ANT-2.45-CHP-x, https://linxtechnologies.com/wp/wp-content/uploads/ant-fff-chp-x.pdf

[19] Based on the *Design Note DN0007* (SWRU120B) provided by Texas Instruments, https://e2e.ti.com/cfs-file/__key/communityserver-discussions-components-files/158/swru120b.pdf

It was already discussed that the physical dimensions are mainly constrained by the ECG electrodes distance and snap-fasteners diameter (35 mm). Likewise, the final height of the device is a sum of the battery's and antenna's height. Commercially available off-the-shelf batteries, matching the physical dimension constraints are (width x height x depth):

- ICP501421PS-01[20] : 22.5 mm x 14.1 mm x 5.2 mm, 115 mA h, 3.7 V
- ICP402025PC-01[21] : 27.5 mm x 20.5 mm x 4.3 mm, 155 mA h, 3.7 V
- LPP 402025 CE[22] : 25.5 mm x 20.5 mm x 4.3 mm, 150 mA h, 3.7 V

The given details refer to the typical capacitance and the nominal voltage.

All components in BG-V5 design have been selected to operate at a minimum supply voltage of 1.8 V, except for the Bosch BNO055. This IMU, which requires a 3 V supply, but is capable of operating its input-output pins with a separate voltage domain. The 3 V domain is also needed to support the native USB functionality of the STM32L4 MCU and to drive the status LED as well as other indication components (buzzer and vibration motor).

In order to guarantee the lowest noise, regarding the analog signal-processing of the ECG, a separate low-noise analog voltage-domain is recommendable. The optimal component would thus combine a battery charger, 2 high-efficient switching-converters (step-down DC-DC convert) for the digital low- (1.8 V) and high-voltage (3.0 V) domain and 1 low-noise linear LDO regulator for the analog voltage-domain (1.8 V). No such component was identified during the design process.

The next best candidates are the *LTC3553*[23] or BQ25120A [42]. Both components are integrated battery-charge and power-management solutions, which meet all requirements but misses a second switching converter.

Otherwise, the components are virtually identical. Comparing both devices, the BQ25120A is smaller by 0.5 mm (2.5 mm x 2.5 mm (width x height)). Furthermore, quiescent current consumption is better in BQ25120A ($<1\,\mu$A compared to $12\,\mu$A). For these reasons, BQ25120A is chosen in favor. It is used in combination with a *TPS62743*[24] buck convert, which is used to provide the secondary high voltage digital power domain (3 V).

[20] Renata, ICP501421PS-01: https://www.renata.com/fileadmin/downloads/productsheets/lithium_polymer/ICP501421PS-01.pdf

[21] Renata, ICP402025PC-01: https://www.renata.com/fileadmin/downloads/productsheets/lithium_polymer/ICP402025PC-01.pdf

[22] Varta, LPP 402025 CE: https://products.varta-microbattery.com/applications/mb_data/documents/data_sheets/DS56416.PDF

[23] Linear Technology, LTC3553, https://www.analog.com/media/en/technical-documentation/data-sheets/3553fc.pdf

[24] Texas Instruments, TPS62743, http://www.ti.com/lit/ds/symlink/tps62743.pdf

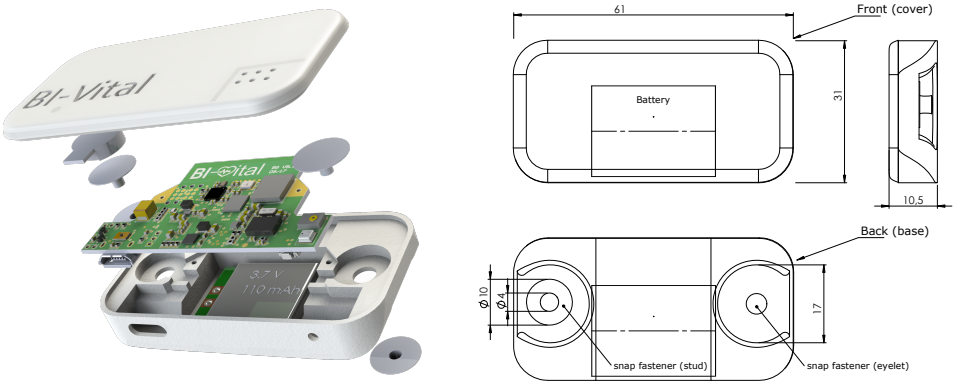


Figure 3.5: Explosion view render and assembly drawing of the BG-V5's enclosure, PCB, battery, and additional mechanical components (i.e. snap-fasteners and vibration motor).

3.3.2 Enclosure and Indication

The BG-V5's enclosure is 3D-printed using transparent polylactic acid (PLA) filament. This material offers a good trade-off between surface quality, printing speed, and price. It is furthermore compatible with the use of polyvinyl alcohol (PVA) as support material. This facilitates the post-production process as PVA can be easily dissolved in water. The support material is thus not needed to be broken out or filed off.

The dimensions of the enclosure follow the absolute minimum dimensions, which are determined by 1. the distance of the ECG-electrodes and 2. the total height of both the battery and the PCB-antenna. Its final dimension is 61 mm x 31 mm x 11 mm (width x height x depth).

The final enclosure (Figure 3.5) consists of a base (bottom) and a cover (top). At the base's bottom, the battery is inserted. The PCB is placed on top. It press-fits into the corresponding cut-out. The breaking edges at the PCB help to fasten it. Besides, optional fixing screws can be used at the same height as the ECG contact areas on the PCB. Conventional spring snap fastener (as used in the textile sector) serve as electrodes. These are pressed into the enclosure. The ECG electrodes and corresponding PCB contact areas are soldered manually.

The enclosure's cover is press-fit onto the bottom part. This is suitable for the use of prototypes because the enclose can easily be opened and closed. This allows presenting the insides of the BG-V5. Other enclosure variants are prepared, which allow a permanent fixation of the cover through snap-hook connectors. Alternatively, screws can be used.

In the base of the enclosure, there are openings for the USB-port and the reset button at the bottom. On top of the snap fasteners, extra space is reserve to mount the vibration motor and piezo buzzer.

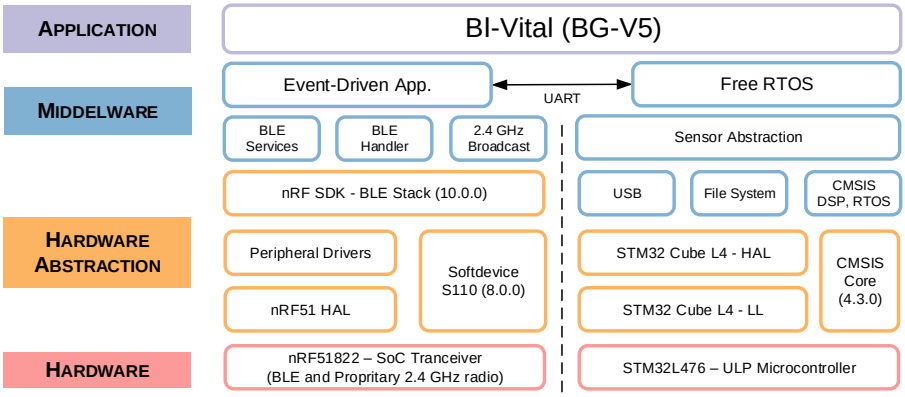


Figure 3.6: Software architecture of the BG-V5. It includes two firmware components, the nRF51 firmware for wireless communication and the STM32L4 firmware for signal acquisition and processing. Both rely on the use of *Software Development Kits* (SDK) and *Hardware Abstraction Libraries* (HAL) provided by the manufacturers. The STM32L4 firmware is based on a real-time operating system (FreeRTOS), which allows encapsulating different software components into *tasks*. The nRF51 firmware is realized as an event-driven application.

The enclosure's cover also acts as a trigger for the push-button on the PCB of the BG-V5. For this purpose, there is a nose on the inside of the cover, which is placed above and is precisely aligned towards the push-button on the PCB.

Having control of the manufacturing process of the enclosure offers full flexibility to change the shape and material in future revisions. For instance, an alternative variant is available, which offers additional attachment loops. These allow using hook and loop fasteners, e.g. to attach the BG-V5 at the wearer's wrist or ankle (for motion monitoring).

Moreover, other variants can be thought off. For instance, the enclosure could be printed using a more flexible filament that offers less tensile stiffness. This could be beneficial to avoid harming the wearer. Additionally, a fully closed enclosure could be printed that would allow under-water use.

3.3.3 Software Design

The BG-V5 software (Figure 3.6) splits up into the firmware for the wireless transceiver nRF51 (sec. 3.3.1.6) and the STM32L4 MCU (sec. 3.3.1.1). The STM32L4 is running the main application and is coordinating all access to the peripheral components, which includes sensors, power management, and USB-connectivity. In this respect, the nRF51 also acts as a peripheral (or slave) device for wireless communication. Its firmware is developed separately. It allows forwarding received messages to the STM32L4.

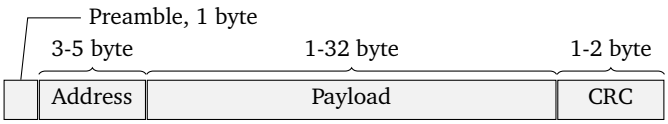


Figure 3.7: Packet structure for a message transmitted with the SB-protocol.

Both firmware components rely on the use of corresponding hardware abstraction library (HAL) or software development kit (SDK) provided by the manufacturers, i. e. Nordic Semiconductor or STMicroelectronics. The firmware development is on the middleware level (adaptations, extensions) and the application level (Figure 3.6).

The nRF51 firmware is built on top of the *nRF51 SDK*[25], and the SoftDevice *S110*[26]. The *SoftDevice* is a closed-source firmware component with interface-level access (application programming interface, API). It offers an abstraction of the BLE stack. The SDK is open-source and provides abstractions to access the radio and other peripherals.

Regarding nRF51 an interrupt-based firmware is implemented. Events can be triggered by incoming BLE messages, which are internally handled by the *SoftDevice* and conceivably are forwarded to the STM32L4, or a request from the STM32L4. A unique feature is the concurrent use of the radio as a shared resource. Incoming messages (if enabled in the configuration) are transmitted via broadcasting using the SB protocol. While BLE is still active, SB packages are transmitted in the spare time between two BLE connection events. The SB communication can either be configured manually at any time or is automatically turned off when a BLE connection is established.

The SB package format (Figure 3.7) is adapted from the BG-V4.2 and was implemented for legacy reasons to maintain compatibility with the BG-V4.2 revision's system setup (PC receiver and software). Each package consists of 21 B payload, which contains the module's ID, a package counter, the sensor data (HR, ECG, acceleration data and more) as well as a cyclic redundancy check (CRC).

In contrast to the original protocol specification, it uses only 2 address bytes (instead of 3 to 5). In addition to the proprietary CRC (as part of the payload), the protocol-specific CRC is included as well (1 B).

In each package, signals with low- and high- temporal resolution (e.g. HR and ECG) are transferred together. Thus, redundancy is high. However, the protocol's over-head is smaller compared to BLE packages (Figure 3.8). As a consequence, SB can be used for low latency, high throughput data transfer, while BLE can be used to transmit other data and configuration parameters.

The inter-MCU communication is based on a proprietary universal asynchronous receiver transmitter (UART) protocol with serial line internet protocol (SLIP)-like

[25] nRF51 SDK v10.0.0, `https://infocenter.nordicsemi.com/topic/com.nordic.infocenter.sdk51.v10.0.0/index.html`

[26] S110 *SoftDevice* v8.0.0, `https://infocenter.nordicsemi.com/pdf/S110_SDS_v2.0.pdf`

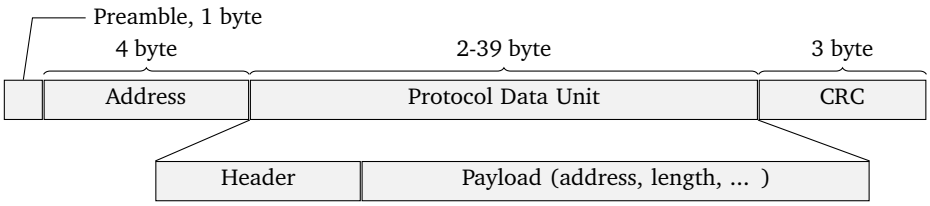


Figure 3.8: Simplified view on the packet structure for a message transmitted with the BLE-protocol. The payload (or Protocol Data Unit, PDU) can be of different types (advertising or data). It again contains auxiliary fields like address, payload length, or a checksum.

encoding. The first byte in the message indicates the message's type (4 different types encoded in the first 2 bits), and its content (6 content fields, one-hot-encoded in the last 6 bits). The content has a fixed order and pre-defined length. In this way, a compact protocol with flexible payload is realized[27].

The main application is running on the STM32L4. It is an event-driven application that is built on top of the *FreeRTOS*[28]. By default, all tasks (i.e. threads) are in a blocked state and wait indefinitely for a notification or message (event). All events are triggered by interrupts, which either occur due to user interaction (motion, button) or are triggered by a timer or other peripheral (e.g. the ADC). The application's functionality is encapsulated in different tasks:

The *sensor-task* is responsible for enabling or disabling the sensors. In this respect, it takes control of the hardware-timers associated with a sensor, and it samples and forwards the raw data. The task also triggers the transceiver- or flash-task. These tasks are responsible for transmitting or storing the data, respectively.

Auxiliary features are handled by additional, low-priority tasks: The *button-task* counts short and long clicks of the user-button. The *indication-task* is responsible for updating the LED, the buzzer, and the vibration motor. the *power-task* controls the battery charging, the power-status, and system voltage-levels. The USB-connectivity of the BG-V5 is used for debugging purposes (UART-over-USB terminal) and file-transfer. Both are handled in the *terminal-task*.

Besides those tasks, a *default-task* is implemented. It takes control of the BG-V5's state-machine and forwards events to the other tasks. This is also where forwarded events from the other tasks are collected and passed on again. As an example, connecting the USB (interrupt forwarded to the default-task) will change the systems state to charging (power-task) and stop all sensors (sensor-task).

[27] In SLIP encoding, each message starts and ends with two pre-defined characters (*START* or *STOP* character). If *START* and *STOP* are part of the message they are replaced by an escape character (*ESC*) followed by a corresponding replacement character for *START* (*ESC-START*) or STOP (*ESC-STOP*).

[28] FreeRTOS Kernel V10.0.1, ©2017 Amazon.com, `http://www.FreeRTOS.org`

3.4 Design Evaluation

The BG-V5 was developed as part of this thesis to serve as a WBS prototype platform for future work. In the following, the fundamental features of the design are evaluated. This includes a test and verification of the ECG circuit (sec. 3.4.1), a comparison of the inertial sensors (sec. 3.4.2), and wireless communication (sec. 3.4.3). It is concluded with the creation of a power profile for typical applications (sec. 3.4.4), which is intended to provide guidelines towards run-time estimations.

3.4.1 Electrocardiogram

The ECG circuit used in the BG-V4.2 has already proven its effectiveness in practical experiments. The schematic design for the ECG's filter and amplification used in BG-V5 is based on that very circuit.

However, with the new layout, also adaptations of the PCB design were carried out, and some components were exchanged. Moreover, the electrode's connection is different, as the housing was changed, which is now based on a custom design. The simulation results already showed that both systems are equivalent (sec. 3.3.1.4). To verify this, both systems are empirically compared to each other (Figure 3.9).

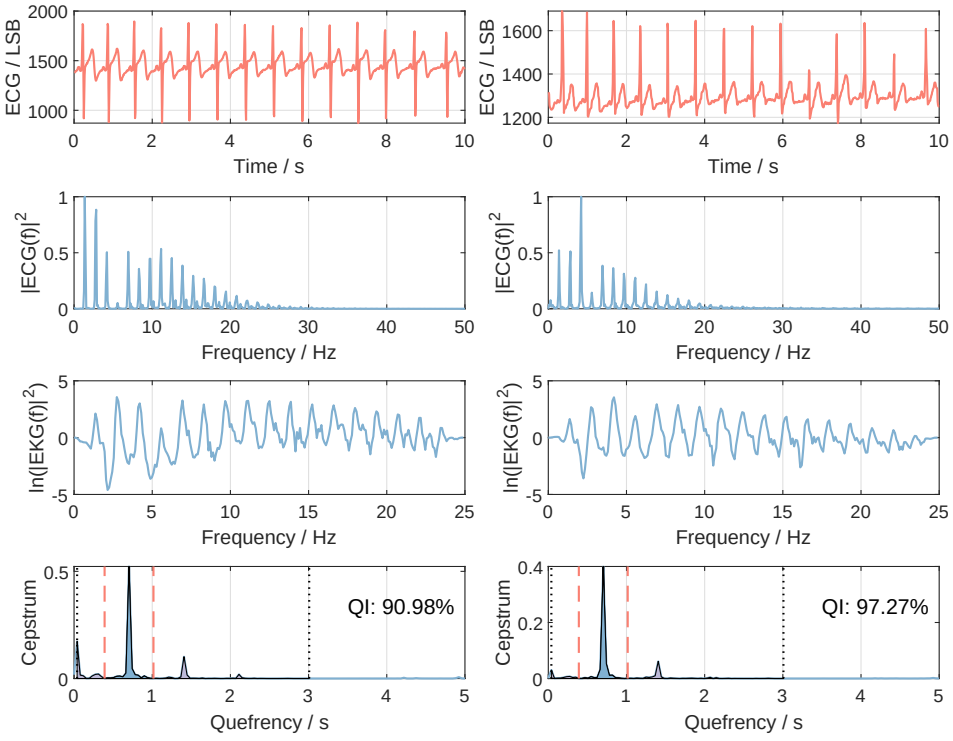
For the comparison of the new design, two ECGs were recorded by one participant using two chest straps at the same time. Inspecting the result, firstly, it is noticed that the ECG morphology is different. This can be explained by the slightly different placement of the electrodes, which are positioned one above the other. Thus different ECG leads are visible.

The spectrum analysis, however, reveals that both recorded ECGs compare. The spectrum of both systems is nearly alike. This is expected due to the fact that both systems share the same circuit design. Also, an unwanted damping around 7 Hz is found in both systems. It is less distinctive in the BG-V5. In return, considering the BG-V5, a ringing is observed around 5 Hz.

Comparing the ECG's quality using the method described in [53], it can be shown that the spectrum contains most energy in the relevant frequency domain. It can thus be concluded that both devices provide accurate ECG readings.

3.4.2 Inertial Sensors

From previous and related work, acceleration is identified as the most important modality for different WBS applications. For some applications, e.g. activity recognition [84, 131] a low-resolution and high-noise, but low-power configuration is preferable (same results with lower power consumption). With the BNO055 integrated into the BG-V5, moreover, the absolute orientation is available. This is interesting, e.g. to experiment with precise indoor navigation [290].



(a) ECG recorded with BG-V4.2 **(b)** ECG recorded with BG-V5

Figure 3.9: Comparison of 2 simultaneously recorded ECGs with separate chest straps in a resting condition. In the first row, the raw ECG data is depicted, followed by normalized and log amplitude spectrum from fast Fourier transform (FFT). The last row shows the ECG's Cepstrum (logarithmic frequency spectrum) and the corresponding quality index as proposed in [53]

The direct comparison of the accelerometer readings of both sensors reveals that they agree in terms of absolute values (Figure 3.10). Differences are in the measurement resolution. For the final application, the designer can choose between both accelerometers. It is thus possible to trade-off resolution (or modality) and power consumption.

3.4.3 Wireless communication

The transceiver used in BG-V5 implements a physical 2.4 GHz radio and comes with a software-defined protocol stack (nRF51). This has the advantage that common standard interfaces (Bluetooth, ANT) are supported. However, the 2.4 GHz band is not optimal

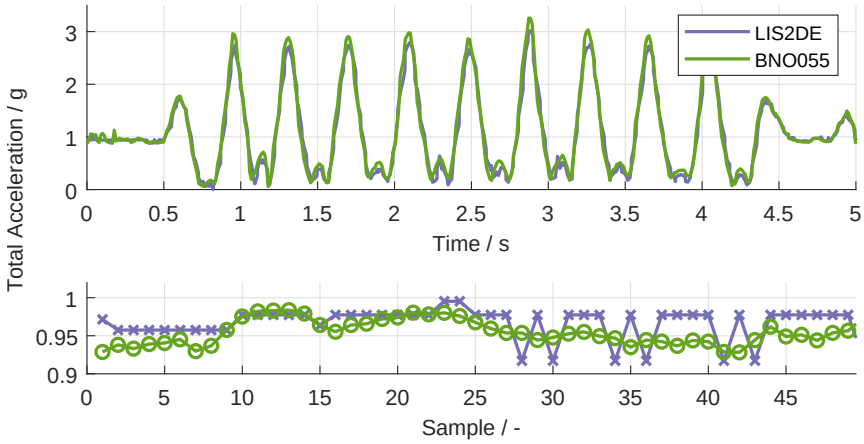


Figure 3.10: Comparison of the low-power accelerometer LIS2DE and the high-accuracy IMU BNO055 (acceleration only). Both sensors were moved and sampled in parallel (upper image). The agreement of the absolute readings and the different resolutions of the two sensors can be seen (lower image).

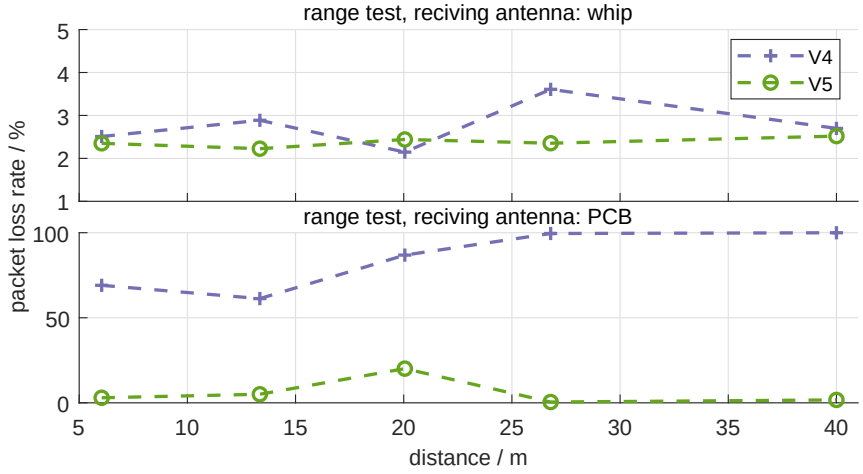


Figure 3.11: Comparison of wireless communication in BG-V4.2 and BG-V5. To compare both, a basic range test was carried out in a sports hall with different receivers (whip or PCB antenna). In each test, 100 packages were sent per second. The packet loss rate is calculated by counting gaps in the package counter.

in terms of range. Consequently, the antenna in the BG-V5 was explicitly granted more space in order to achieve a high range still.

To test the antenna's and transceiver's performance, a simple range test was carried out. Therefore, 2 different receivers were used at the receiving side. That is a small PCB antenna (0 dB, *nRF51 Dongle*[29]) and a 19 cm whip antenna (4 dB, *Linx ANT-2.4-OC-LG-SMA*[30]). The senders were placed at the experimenter's chest, who has positioned himself at different distances to the receiver (6 m to 40 m, line-of-sight conditions). At each distance, 30 s of data were recorded, where the devices were configured to transmit an incremental package counter in the payload. Based on this counter, the package loss rate at the receiving side is calculated in % (by counting gaps in the incremental package counter).

As a result, it can be seen that for both transmitting devices, the packet loss rate is low up to a range of approximately 40 m. That is if the whip antenna is used at the receiving side only. If the gain at the receiving side is lower, only with the BG-V5, the transmission rate remains acceptable.

The direct comparison of the BG-V5 and the BG-V4.2 shows the effectiveness of the changed antenna selection (Figure 3.11). The improved antenna performance becomes critical when the receiving antenna is small. As a consequence, the BG-V5 can also be used with conventional receivers (e.g. a BLE antenna embedded into a smartphone or laptop). That relaxes the constraint to use custom hardware for wireless communication.

3.4.4 Power profile

Average power consumption is crucial in WBS design. It must be kept at a minimum in order to allow long run-times (autonomy of the WBS). In the following, a power profile of the BG-V5 for different use-cases is given (Table 3.3).

All measurement results presented were made using a *Source Measurement Unit*[31], which allows acting as a power source and measurement device simultaneously. The measurement results are averaged over a period of 60 s. To identify the individual components' power consumption (sensors, transceivers, signal processing), they were switched off one after the other. Based on the differences between the separate measurements, the power consumption of the respective isolated component is calculated.

Two different use-cases are considered. First, the use-case as a high-resolution diagnostic data-logger with a sample rate of 100 Hz is considered. In this case, the same behavior as in the predecessor version (BG-V4.2) is replicated. The packet size of the wireless data transmission is 22 B. As a comparison, also, the application of the BG-V5 as a simplistic BLE based HRM is considered. In this mode only ECG and QRS

[29] nRF51 Dongle, https://infocenter.nordicsemi.com/pdf/nRF51_Dongle_UG_v1.0.pdf
[30] https://linxtechnologies.com/wp/wp-content/uploads/ant-2.4-oc-lg-fff.pdf
[31] Source Measurement Unit, Keithley 2450, https://www.tek.com/keithley-source-measure-units/keithley-smu-2400-series-sourcemeter

Table 3.3: Power profile of the BG-V5 for typical application as a data-logger (100 Hz) or BLE HRM (advertising only). Average (avg) and standard deviation (std) for each scenario are given. Power consumption for each component is calculated from differential measurements.

Scenario or component	**power consumption / mW**	
	avg.	**std.**
BLE HR-Monitor	10.936	0.812
Data-Logger (100 Hz)	13.185	0.819
Environmental sensor (temperature)	0.028	-
QRS-Detection	0.016	-
ECG (sampling, ADC and DAC)	0.199	-
ECG (dissipation, calculated)	0.116	-
Low-power acceleration sensor	0.465	-
Transceiver (2200 B/s)	2.609	-
Micro-controller	5.168	-

detection are active. The transceiver is used in advertising mode only (advertising interval is 1 s).

The average power consumption of the BG-V4.2, used as a data-logger for HR, acceleration, and temperature, is 19.87 ± 0.21 mW. In comparison, the average power consumption of the BG-V5 is 13.18 ± 0.81 mW. In this example, 19.8 % is accountable for the wireless transmission of the data. A further 6.3 % can be assigned to the sensors. The remaining power consumption is due to the operation of the MCU and static power losses. The average power consumption as a HRM is found to be 10.9362 ± 0.8120 mW.

The comparison of the HRM and data-logger scenario highlights the dominant power consumption of the MCU. In conclusion, it can be seen that further optimization of the MCU code is required to reduce the power consumption of the BG-V5. In the current version, the MCU (STM32L476) is running clocked with 16 MHz in active mode, and switches to sleep mode, which turns off CPU only. Specified current consumption in active mode is 2.150 mA and 0.671 mA in sleep mode [223, p. 25]. Optimizing code to use, e.g. STOP modes, would allow better power savings, as current consumption can be as low as 0.007 mA or 0.001 mA in STOP1 or STOP2 mode, respectively.

Considering these measurements, the BG-V5's operating time can be calculated. The energy capacity of the Li-Pols used for the BG-V5 is in the range of 370 mW h to 555 mW h (sec. 3.3.1.7, capacity 100 mA h to 150 mA h, nominal voltage 3.7 V). Thus, the estimated operating time is in the range of 33.8 h to 50.3 h for the HRM scenario, 28.1 h to 41.7 h for the data-logger scenario, or 8.5 h to 12.6 h for the data-logger scenario using the high-accuracy IMU.

In summary, the power consumption for the data-logger scenario is reduced by 33.7 %, comparing the BG-V5 to the BG-V4.2. Moreover, there is potential to further reduce the power consumption by optimizing the low-power management of the controlling MCU.

The change from coin cell to Li-Pol battery also increases the operating time, as the energy storage can now be better used. The theoretically estimated operating times thus correspond to the real actual times. Further optimization of the power consumption can thus be regarded as a secondary goal. Since the current version of the BG-V5 already offers an operating time of more than one day, for most scenarios, this creates space to concentrate more on the implementation of future applications.

3.5 Selected Application: Wearable ECG Classification

The BG-V5 is designed to fit a wide range of application use-cases. For instance, it is possible to select the low-power accelerometer for motion analysis [285] or to select the high-resolution absolute orientation IMU for indoor navigation [290]. Likewise, the ECG sample-rate could be changed depending on the accuracy needed by a specific application.

As a consequence, the respective application can either benefit from lower power consumption, which furthermore goes with increased recording time, or prefer a higher temporal resolution instead (with the drawback of having a higher power consumption and lower recording time). Here it can be seen that power consumption, storage capacity, and autonomy (run-time) mutually influence each other. A trade-off has to be found concerning these variables. For instance, this applies to the selection of a sensor or its sample rate.

Moreover, the architecture of an application is affected by this very same trade-off as well. More precisely, it especially exists for computational-expensive applications and the use of *embedded algorithms*. This is highlighted in the work of Rault et al. [189], where different scenarios are identified:

1. Sampling, and transmission of the raw sensor data.
2. Sampling, feature extraction, and transmission of *compressed* data.
3. Sampling, classification, and transmission of extracted information.

The benefit of embedded signal-processing or feature-extraction becomes apparent with the example of embedded QRS-detection using the Pan-Tomkins algorithm (sec. 2.3.2). From the power profile of the BG-V5 (sec. 3.4.4), it can readily be seen that forgoing raw ECG data transmission saves energy consumed by the transceiver. At the same, the increased power consumption of the MCU (due to the QRS-detection) is negligible. In total, power consumption is decreased by approximately 17.1 %. Indeed, a similar effect is also described by Rault et al. [189], using the Shimmer 3 platform (sec. 3.1.2.2) for an activity recognition scenario using DT based on accelerometer data (HAR, sec. 5.2). There, power consumption was decreased by 63.5 %.

Yet, power consumption is only one argument for the implementation of *wearable algorithms*. Another reason is that WBSs are primarily used to record private (and possibly sensible) data of their wearer. It is thus easily argued that it is beneficial to keep the data private and avoid any transmission. This is a topic that is already broadly discussed in the literature [166]. A real-world example where this can be learned from is the application of embedded cardiac arrhythmia detection for WBS equipped with an ECG based heart rate sensor [202].

In this section, the example of an embedded classification for cardiac arrhythmia detection is demonstrated as an excursus. In this way, the necessity of having a power-efficient yet high-performance WBS like the BG-V5 is highlighted. In the following, specific characteristics of the ECG indicating cardiac arrhythmia, public databases containing exemplary data, and related work on classifying cardiac arrhythmia are presented. Subsequently, a method is selected, re-implemented for the BG-V5, and evaluated in terms of latency and power consumption.

3.5.1 Problem statement: cardiac arrhythmia

In HRM, the raw ECG sensor data is *compressed*, leaving heart rate information only. This can be done in multiple ways [82]:

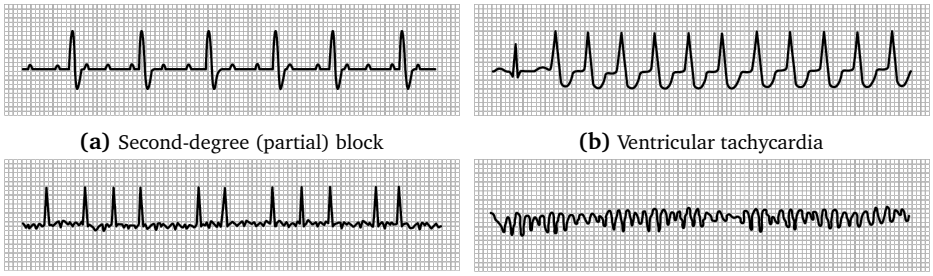
1. Simplistic peak detection using amplitude only.
2. Advanced peak detection using multiple thresholds.
3. Peak detection combined with pre-processing (e.g. filtering).
4. Approximation methods, like template matching or wavelet transformation.

An example of the 3. method, the Pan-Tompkins algorithm that is also used in BG-V5, has already been addressed in (sec. 2.3.2). An overview of these methods and their qualitative computational complexity can be found in [82].

All QRS detection algorithms aim at extracting the position of the QRS complex and thus derive heart rate from the raw ECG signal. Indeed, this *compressed* information can readily be used to extract more than heart rate alone. It can be seen as a first step towards detecting cardiac arrhythmia that become apparent in the form of abnormal heart rhythms.

A common example is atrial fibrillation (AF). Typically, the signal from the sinus node leads to a synchronized contraction of the atrial muscles, followed by a contraction of the ventricular muscles (sec. 2.3.2). With AF and other abnormal rhythms, this ever same process is disturbed. It can be analyzed inspecting the apparent (distorted) changes in the characteristic of the ECG.

With the example of AF, electrical spiral waves, chaotically propagating in the atrial occur. This fact leads to a distinguishable set of methodologies to detect AF, either by inspecting *atrial activity* (chaotic electrical activity in the atrial) or by inspecting irregular *ventricular response* (contractions of the ventricular muscles following the chaotic electrical activity). [153]



(a) Second-degree (partial) block

(b) Ventricular tachycardia

(c) Atrial fibrillation

(d) Ventricular fibrillation

Figure 3.12: Cardiac arrhythmias show certain distortions or missing parts compared to the normal waves and peaks (P, Q, R, S, T). E.g., in (a) not every P wave is followed by a QRS complex, in (b) the shape of the QRS complex is distorted, (c) is irregular and shows electrical fluctuations (f-waves), and in (d) no normal electrical activity can be seen at all. Illustrations are based on OpenStax, 19.2 Cardiac Muscle and Electrical Activity. OpenStax CNX. 2. Mai 2019 `http://cnx.org/contents/302812e9-2d2d-4e44-8075-4bb75db53f36@8`, CC BY 4.0

With the *ventricular response* analysis, QRS detection is a first step to identify AF episodes, characterized by the irregular rhythms of the heart rate. An example can be found with Poincaré-, or Lorenz plot analysis [204]. They can be used to visualize and classify cardiac arrhythmia that become visible by the irregular distances of successive heartbeat intervals (RR-intervals, sec. 2.3.2). Yet, such methods are discarding any other abnormalities besides irregular QRS complexes. Therefore, these could be considered a *lossy compression* technique, preserving only a fraction of the raw ECG signal, where other significant information is lost.

With the *atrial activity* analysis, morphological features or generalized stochastic metrics are extracted from the ECG. As an example of AF detection, atrial fibrillation waves (*F-waves*) are of interest, which become visible as a sinusoidal like waveform in the ECG.

All types of cardiac arrhythmia can be detected based on the ECG (Figure 3.12), some of which require multiple-lead analysis to allow for spatial analysis of the ECG. The procedures can be based on using certain thresholds, template matching, or wavelet transformation, just like it is done with QRS detection. However, in the case of cardiac arrhythmia detection, the ECG analysis is more challenging since fluctuations or abnormalities can become apparent in merely any patterns. Therefore, today, often ML techniques are used for such classifications tasks.

Cardiac arrhythmia classification from ECG gained massive attention with a Physionet challenge exposed in 2017 [66][32]. The task was to classify *Atrial Fibrillation* against other, normal (healthy) and noisy rhythms from 1-lead raw ECG signals recorded with

[32]AF Classification from a Short Single Lead ECG Recording - The PhysioNet Computing in Cardiology Challenge 2017, `https://www.physionet.org/content/challenge-2017/1.0.0/`

a WBS. Among the submitted solutions, the top-4 solutions produced scores in the range of $F_1 = 0.825$ to 0.831[33]. This is remarkable since these results are close to the performance of cardiologists.

The latest work focus on the use of recurrent or convolutional ANNs. For instance, Kamaleswaran et al. [122] presented a 13-layer convolutional neural network (CNN) to classify *Atrial Fibrillation* using the PhysioNet 2017 challenge's data set. With this approach, they achieve the same accuracy as the original winners of the challenge. The success of CNN generalizes for multiple problems related to ECG classification [7]. For instance, Strodthoff et al. [226] state that their work "outperforms state-of-the-art approaches and reaches the performance level of human cardiologists for the detection of myocardial infarction" on the *PTB Diagnostic ECG Database*[34].

From the engineering perspective, it is also to be highlighted that these solutions do not require elaborate pre-processing steps but are suitable to process the raw data directly. As stated by Kamaleswaran et al. [122], that makes these solutions more applicable and robust. Remarkably, they find that the accuracy of their approach can be reproduced using 2.5 ECG segments only. This is in agreement with the findings presented by Kachuee et al. [120]. In their work, a generalized network is presented that allows detecting different arrhythmia on the level of single heartbeats using *PhysioNet MIT-BIH Arrhythmia Database*[35], as well as myocardial infections using the *PhysioNet PTB ECG Database*. Their solution allows to "accurately classify five different arrhythmias in accordance with the AAMI EC57 standard" [120] (93.4 % arrhythmia).

If it is intended to port an ANN for use by an ES (MCU), its size (i.e. memory requirements) is a crucial constraint. The model's size is also a reasonable estimate for its computational complexity because, in a first approximation, these can be assumed proportional to each other. In this respect, the model's size has a direct influence on the latency and energy consumption.

Regarding state-of-the-art MCUs, Zhang et al. [267] defined 3 classes of neural network complexities that are either small (limit of 80 kB, 6 MOps), medium (limit of 200 kB, 20 MOps) or large (limit of 500 kB, 80 MOps). Zhang et al. [267] compare these different sized models, which are either utilizing fully-connected, convolutional, or recurrent layers. This comparison is in terms of the accuracy for a key-word spotting task (classify audio streams and detect key-words, e.g. "Yes", "No", "Stop"). The STM32L476 used in the BG-V5 (sec. 3.3.1.1), is part of this comparison.

As a result of the work of Zhang et al. [267], it can be concluded that the STM32L476 is capable of processing small to medium-sized models. Larger models could be implemented as well. However, this would require to use external flash-memory in order to store the model's weights. Nevertheless, given the use-case considered by Zhang et al.

[33] The F1-score is a combined measure of both, sensitivity (or recall) and precision of a binary classification. It is defined as $F_1 = 2 \cdot \text{precision} \cdot \text{recall} / \text{precision} + \text{recall}$. In the PhysioNet challenge the F_1 is the mean of the F_1 scores for all classes except noisy, which was excluded.

[34] PhysioNet PTB Diagnostic ECG Database, `https://physionet.org/content/ptbdb/1.0.0/`

[35] PhysioNet MIT-BIH Arrhythmia Database, `https://physionet.org/content/mitdb/1.0.0/`

[267], small- to medium-sized models already resulted in an accuracy of up to 94.9 %. This is close to the best results found with larger models (95.4 %).

Likewise, for arrhythmia classification, it can be seen that an *optimal* region for the complexities of beat-level classification models exists [128]. This complexity is in the region of 1000 to 100 000 trainable parameters, which is a result of a benchmark comparing F1-scores against trainable parameters in CNNs. In comparison, the complexity of methods using longer episodes of the ECG is approximately larger by 1 to 2 orders of magnitude. For instance, re-implementing the architecture presented by [122] (input of 60 s of raw ECG data points sampled at 300 Hz), it is found that the model consists of 5 491 136 trainable parameters.

For the subsequent analysis, the model presented by [120] is chosen because it perfectly falls into the complexity *sweet-spot* of beat-wise ECG classification (*atrial activation*) having 55 013 trainable parameters. The 13-layer CNN architecture is similar to other presented methods and showed excellent performance in arrhythmia detection [120][36].

3.5.2 Latency and power consumption

The approach presented in [120], is stacked on top of a QRS-detection, which is used to provide beat-synchronized inputs to the CNN classifier. In this respect, the input of the classifier is a normalized ECG signal with a length of 1.5 s. The network architecture consists of multiple convolutional layers with 32 filters of size 5 (Figure 3.13).

The output of this first layer is forwarded to a block of a convolutional layer followed by rectified linear unit activation and another convolutional layer. The output of this block and its former input are then summed up and again forwarded to a rectified linear unit activation and pooling layer using the maximal function. This block is repeated 5 times before its output is passed to a 3-staged fully-connected layer. The output is finally forwarded to a softmax activation layer.

The final model (Figure 3.13, [120]) is translated into C-code using the *X-CUBE-AI*[37] expansion pack provided by STMicroelectronics. It allows analyzing memory footprints and estimating the count of multiply-accumulate operations needed to carry out the inference of a given ANN model.

In the following analysis, the fitted model is directly tested on the BG-V5 MCU (STM32L476). Power and energy consumption are calculated on the values provided in [223, p. 25] (80 MHz - 19.08 mW; 48 MHz - 11.52 mW; 16 MHz - 3.33 mW). The clock cycles are obtained from the internal clock cycle registers of the MCU. These are used to calculate the duration of each inference, assuming a perfectly stable clock. The inference is tested with different clock settings. All results are averaged for 5 runs of

[36] An example implementation can for instance be found here: `https://github.com/CVxTz/ECG_Heartbeat_Classification`

[37] STM32Cube.AI, `https://www.st.com/resource/en/data_brief/x-cube-ai.pdf`

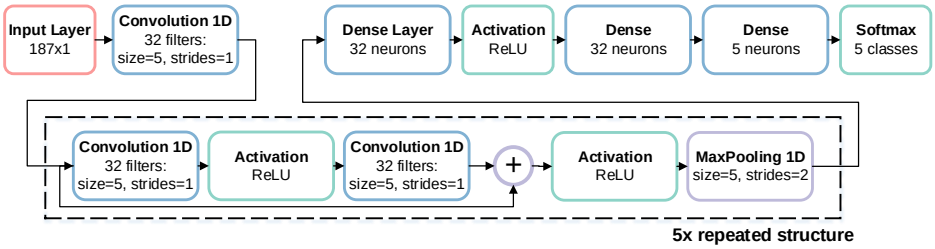


Figure 3.13: The CNN architecture, as proposed by Kachuee et al. [120], to classify cardiac arrhythmia. The classification allows detecting arrhythmia according to the EC57 standard with an accuracy of 93.4 %. It consists of a residual block of 2 convolutional layers, which is repeated 5 times. Its output is forwarded to 2 dense layers and is finally activated by a softmax layer. The model contains 50 853 trainable parameters using 32-bit floating-point representation.

blocks of 16 inferences. As a result, memory, latency, and energy requirements can be analyzed.

The number of trainable parameters in the model presented by [120] is 55 013 and requires 3 614 635 multiply-accumulate operations for a single inference. The static memory needed for the model is 220 052 B (32-bit float values) of read-only memory (ROM) and additionally 71 040 B in the random-access memory (RAM). It should be noted that by using fixed-point representations, these memory requirements can be reduced by a factor of approximately 4. This, however, comes with the risk of loss in terms of accuracy, which in turn must be examined separately. It is thus not part of this excursus.

The average number of clock-cycles, needed to process a block of 16 inferences reaches from 35 667 037.2 ± 103.0 at 80 MHz, over 32 079 040.0 ± 124.8 at 48 MHz to 32 079 050.8 ± 119.0 at 16 MHz. Noticeably, the clock cycles needed with 80 MHz are higher (Figure 3.14). This can be explained by additional wait-cycles needed to load data from ROM. Still, the inference's latency lower bound of 27.9 ms is found with the clock set to 80 MHz. With the lowest clock rate of 16 MHz latency is 125.3 ms.

Taking the latency and the static power consumption of the MCU (given a certain clock-rate), the energy per inference can be calculated. In summary, the energy required for a single inference is 476.7 µJ on average. Due to the increased number of clock cycles, energy consumption is highest when the MCU is running at 80 MHz (531.7 µJ).

Moreover, the MCU's efficiency is maximal if it is running with a reduced internal voltage (RUN2 mode with approximately 0.21 nJ per cycle, compared to RUN1 mode with approximately 0.24 nJ per cycle). This mode, however, allows clock-rates smaller than 26 MHz only. Consequently, the minimal energy consumption per inference is found with the MCU utilizing RUN2 mode. Here, this is the case for inference running at 16 MHz. This is the default setting of the BG-V5.

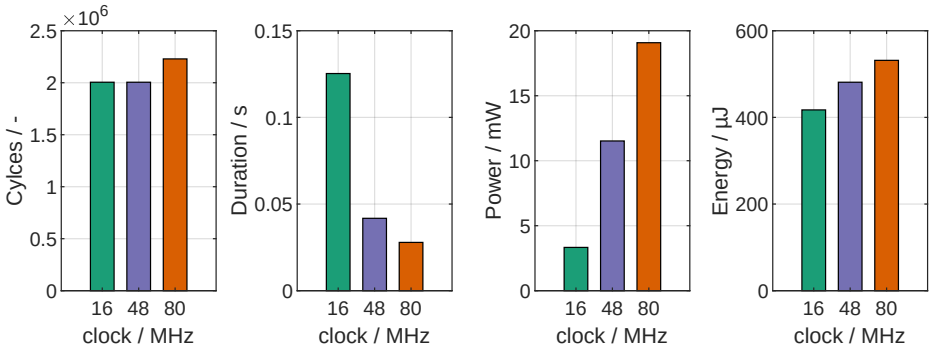


Figure 3.14: Characterization of the ANN-based inference for ECG classification in terms of clock cycles, duration (latency), the average power consumption of the MCU, and finally, the estimated energy per single inference.

3.5.3 Summary and Discussion

Within this excursus, an embedded ML application example utilizing the BG-V5 was shown. It goes beyond classical state-of-the-art heart rate detection but allows for embedded detection of cardiac arrhythmia on the level of distinctive heartbeats. It outlines the future application of WBS like the BG-V5 using advanced embedded signal processing techniques, here ANN-based inference, that allow for an online-detection of critical events.

Having real-time constrains on the ECG analysis, it can be seen that assuming a typical heart rate in the range of 60 bpm to 200 bpm, each inference must be completed within 1000 ms to 300 ms. The inference on the BG-V5 (STM32L476) is completed within 28 ms to 125 ms on average, which matches the time constraint. Thus, real-time applications are enabled. Furthermore, the additional power consumption of approximately 3.3 mW (assuming repeated ongoing inference), still allows run-times above 24 hours (comparable to that of the data-logger scenario, sec. 3.4.4).

Still, the detection presented here is on the beat level only. For future applications, it would be interesting to examine the feasibility of analyzing longer time-series online. By this means, more complex arrhythmia could be detected. However, this increases the complexity of the models by 1 to 2 orders of magnitude (since the time-series changes by this order of magnitude). Thus, memory and latency constraints are violated. In this regard, the question arises if the extracted beat level information (as it was demonstrated here) could itself be forwarded and used in a subsequent classification step. In this step, additional information, e.g. the RR-intervals (distance between two successive QRS-complexes) could be added.

The use of ANN in this manner, i.e. to pre-process or *compress* raw data is also interesting for other WBS applications. A related example, i.e. the compression of

motion data by means of ANN-based auto-encoders, was examined in student's work [291] and showed promising results on which future work can be based.

For the future design of WBS and WSN in general, specialized hardware components, e.g. CNN accelerators could make a significant contribution. An initial approach towards these direction can be found in [145], who are working on accelerators for approximate feature-based methods. Also, hardware accelerators for specialized applications like key-word spotting (speech recognition) were presented recently. An example is found with the *Syntiant NDP100*[38]. These components allow building heterogeneous architectures using MCU combined with a programmable accelerator. Indeed, research results on this topic were already presented and demonstrated offloading an STM32L476, which is also used in BG-V5. In this way, they are boosting efficiency for matrix multiplications, which are heavily used in CNN applications. The energy efficiency is increased by 2 orders of magnitude from approximately 1 GOPS/W to 100 GOPS/W.

Optimizing energy consumption in this way paves the path for autonomous real-time classification of the ECG. Taking the results here (476.7 µJ per classification) and assuming further advantages in power-efficiency (approximately 2 orders of magnitude) as was just discussed, an online classification system with an energy consumption of approximately 5 µJ is realistic. The heart itself has an energy turn-over of approximately 1.6 J per beat [147][39]. By taking both numbers into account, it can be seen that harvesting about just one-millionth (0.000 312 %) of the heart's energy turnover would be sufficient to operate an online ECG classification system permanently.

3.6 Conclusion and Future Directions

The BG-V5 presented in this chapter is a compact, highly-integrated, and scalable WBS (Figure 3.15, Table 3.4). It can be used as a low-power heart rate monitor, high-frequency data-logger, or experimental platform for advanced wearable algorithms (e.g. embedded CNN for ECG analysis or HAR). Its total dimensions are 61 mm x 31 mm x 11 mm (width x height x depth), it weighs less than 16 g and has a typical power consumption of 14 mW (data-logger scenario).

The evaluation of the BG-V5 has only shown its basic functionality, with the primary intention to compare the BG-V5 against its predecessor. It could be shown that both modules provide comparable readings and are interchangeable in this regard. Also, advantages of the BG-V5 were highlighted, namely the standardized wireless communication interface using BLE or the availability of a full IMU.

Moreover, different secondary requirements were improved. That, for instance, includes maintainability, as the enclosure can be manufactured in-house. Also, the

[38] Syntiant Corp., NDP100 Neural Decision Processor™, `https://www.syntiant.com/ndp100`

[39] Here the cardiac work is estimated based on mechanical work, chemical energy conversion efficiency and heat dissipation.

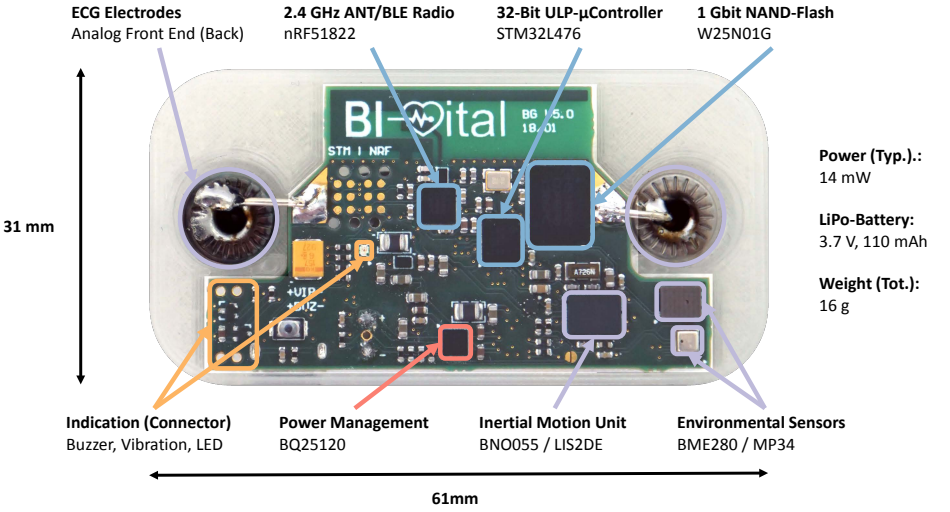


Figure 3.15: Annotated photograph of the BG-V5 with outer dimensions. The electronic components of the ECG circuit and the *LIS2DE* accelerometer are located on the back of the PCB and are thus not visible.

BG-V5 has USB-connectivity, which allows reading out the internal flash in future software revisions. That facilitates the work with the module, as no expert is needed to read out the module.

Regarding the design of the BG-V5, it can be considered a scalable research platform suitable for a broad range of body sensor applications. Its design was carried out with priority to ultra-low power consumption in mind but offering capabilities for high-performance applications as well. An example is the integration of the low-power accelerometer and the advanced, high-precision IMU. Depending on the specific application and power constraints, they can be used in exchange.

Scalability is likewise found with the MCU, which offers spare computational capacities. These can be used for advanced wearable algorithms, e.g. machine learning inference, which was already touched in [285][40] or in the excursus on embedded ECG classification (sec. 3.5). Another example of the computational capacity of the STM32L4 can be found with [267], who demonstrated inference for key-word spotting utilizing deep ANN.

Currently, 3 MCUs are part of the system's design. That is 1 MCU based on the ARM Cortex-M4 architecture (STM32L4) and additionally 2 MCUs, which are based on the ARM Cortex-M0 architecture (nRF51, BNO055). The latter, however, are not used

[40] In that work ML was applied for inference on the MCU utilizing DT or SVM based on data taken from [65].

for signal processing nor to control the application flow. They belong to the nRF51 transceiver and the BNO055 IMU, both as part of their SoC-architecture.

The nRF51 is freely programmable, and its feature set can be extended on the basis of the software-defined protocol stack. Thus, the design can be adapted to other, new wireless protocols. In contrast, the BNO055 is not freely programmable (although firmware updates that are provided by the manufacturer can be uploaded to the chip), but embeds a software motion-processing unit. Both off-loads the main MCU from handling protocol specifics (wireless communication) or sensor fusion tasks. This is beneficial since it makes room for advanced, computationally intensive applications.

Here, it must be mentioned, that in future WBS designs, even the computational aspect could be *outsourced* by integrating specialized hardware accelerators to increase inference and furthermore boost power-efficiency (sec. 3.5). On the downside of integrating more and more accelerators, the system's overall complexity increases.

The latest developments focus on the integration of multiple devices on a single chip (SoC). A first example is found with the nRF51 transceiver used for the design of the BG-V5. It combines a radio and a freely programmable MCU (Cortex-M0). Other examples are the *STM32WB*[41] or *CC2652R*[42], to name a few. In contrast to the nRF51, both offer a co-processor design similar to the one presented here (Cortex-M4 and Cortex-M0), but on a single chip (big-little architecture). Similarly, smart-sensor hubs are on the rise (e.g. *BMF055*[43] or *KX23H-1035*[44]) combining an MCU and a MEMS sensor on a single chip.

It is becoming apparent that with upcoming MCU generations, co-processors, radios, MEMS sensors, or accelerators (e.g. for ANN-based inference), will most probably become default peripheral components just like ADCs, DACs, hardware timers, and others are already today. These architectures are certainly interesting for the design of upcoming WSN generations and WBS in particular because they allow to further reduce cost, size, and architectural complexity all at the same time.

[41] STMicroelectronics, STM32WB series, `https://www.st.com/en/microcontrollers-microprocessors/stm32wb-series.html`

[42] Texas Instruments, CC2652R, `http://www.ti.com/product/CC2652R`

[43] Bosch Sensortec, BMF055, `https://www.bosch-sensortec.com/bst/products/all_products/bmf055`

[44] Kionix, KX23H-1035, `https://www.kionix.com/product/KX23H-1035`

Table 3.4: Summary of the BG-V5 components and characteristics. Power consumption values are to be understood as limits (<) or typical value for the data logging application (100 Hz.

Domain	Material / Component	Dimension (w x h x t) / mm	Weight	
PCB	-	26.5 to 49.5 x 28.0 x 0.9	3.3 g	
Battery	LPP 402025 CE	25.5 x 20.5 x 4.3	3.9 g	
Enclosure	PLA filament	31 x 61 x 11	8.6 g	
Total			<16.1 g	
Information processing, communication, and storage			**Interface**	**Power**
System MCU	ST STM32L476	Cortex M4 80 MHz, 512 kB ROM, 128 kB SRAM (FPU, ADC, DAC, I2C, QSPI, PDM, USB)		<40 mW
Wireless	Nordic nRF51822	2.4 GHz transceiver (SB, BLE)	UART	<23 mW
Flash-memory	Winbond W25N01GW	1 GB, NAND-flash	QSPI	<45 mW
Sensors: analog and digital; typical data output rate 100 Hz				
Inertial	ST LIS2DE12	acceleration (±2 g to ±16 g)	I2C	15 µW
	Bosch BNO055	acceleration (±2 g to ±16 g)	I2C	0.3 mW
		angular velocity (±2000 deg / s)		12.0 mW
		magnetic field (2500 mT, 16-bit)		11.8 mW
		orientation (Euler angel, 16-bit)		36.9 mW
Physiological	discrete	ECG (1-Lead, 12-bit)	DAC, ADC	315 µW
		respiration curve (prototype, 12-bit)	DAC, ADC	<1.4 mW
Sensors: analog and digital; typical data output rate 1 Hz or event-based				
Physiological	calculated	heart rate (Pan-Tompkins algo.)	CPU	16.1 µW
Environmental	Bosch BME280	temperature (−40 °C to 85 °C)	I2C	3.2 µW
		relative humidity (0 % to 100 %)		5.0 µW
		air pressure (300 hPa to 1200 hPa)		6.5 µW
	ST MP34DT04	sound pressure level (max. 120 dB)	PDM	1.1 mW
Indication: user indication and interaction				
RGB-LED	SML-LX0404SIG	30 mcd, 40 mcd and 20 mcd (red, green, blue)	Soft-PWM	4.7 mW
Vibration motor	C1026B002F	9000 rpm	GPIO	270 mW
Piezo buzzer	CPE-163	20 Hz to 20.000 × 10^3 Hz, max 80 dB at 4.8 kHz	PWM	30 mW
Push-button	KSR223GNCL	user event	GPIO	-
Push-button	PTS840 GK	hardware reset	GPIO	-

4 Monitoring Cognitive Workload

This chapter provides insights into the use of body sensors to estimate psychological or cognitive workload (CW). At first, the concept of CW is outlined (sec. 4.1). In the following, a summary of possible methods to measure CW (sec. 4.2) is presented.

Subsequently, the potentials and limitations of using WBS to estimate CW are presented and discussed on the basis of experimental data collected for this thesis (sec. 4.3). Parts of this chapter (software environment, selection, and comparison of sensors) are based on the master's thesis of Christian Menßen [163]. Experimental work was done in collaboration with Torben Tönniges, who made use of the data set for a vision-based affective computing approach. The results presented here already have been partially published in [273, 274]. Since then, the data set was expanded with an additional test run to increase the number of participants and give more detailed insights regarding different dimensions of CW.

4.1 Background and Motivation

Cognitive workload is of interest in public and occupational health. Topics linked to CW are arousal or fatigue as well as the catchphrase *stress*. The concept of "stress" was originally proposed by Selye [210]. It was used to summarize bodily responses to external stressors, e.g. cold, heat, or fear [210, 211]. Before the term became part of everyday language, in its original definition, it was summarized as "general adaptation syndrome" [210]. Thus, in its broadest definition, stress is understood as the "nonspecific response of the body to any demand made upon it" [211]. It is a collective term nowadays often associated with the idea of an over-strain in regard to mental work. As such, the term *stress* has predominantly negative connotations. In its original understanding, it was seen more neutrally, as an activation of the body. In this way, stress is a normal reaction of the body that is found often as a response to a potentially dangerous situation. Therefore, the stress reaction is to be considered quite positive since the body is prepared in such a way that allows reacting optimal (sec. 2.1).

Concepts related to stress are the psychological, mental, or cognitive workload. It should be noted that these terms are often used synonymously. Also, there is no single definition of one of these terms to date. The phrase "cognitive load", however, is based on the work of Sweller [230]. In his work, Sweller focused on the capacity of cognitive processing in reference to learning. Later Sweller et al. proposed the "cognitive load theory" [231], in which it is assumed that the working memory (in the brain) is limiting

success in learning. Cain [48], however, proposed to summarize the term mental workload as "the capabilities and effort of the operators [or users] in the context of [a] specific [situation]". Hence, CW is not a uni-variate, but a "multifaceted" [48] entity.

Following this definition, it becomes evident that CW is a subjective quantity, as it depends on personal capabilities. Also, the definition is broad, as it encapsulates any situation and task. Correspondingly, the comprehensible definition of [40] states workload (as a cause of "occupational stress") to be: "an all-encompassing term that includes any variable reflecting the amount or difficulty of one's work". This chapter will follow that definition and refer to it as CW. Its definition is close to the one used for stress. However, in this chapter, the cause of the stress-reaction is focused on, which is the execution of specific tasks that differ in their difficulty. The type of task and difficultly remains without restriction and could be of any kind, e.g. memory or time demands, clear or unclear presentation of information but also frustration, e.g. caused by an unsolvable task.

The operationalization of stress (here understood as a general concept of CW) is of uttering importance because chronic stress is linked to cardiovascular diseases, headaches, and long-term effects like depression. It is also known that excessive stress causes a majority of work-related illnesses. According to Chisholm et al. [59], the annual loss in the global economy is 2.5 to 8.5 trillion USD per year. In this respect, the effects of both constant over- and under-load are addressed in research [34, 87].

In occupational contexts, the opposite of stress is boredom, which is often found due to monotony. Like stress, boredom is causing mental fatigue. In conjunction with work-related illnesses, the extreme is often referred to as burn-out syndrome [157]. In German literature also the contrary phrase "bore-out" [224] was coined. Both are seen to restrict performance at the workplace. An often-cited model depicting this effect is the so-called Yerkes-Dodson Law [263]. In a simplified version, it is explained as an inverted U-shaped relation between arousal and performance. In that way, there is an optimal point between arousal and performance. Under- or over-load would decrease performance. That, however, as stated by Diamond et al. [75], is an often misunderstood and over-simplified view. According to the original version presented by Yerkes et al. [263], it applies to difficult tasks only. For simple tasks, higher arousal would further increase performance instead.

Regardless of the conceptual model used to explain CW, it can be seen that tools allowing to measure the same are needed. For example, to improve workplace design, processes, or occupational health management.

In addition to the long-term consequences (health-related effects) of stress, short-term effects are of interest as well. This is in the area of *human factors* or *affective computing*, which addresses the design and development of human-machine interfaces or emotion aware computing software, respectively[1]. Short-term effects are important since estimating a person's affective state or CW can be seen as an enabler towards

[1] As a source of further information on this topic is the book by [117]

responsive *affective computing* applications. These include the realization of *adaptive Human-Machine-Interaction* [273] or building "empathic machines" [207] (the latter is more about estimating emotion than CW).

The idea behind this is to allow or qualify software (and technical systems in general) to reacting user-oriented. For instance, a software application could provide additional explanations, or a robot could interrupt its current task if the user reacts with a particular emotion or change in CW. Also, in regard to the improvement of software, feedback on CW perceived when using individual software components can potentially provide valuable feedback for usability optimizations [127].

In the context of modern working environments, the interaction of arousal and performance is actively discussed [264], i.e. to improve the use of assistant systems. A modern working environment here is understood as a workplace that requires time and task-related flexibility (sequence of processes cannot be planned ahead because of on-demand production). An assistant system here is understood as any technical system (computer or robot) that assists a human worker. It is argued that for the human user to fulfill the requirement of flexibility, an assistant system needs to know the human user's cognitive capacity. Therefore, in order to adjust correspondingly to the user's needs, it is important to precisely model the user's perceived CW in detail and in real-time. A similar but more common example or use-case for real-time estimation of CW is car driving [201]. In order to prevent accidents, a driver needs to keep a constant high level of arousal. A technical system (car) being capable of monitoring CW or estimating a decrease in arousal could warn its driver beforehand.

4.2 Operationalization of Psychological Workload

The measurement of CW is as divergent as its definition. Firstly, CW can either be measured subjectively (e.g. by a self-report) or objectively via performance measures (e.g. hours on the job). Since CW is of a subjective nature, the subjectively reported perceived CW is of primary interest. Here again, objective measures are sought that can be used or serve as a predictor, e.g. the error rate or time-on-task. This information can then be compared and brought together with subjectively perceived CW.

Also, measuring brain activity through electroencephalography (EEG) or utilizing psycho-physiological measures [48] (heart rate, pupillary diameter, perspiration) are commonly found and used as predictors for CW. In addition to psycho-physiological responses, behavioral changes, like facial expressions or the use (interaction) with a device, e.g. a smartphone or computer-mouse as well as contextual information, i.e. location, e-mail usage are used as predictors as well [8].

Self-reported measures are typically questioned in the form of a Likert scale. They are used in an uni-variate manner, e.g. asking a single question on the perceived difficulty of a task, the perceived stress level, the level of fatigue, or the emotion. Also, multidimensional tools for stress or CW assessment are available. Commonly found

tools, inventories scales or indices are the “Perceived Stress Inventory” (PSI) and its predecessors, e.g. the “Perceived Stress Questionnaire” (PSQ), the “Stress Response Inventory” (SRI) or relative tools like the “State-Trait Anxiety Inventory” (STAI) or the “Perceived Stress Scale” (PSS), to name a few [135]. In addition to these tests, which explicitly measure stress, the “NASA-Task Load Index” (TLX), as proposed by Hart et al. [103], is also used. As the naming suggests, its focus is on workload consisting of multiple dimensions including, mental, physical, and temporal demands as well as performance, effort, and frustration.

Psycho-physiological measures are based on the physiological responses of the human body. Because CW is a process of the CNS (i.e. the brain), direct observation of the brain’s activity through imaging methods like positron-emission tomography, near-infrared spectroscopy, magneto-encephalography or electrical methods (i.e. EEG) are used. These methods differ in their temporal and spatial resolution. For instance, EEG offers high temporal resolution but cannot be used to distinguish activity in different brain areas. Furthermore, most methods are limited to lab-use only as the devices are too big, costly, or require the participant not to move.

Because of its relatively good availability, size, and cost, EEG is an often-used technique in stress or workload studies. The EEG is used to measure the electrical activity of the brain. Several electrodes are placed on the head for this purpose. For optimal signal quality, additional electrode gels must be used. The recorded EEG signals are then typically examined using frequency analysis. Often, 5 frequency bands are distinguished, including alpha (8 Hz to 13 Hz), beta (13 Hz to 30 Hz), delta (0.1 Hz to 4 Hz), and theta (4 Hz to 8 Hz) band. The frequency bands are associated with different mental states, e.g. alpha waves are found in relaxed participants, while beta waves are more frequently observed when the participant is in a stressful situation [8]. However, the construction of a wearable EEG-system to reliably capture EEG in every-day life is still a challenging and active research object.

Instead of analyzing brain activity directly, bodily functions can serve as an indirect measure of CW. In CW studies, what one would like to observe is the shift of predominant PNS to SNS control over bodily functions. Since these are controlled by the ANS or CNS (sec. 2.1), they allow making conclusions on CW. It is the anticipatory effect, i.e. adjustment of bodily functions, e.g. the stress-reaction as presented by Selye, which is observed and allows to make conclusions on the arousal, emotion, stress, or CW.

Well known and most used signals of bodily functions to quantify stress or estimate CW are based on HR or HRV [118] as well as on the EDA[2] [116]. Additionally, other physiological signals are used, namely blood pressure, skin temperature, or respiration rate. In addition, hormone levels can be monitored in order to capture CW. The most common method is to measure the salivary cortisol level. A comprehensive overview of these methods is given in [8, 215].

[2] electrodermal activity is also referred to as “galvanic skin response” (GSR). This notation however is outdated.

4.2.1 Electrodermal Activity

To reflect the activity of the SNS, i.e. arousal or CW, EDA is a well-known and often-used psycho-physiological parameter. An introduction is given in [47], which will be briefly summarized here. The term electrodermal activity (EDA) is a general term, which reflects all electrical changes of the skin. It is an interesting estimator for the activity of the SNS because EDA changes in response to sweat secretion, which is exclusively controlled by the SNS. It is usually measured using a bipolar electrode configuration (2 electrodes). Therefore, electrodes are placed on the index and middle finger or the palm. A constant voltage is then applied between the electrodes. The resulting current flow, which is the skin conductivity, is measured. The corresponding SI unit is Siemens (S). With respect to EDA, typically reported in the order of μS. Also, skin resistance is reported instead. It is given in Ohm (Ω), which is the reciprocal value of the conductivity.

For later analysis, the raw EDA signal is typically separated into a low- and a high-frequency component, which are named tonic- or phasic-level[3]. The energy of the tonic- and phasic-level ranges from 0 Hz to 0.05 Hz or 0.05 Hz to 1.5 Hz, respectively [266].

The tonic level, the skin conductance level (SCL), changes very slowly, and strong inter- and intra-individual differences can be found. It is often observed that SCL decreases during resting conditions, while it slowly rises with increased arousal. In contrast, the skin conductance response (SCR) denotes rapid changes in EDA. It is observed as a short peak in the signal, which occurs about simultaneously with an external stimulus. The latency between the stimulus and the SCR is about 1 s to 3 s. In the absence of an external stimulus, these peaks are referred to as a nonspecific SCR (NS-SCR).

To analyze EDA, absolute values are normalized participant wise, e.g. in an initial learning phase. Afterward, statistical measures such as mean, median, standard deviation, or kurtosis and skew are calculated to examine the signals. Regarding SCR, the peaks are analyzed with respect to their rise time (duration from minimum to maximum value), recovery time (duration from the maximum to half of the maximum value), amplitude, or area.

4.2.2 Heart Rate Variability

Heart rate variability (HRV) is another widely used parameter to reflect the activity of the ANS in general. In contrast to EDA, HRV is doubly controlled by SNS and PNS. It is measured on the basis of an ECG recording, calculating the distance between 2 successive QRS-complexes, known as RR-interval. The variation in the RR-interval reflects changes in the ANS, which in turn are used to estimate stress. The maximal

[3] In physiology, receptors that are slowly or rapidly adapting to a stimulus are referred to as tonic or phasic receptors, respectively.

energy of the HRV signal is in the range of 0.003 Hz to 0.4 Hz [154, 214]. In general, the low or high frequencies band is associated with SNS or PNS activity, respectively.

HRV is influenced by a multitude of physiological effects. For instance, HRV is directly linked to respiration, which is known as respiratory sinus arrhythmia. This is an effect where the heart rate increases while the participant inhales and decreases while the participant exhales. It is explained as an effect of atrial pressure, which is also lowered when the participant inhales. This triggers the baroreflex, and in turn, it can be observed that HR is modulated by and synchronized with respiration. Likewise, other causes for a change in blood pressure act directly on HRV. [188]

To analyze HRV, multiple measures have been proposed in the literature. Among them are statistical measures in the time domain (e.g. standard deviation, or root-mean-square), frequency analysis (e.g. as a ration of low and high frequencies), or geometrical features (e.g. utilizing histogram information) [154, 188].

4.3 Experimental estimation of Cognitive Workload

In this section, the use of WBS to estimate CW is presented.

At first, related work is summarized (sec. 4.3.1). Afterward, the methods are outlined (sec. 4.3.2). This includes sensors used to measure psycho-physiological data and the hardware used for the conducted tablet-interaction study (sec. 4.3.2.1 - 4.3.2.2). In summary, subjective cognitive workload reports are queried during the experiment (sec. 4.3.2.3), and in parallel, HR, and EDA are measured. Based on these data, ML-models that allow for detailed estimation of short-term changes in cognitive workload are presented (sec. 4.3.2.4 - 4.3.2.6).

As parts of the result section (sec. 4.3.3), a detailed analysis of the perceived CW is given (sec. 4.3.3.1), feasible psycho-physiological features are identified (sec. 4.3.3.2), and fine-grained estimation models with 5 levels of graduation are introduced (sec. 4.3.3.3). In the discussion (sec. 4.3.4), the results are critically reflected. Also, uncertainties in both the subjectively reported CW and the measurement data obtained through WBS are discussed. This section concludes with a summary of the results and a prospect of the future use of WBS for CW estimation (sec. 4.4).

4.3.1 Related Work

The possibility of estimating CW has been presented frequently in recent research. The estimated demands or strain are often referred to as “mental workload”, “cognitive workload”, “cognitive load”, or more generally, as stress. A survey of recent work on the estimation of stress can be found in the methodological review given by Alberdi et al. [8]. Selected results of related work are presented in the following:

The effectiveness of HRMs in detecting mental stress was demonstrated by Choi et al. [62]. They highlighted the importance of an unobtrusive design to obtain a high user

acceptance rates. With their approach, they were able to distinguish between stressed and non-stressed mental states with an accuracy of 69 %.

The feasibility of using ultra-short-term HRV features was also investigated by Salahuddin et al. [203]. They validated numerous short-term metrics and evaluated the minimum duration of RR-Intervals needed (10 s to 50 s) to distinguish between stress and non-stress situations in 24 participants. To induce stress, they made use of a Stroop test.

Within the work of Wijsman et al. [253] and Choi et al. [61], it was shown that the combination of HR and additional predictors (here respiration rate and EDA) improves the estimation's accuracy (79 % and 81 %). The effectiveness of EDA was also verified against objective measures (salivary cortisol) by Amalan et al. [17] in a controlled setting. They further highlighted the importance of labeling false-positive EDA peak signals that are introduced by physical activity, in the form of motion artifacts.

Looking at physical activity the other way around, the problem of detecting CW (or stress) by considering physical activity as an additional predictor is emphasized by Karthikeyan et al. [123]. They used physical activity information in order to prevent it from becoming a confounding factor, and report an accuracy of up to 92.4 %.

Additionally, Sun et al. [228] focused on short-term signal processing, which enables the detection of short-term stress events. They presented remarkable results with a classification accuracy of up to 95 %. Moreover, fine-grained estimation of stress has been addressed in the work of Healey et al. [109]. They estimated the perceived stress level of drivers in 3 distinct gradations with an accuracy of up to 97 %.

Similar work was presented by Li et al. [141]. They reported an accuracy of 96 % for a 5-leveled classification task. The reported accuracies, however, differ. For instance, the very same problem was also examined by Manawadu et al. [155]. They considered the fine-grained estimation of CW in the setting of semi-autonomous driving and reported an accuracy of 74.5 %.

The outlined experiment aims towards a combined contemplation of CW estimation. It is intended to provide a detailed (or fine-grained) estimation that is also based on short-term signals and non-laboratory but consumer-grade sensor equipment. The solution to be found is also intended to be suitable for non-laboratory use and, in this respect, has to be portable onto a consumer-oriented device (e.g. WBS). This chapter thus focuses on both the short-term signal processing of multiple parameters and the fine-grained estimation of CW. Both are mandatory requirements in order to enable adaptive assistant technology.

4.3.2 Methods

In the following, the hardware (sec. 4.3.2.1) and experimental setup (sec. 4.3.2.2) is described. This is followed by an explanation of the assessment of ground truth data (sec. 4.3.2.2) and the methods used to extract (sec. 4.3.2.4) and select (sec. 4.3.2.5)

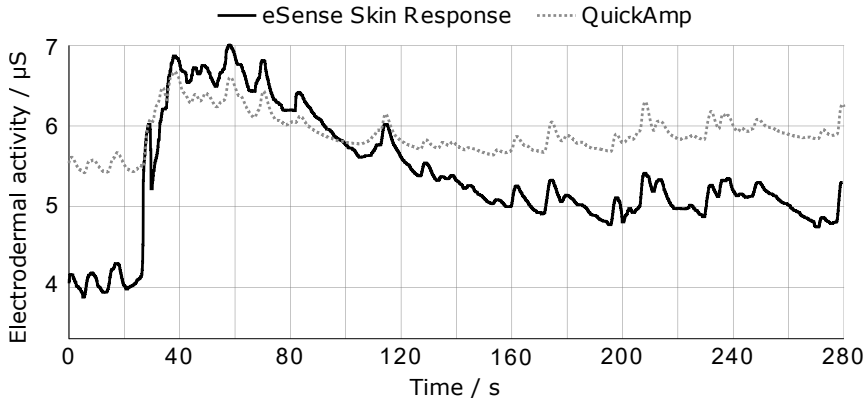


Figure 4.1: Comparison of two concurrent measures of EDA by a mobile (Mindfield eSense) and a reference system (Brainproducts QuickAmp). Signals differ in absolute values but show a high linear correlation (Pearson correlation coefficient > 0.8. Previously presented in [163]. Reprinted with permission, ©2017, Springer Nature, [273]

important features. Finally, an overview of the methods used to classify the data is given (sec. 4.3.2.6).

4.3.2.1 Hardware Setup

The hardware setup is based on the Google Nexus 10 tablet computer[4], which has sufficient computing power for the desired task and allows easy integration of the external sensors. It was used to run the experiment's software, log the sensor data, and retrieve the participants' perceived stress level.

The EDA is captured by using the Mindfield eSense Skin Response system, which is a portable solution designed for tablet computers and smartphones. It is connected to the tablet computer by the microphone jack. On the participant's side, electrodes (hook and loop) are placed around the participant's index and middle finger.

In order to validate the functionality of the Mindfield system, it was compared to a Brainproducts EDA sensor connected to an appertaining QuickAmp Amplifier[5] as a reference system (Figure 4.1). Both systems showed different outputs in terms of absolute value. Nevertheless, signals showed close agreement (Pearson correlation coefficient > 0.8). Therefore, the mobile and inexpensive Mindfield system is used for this study.

[4] GT-P8110; Google Inc., Samsung Electronics
[5] Brain Products GmbH, `http://www.brainproducts.com`

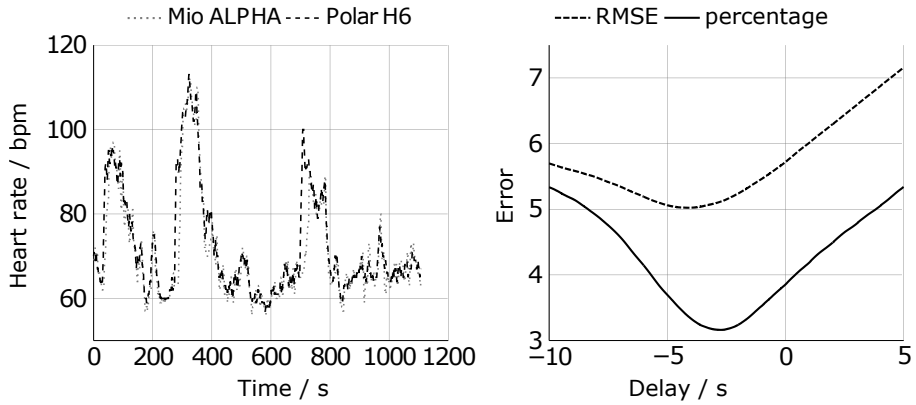


Figure 4.2: Comparison of two concurrent measures of heart rate by an optical sensor (Mio Alpha) and an ECG-based system (Polar H6). The signals show a temporal delay, which is why it is assumed that the optical measures are smoothed. Previously presented in [163].

The HR was captured using two redundant systems. Firstly, the ECG-based Polar H6[6] HRM is used, which is attached to a chest strap. Secondly, the PPG-based Mio Alpha watch[7] was used, which is worn around the wrist. Both HR sensors communicate wirelessly with the tablet computer via BLE. Measurement readings from both devices were comparable (mean deviation of 3.85 %). However, it was noted that the Mio Alpha smooths the measured values (Figure 4.2). In contrast to the Polar H6, it does not provide R-peak intervals, which are necessary for HRV calculation. Albeit initial methods to extract HRV features from PPG signals were presented in the past [30], these are still under development. The current consensus is that they cannot fully replace ECG-based measurements [183]. For this reason, only data obtained from the Polar H6 module is used in the following.

4.3.2.2 Experimental Setup

The conducted experiment was designed to induce different levels of CW during the interaction with a tablet computer. In total, 31 participants volunteered to participate in the experiment (20 male, 11 female, mean age 28.2 ± 9.1 a). Of the 31 participants, 15 or 16 were recruited during the 1st or 2nd test run, respectively. In the 1st test run, participants were mainly male students (14 male, 1 female, mean age 25.9 ± 2.1 a). In the 2nd test, more female participants could be recruited (6 male, 10 female, mean age 30.4 ± 12.3 a).

[6] Polar Electro Oy, `http://www.polar.com`
[7] Physical Enterprises Inc. (Mio Global), `http://www.mioglobal.com`

All participants were informed about the experiment's design and gave their informed consent. The experiment lasted approximately 20 to 25 minutes for each participant and was repeated after a short break. During the break, the sensors were reapplied to increase robustness in terms of repeatability concerning the differences in the sensors' attachment. The participants' hands were filmed during recording to find possible motion artifacts in the EDA-signal afterward. The electrodes were applied to the index and middle finger of the non-dominant hand.

Each trial of the experiment was divided into the following 5 phases:

1. Relaxation video (2 minutes)
2. Memorize items (3 to 4 minutes)
3. Stroop test (3 to 4 minutes)
4. Recall items (4 to 5 minutes)
5. Memory and reaction test (3 to 4 minutes)

At the beginning of the experiment, the equipment was introduced to the participants. The sensors were then attached by the participants themselves and tested afterward by the experimenter. After this setup, the experiment started with a presentation of short sequences of relaxation videos. This was done in order to prevent possible effects resulting from the excitement of the ongoing experiment. Thereby, the participants were given the possibility to test the tablet computer and to familiarize themselves with the sensors attached to them. The participants were then asked to select the video they found to be the most relaxing. Afterward, the measurements were started.

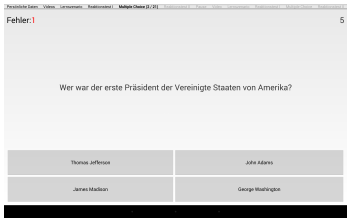
At the beginning of the experiment (phase 1, Figure 4.3a), the previously selected relaxation video was presented to the participant (phase 1, video duration 90 s). It is intended to record a baseline measure of HR and EDA in this way.

Next, a memory test was initiated (phase 2). During this phase, 12 items of learning content were provided to the participant. The learning content consisted of demographic and economic data of the United States (during the 1st trial) or the Czech Republic (during the 2nd trial). For each item, the time to memorize the provided information was limited to 10 s. Before the memorized content had to be recalled by the participant (phase 4, Figure 4.3b), a Stroop test [227] was carried out (phase 3).

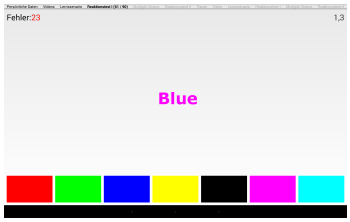
During the Stroop test (phase 3, Figure 4.3c), the participant had to touch the button with the color that is identical to the color of a text shown on the screen. The background color, the number of possible solutions (buttons), and the time available to answer were altered randomly. Hence, the Stroop test challenged the user with varying intensity levels. Overall, the participant was asked to reply to 90 Stroop items during 6 repetitions (15 items each). A short break preceded every repetition.

Subsequently, the participant was asked to recall (phase 4) the learning content from phase 2. This was done by offering multiple-choice questions. In total, 7 questions were composed into 3 blocks of varying difficulties. To increase the CW for the multiple-choice test in each block, the available time to answer was reduced (7 s, 6 s, and 5 s). Additionally, in the last block, only invalid answers were provided.

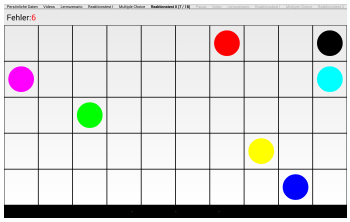
(a) Relaxation video; phase 1



(b) Memorize and recall; phase 2/4



(c) Stroop test; phase 3



(d) Memorize and reaction; phase 5

Figure 4.3: Screenshots of the relaxation video (a), memory test (b), Stroop test (c), and the checkerboard (d) presented to the participants during the tablet-based CW experiment. Previously presented in [163]. Reprinted with permission, ©2017, Springer Nature, [273]

At the end of the experiment (phase 5), the participant had to perform a mixed memory and reaction test. For this test, colored circles were consecutively drawn on to the screen. The participant's task was to memorize the color sequence and to recall it afterward immediately. The difficulty was altered by changing the number and duration of the circles shown (3 to 7 circles were shown for a duration of 700 ms to 500 ms each). Moreover, the number of colors used was changed randomly (3 to 7). A checkerboard was presented to the participant (Figure 4.3d), which allowed them to enter the recalled color sequence. The checkerboard was sparsely filled with colored circles (randomly distributed). The participant was asked to recall the color sequence, which was shown beforehand, by touching the corresponding circles.

The proposed experiment abstractly covers typical tasks with which workers are faced. The abstraction focuses on the tasks to memorize and recall various working steps, e.g. while assembling a workpiece or wiring a cable harness at the production line (mixed reaction and recall test, phase 5). The worker has to recall a new working process under time pressure. Another example is performing and following a diagnostic sequence. In this case, the worker has to memorize facts and later on recall and compare the results (memory test: phase 2 and 4).

4.3.2.3 Ground Truth

In order to obtain ground truth data, all participants were asked to self-report their perceived CW on a scale from 1 (lowest) to 5 (highest). This self-report was inquired directly after a specific task was finished during every phase of the experiment (sec. 4.3.2.2). Thereby, during each trial of the experiment, each participant was asked 17 times to give a self-report of the perceived CW. This self-report was then assigned as ground truth (target label) for the previously performed task.

In addition, in the new (2nd) test run of this experiment, the participants were asked in more detail about their perceived CW. Therefore, the NASA-TLX score ([103], sec. 4.2) was used. Questions were out-handed to the participants on a printed piece of paper, and the different dimensions (items) were briefly explained to the participants (Table 4.1). In the tablet's application, only the short title was mentioned as a reference.

The NASA-TLX items were added to examine the source of CW in more detail. This is because there is a suspicion that the uni-modal Likert scale, which was used during the 1st test run [273], was maybe too unspecific. However, in order to compare both test runs, both metrics are kept. Thus, also in the 2nd test run, the participants were asked to answer the more general question of perceived CW (Likert scale) without modification. Moreover, difficulties were not adapted (e.g. for the Stroop, or memory and reaction test), although it was found that only a few participants reported very high CW during the 1st test run. This was also done to keep both runs comparable.

4.3.2.4 Pre-Processing and Feature Extraction

The utilized Polar H6 provides HR and the RR-interval for each recognized heartbeat. For this reason, the data stream is recorded in non-uniform time intervals. To enable a conventional frequency-based analysis of the data, it is re-sampled to 4 Hz as proposed by Singh et al. [217]. For this transformation into the frequency-domain, Welch's method, in combination with a Hamming-window, is used. Prior to the feature extraction, the RR-interval is normalized, and polynomial trends are removed (detrending), as demonstrated by Tarvainen et al. [233]. Furthermore, HR and EDA for each participant are min-max normalized (sec. 2.3.3.2, eq. 2.20), to increase inter-subject comparability.

The EDA is captured with a sample rate of 10 Hz. In order to remove outliers, a low pass filter with a cut-off frequency of 0.5 Hz is applied to the raw signal. Furthermore, the raw EDA signal is decomposed into SCL and SCR, as described by Choi et al. [61]. Their method is based on the approach from Tarvainen et al. [233], which was also used for detrending the RR-interval beforehand.

Statistical data (minimum, maximum, mean, standard deviation) is calculated from HR, HRV, EDA, SCL, and SCR. In addition, amplitude, duration, area, and frequency of the EDA and SCR signals are computed. Furthermore, commonly known features based on HRV are used [154, 243]. To extract those, the *HRV-Toolbox*[8] was used.

[8] HRV-Toolbox by Marcus Vollmer, version 1.0 `www.github.com/MarcusVollmer/HRV`[243].

Table 4.1: Title and description of the NASA-TLX items used in the 2nd test run of the experiment and presented to the participants.

Title	Description
Physical Demand	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Mental Demand	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Frustration Level	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?
Temporal Demand	How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	How successful do you think you were in accomplishing the goals of the task? How satisfied were you with your performance in accomplishing these goals?

Table 4.2: Overview of all signals used and corresponding features extracted.

Source	Feature
HR	mean, standard deviation, minimum, maximum
HRV	meanNN, pNN50, RMSSD, SD1, SD2, SD1/2, SI, skew, kurtosis, TRI, TINN, RRmed, RRqr, VLF, LF, HF, nLF, nHF, LF/HF
EDA, SCL, SCR	mean, minimum, maximum, standard deviation
SCR peak	count, duration mean, duration sum, amplitude mean, amplitude sum, area mean, area sum

The experiment was carried out using a tablet computer. Therefore, the mean pressure, mean duration, and the total count of touch events on the screen were additionally recorded during the experiment. These features were intended to be used to reflect behavioral changes of the users. However, as already noted in previous work [274], these features show a spurious relationship with the different experimental phases. This is because no normalization strategy was applied to adjust the number of interaction events, neither during the experiment nor for later analysis. For this reason, there is a correlation between the number of touch events and the task or its difficultly. Hence, touch features are excluded from the analysis[9].

In total, 42 features are extracted (Table 4.2; Detailed information and explanations can be found in Table 4.5 located at the end of this chapter.) from the different sensor elements (HR, EDA). Because the extracted features are not all commensurate, min-max scaling or z-transformation is used (depending on the classifier used, sec. 2.3.3.1, eq. 2.20 - 2.21).

4.3.2.5 Feature Selection

To identify the optimal window size and overlap, multiple feature subsets are derived based on the corresponding sensory element (HR, EDA). These subsets are then empirically explored by comparing the predictive performance for each combination of subset, window size, and overlap in a grid search. The window size was set to 10 s to 120 s in steps of 5 s. The overlap was set to 0 % to 75 % in steps of 12.5 %. To evaluate the predictive performance, the mean accuracy from a stratified 10-fold CV of DTs with a maximum of 100 splits is refereed to.

4.3.2.6 Classification

With a comparative analysis, the potential of the selected feature set for fine-grained and short-term estimation of CW is to be analyzed. The analysis consists of a comparison of multiple fine-grained supervised classification models. These models make use of the same feature set and window size configuration, which was evaluated beforehand (sec. 4.3.2.5).

For comparison, well-known classifiers are selected. The ML-models are trained using the correspondent *MATLAB*[10] toolbox implementations. The evaluated methods are: naive Bayes (NB), decision tree (DT), k-nearest neighbor (KNN), support vector machine (SVM), and Gaussian process (GP). For each method, hyper-parameters were tuned using Bayesian optimization.

The predictive performance measures referred to are accuracy, sensitivity, specificity, and precision (sec. 2.3.3.1). In order to prevent an over-fitting of the classifier, 10-fold

[9] It is to be noted that, using a normalization strategy touch input could be used to estimate CW as demonstrated in Hernandez et al. [112]

[10] The MathWorks, Inc., Version 2018b, `https://www.mathworks.com/products/matlab.html`

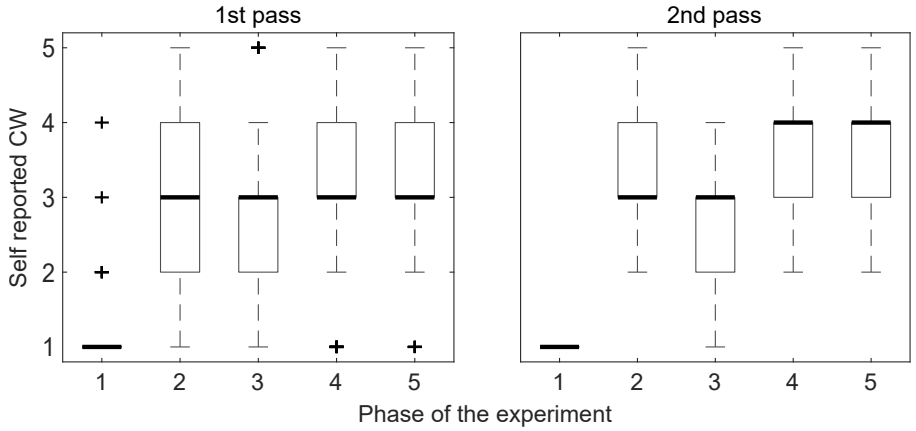


Figure 4.4: Distribution of the self-reported CW level (ground truth) during the 1st and 2nd trial of the experiment, grouped by the experimental phase. The phases are: 1. relaxation video, 2. memorize items, 3. Stroop test, 4. Recall items, 5. memory and reaction test.

CV is applied. To evaluate the generalization of the classifier, results utilizing LOGO CV are considered additionally.

4.3.3 Results

In this section, findings from the experiment are presented (sec. 4.3.2.2), and the selected feature subset is revealed (sec. 4.3.3.2). Finally, a comparison of the trained classifiers is given (sec. 4.3.3.3).

4.3.3.1 Experiment

To verify that the participants were adequately challenged (i.e. CW was induced) during the experiment, the self-reported CW (ground truth) from all phases of the experiment is compared (Figure 4.4).

As expected, a significant difference between the relaxation phase (phase 1) and all other phases of the experiment can be observed. Also, it is noticed that the participants rate the CW for the phases 2 (memorize items, CW: 3.0 ± 0.9) and 3 (Stroop test, CW: 2.7 ± 0.9) equally (no significant difference). The same applies to the phases 4 (recall items, CW: 3.4 ± 1.0) and 5 (memory and reaction test, CW: 3.3 ± 0.9). It is concluded that the participants were equally challenged during both tasks or groups of tasks (phases 2 and 3, or phases 4 and 5, respectively).

Moreover, it was observed that only a few participants reported very high CW during any of the phases, e.g. level 5. This was already observed in the 1st run of the

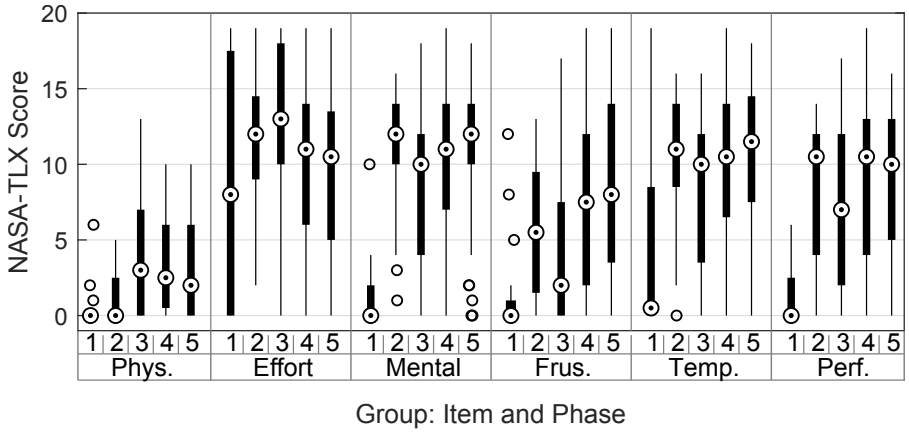


Figure 4.5: Tukey plot for each TLX item during both trials of the experiment grouped by the experimental phase. The phases are: 1. relaxation video, 2. memorize items, 3. Stroop test, 4. Recall items, 5. memory and reaction test.

experiment. However, to keep both runs comparable, the difficulty was not increased. In all, 41.9 % of the participants reported 5 different levels of CW during the experiment, and 84.4 % of the participants reported at least 4 different levels of CW during the experiment. The remaining participants reported 2 or 3 different levels of CW.

Additionally, the repeatability of the experiment is verified by comparing the 1st and the 2nd trial of the experiment (both runs). Similar mean and variance are found concerning the self-reported CW levels during the different experimental phases (Figure 4.4). With paired t-test, the null hypothesis that the self-reported CW between both trials is equal was not rejected ($p = 0.93$). Thus, it is concluded that there is no significant difference in the perceived CW during both trials of the experiment. It should be noted that in the 1st trial of the experiment, the mean CW during the relaxation phase was higher and also showed higher variance (1.3 ± 0.5) compared to the 2nd trial (1.2 ± 0.2). From this, it can be reasoned that a longer relaxation phase is necessary to allow the participants to accustom themselves to the situation.

Regarding the NASA-TLX items, group differences were tested using analysis of variance (ANOVA) (Figure 4.5).

Significant differences for physical demands were reported between phase 1 and the last 3 phases ($p<0.05$), and also between phase 2 and 3. The first 2 phases did not require any interaction with the tablet computer, but clicking to get to the next screen. For all other phases, however, interaction with the tablet computer was needed.

Besides, differences between the Stroop test (phase 3) and the other phases are striking. Regarding the Stroop test, *effort* was rated higher than for any other phase. This applies vice versa to the *mental*, *frustration*, *temporal*, and *performance* item. Here,

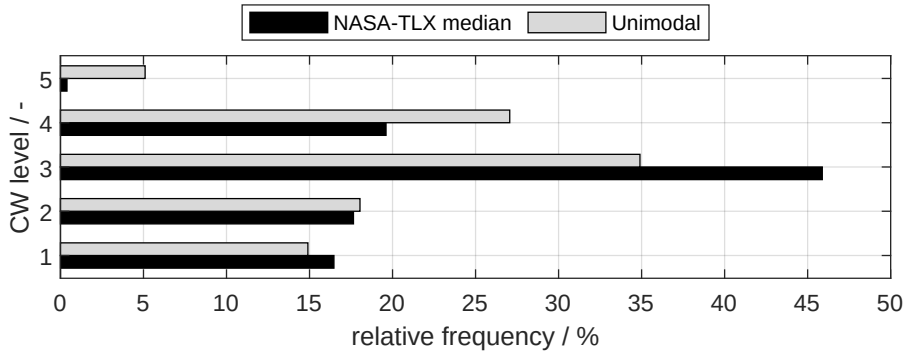


Figure 4.6: Comparison of the unimodal CW metric and the median of all NASA-TLX items.

the Stroop test was rated the lowest (excluding the relaxation phase). The differences for phase 3 are significant ($p < 0.04$) compared to phases 4 and 5, given the *mental* and *frustration* item. For all remaining items (*physical*, *effort*, *temporal*, *performance*), no significant differences between the phases of the experiment could be found.

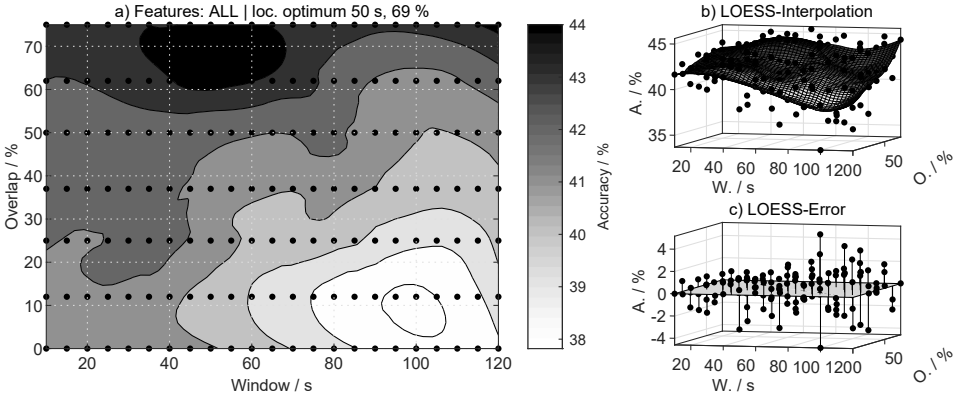
Furthermore, by comparing the uni-modal CW item against the median of all NASA-TLX items, no difference could be found (Figure 4.6). It can be concluded that the details with respect to the different items of the NASA-TLX are limited. Also, from the 1st run, it was seen that a more precise estimation of CW, i.e. with a gradation of more than 5 classes, is hardly possible. Thus the uni-modal CW measure is used in the following as ground truth.

4.3.3.2 Feature subset

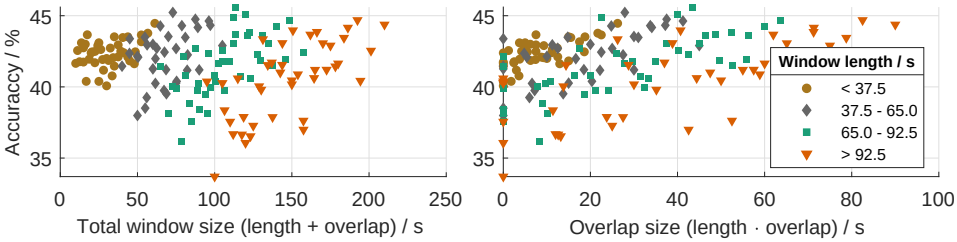
Based on the data of the 1st run alone, the best feature subset was found using sliding windows with an overlap of 75 %. This result was reported in [274]. It applied to all tested feature subsets, including HR and EDA features, regardless of the window size (grid search between 10 s to 60 s).

Concerning the window size, however, the results were not equally consistent. For the feature set containing HR alone, a significant trend or correlation between the classifier's performance and the window size could be identified (Pearson's $r = 0.9503$, $p < 0.05$). However, with regard to the EDA feature subset, no trend, but a local optimum, was found around 40 s to 45 s. It was concluded that there is no all-encompassing optimal window size or overlap, but each subset has its own optimum.

The data set under consideration in this chapter contains additional raw data from 16 participants. It is thus about twice as large as the data set used in [274]. Similar to how it was done in [274], in order to estimate usability and to determine the optimal



(a) Accuracy map for different window sizes and overlaps.



(b) Detailed evaluation of window size and overlap effect on accuracy variance.

Figure 4.7: CW estimation accuracies found with grid search across window size and overlap (a). Results were generated using pruned DTs trained on the complete feature set. Evaluated results were interpolated using locally weighted scatterplot smoothing (LOESS). Alongside, the detailed view on the distribution of accuracies with altering total window sizes (length + overlap) and amounts of re-used data due to window overlapping.

window size and overlap in advance, the accuracy of 10-fold CV DTs are evaluated for each feature set.

In general, the findings on the here evaluated data set are in agreement with previous findings from [274] (Figure 4.7a). Previously, a local optimum was found around a sliding window with a length of 40 s and maximum overlap (75 %). Again, it can be seen that the window size does have little influence on the classification's accuracy (for the range that is part of the grid search). Instead, the window's overlap determines the classification's accuracy. Here again, a trend is found, which is that the accuracy increases at the same time as the overlap is increased. The region with the highest accuracy is found for window sizes between 30 s to 70 s and overlaps between 50 % to 75 % (found during a grid search). Differences regarding HR and EDA features are found to be negligible.

Table 4.3: Average ranking with standard deviation for the top 24 ranked features based on relative entropy. Selected features for the final sparse feature subset are printed bold.

Feature	avg. rank	Feature	avg. rank
HR, minimum	1.0 ± 0.0	SCL, mean	13.6 ± 2.0
HR, mean	2.5 ± 0.5	EDA, mean	13.8 ± 2.1
SCR, peak amp. sum	3.6 ± 1.1	SCL, minimum	14.4 ± 1.5
HR, maximum	4.4 ± 0.6	SCR, std.	16.0 ± 1.7
EDA, minimum	4.7 ± 2.3	EDA, maximum	16.8 ± 2.9
HRV, meanNN	5.8 ± 1.5	SCR, peak amp. mean	16.8 ± 1.5
SCR, peak area sum	7.1 ± 1.7	SCL, max	17.3 ± 1.7
SCR, max	9.0 ± 2.2	**HRV, RRmed**	21.4 ± 1.8
HRV, RMSSD	10.1 ± 2.8	**SCR, peak dur. sum**	21.4 ± 0.8
HRV, SD1	10.6 ± 1.9	SCR, peak area mean	22.4 ± 2.2
HRV, SD2	12.8 ± 3.5	HRV, pNN50	22.5 ± 0.9
SCR, minimum	13.6 ± 2.0	SCR, peak count	22.6 ± 1.6

Regarding the total window size and accuracy, which is the window's length plus its overlap (Figure 4.7b), no trend can be found. This is in accordance with previous results. However, it is found that the estimation results on longer windows (around 50 s) show more variability. Peak accuracy is found for total window sizes of 50 s to 120 s, re-using between 20 s to 50 s of data (overlap).

For further analysis, the window size is therefore adjusted. It is extended from 40 s used in [274] to 60 s. This is done in accordance with the procedure in [228] and [123]. The overlap is kept and fixed at 75 %. In this way, a smoothing between the feature windows and the CW is achieved. With the given settings of a window size of 60 s and an overlap of 75 %, a new estimation of CW is possible every 15 s.

In total, 42 features have been included in the comparison. In order to reduce inter-dependencies and redundancies within the full feature set, the most important features are identified to deduce a sparse feature subset. Ranking of the features is based on Kullback-Leibler divergence (relative entropy). The topmost relevant features are based on both sensors, EDA, and HR. Furthermore, simplistic features like minimum, maximum, or mean values, derived on the raw data outperform sophisticated features.

Due to the redundancy between minimum, maximum, and mean values, only the corresponding top-ranked feature is kept. The same applies to the mean or the sum (integral) of the peak area of the SCR. Also, for other features that are correlated to each other, e.g. RMSSD and SD1 (or SD2), only the top-ranked are kept. For the final sparse feature subset, the 10 top-ranked (and not correlated) features were selected (Table 4.3).

Table 4.4: Comparison of the results (mean and standard deviation) found with the different classifiers using 10-fold CV and LOGO CV.

Classifier	Accuracy / %			
	10-fold CV		LOGO CV	
	5-class	2-class	5-class	2-class
KNN	72.60 ± 1.97	93.87 ± 1.09	30.33 ± 08.79	81.38 ± 17.25
SVM	69.35 ± 2.17	90.07 ± 1.04	37.93 ± 13.62	83.63 ± 19.30
GP	66.83 ± 1.14	94.57 ± 1.05	32.53 ± 09.51	83.68 ± 15.69
DT	52.91 ± 3.28	87.51 ± 2.17	35.80 ± 13.31	81.87 ± 17.42
NB	50.45 ± 2.13	85.10 ± 1.44	38.68 ± 13.96	80.51 ± 18.17
Mean	62.43 ± 9.00	90.22 ± 3.63	35.05 ± 3.18	82.21 ± 1.25

4.3.3.3 Classification accuracy

For evaluation, multiple classifiers are used to train models, given the selected sparse feature subset (Table 4.4). Results are obtained using 10-fold and LOGO CV. First, results from 10-fold CV are presented:

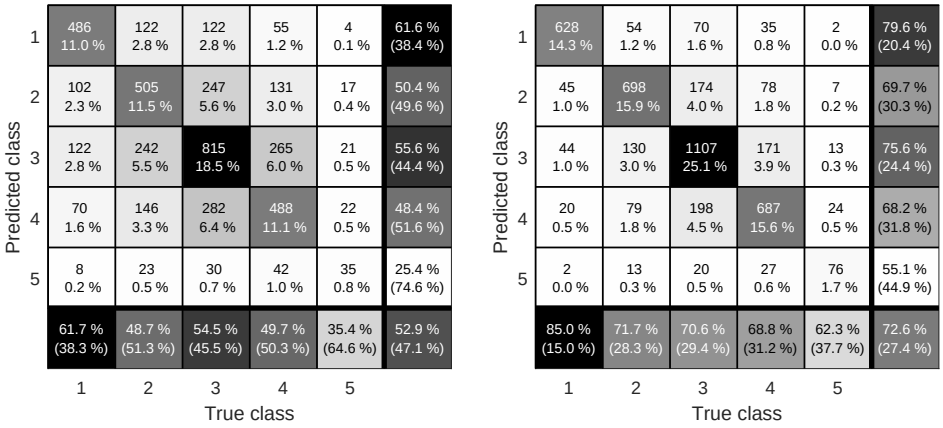
The lowest accuracy was found with NB classification (50.5 ± 2.1 %). The maximum average sensitivity is found on level 1 (60.17 ± 6.08 %). Lowest average sensitivity is found on level 5 (7.97 ± 5.21 %). Mean sensitivity considering levels 2, 3, and 4 is 48.90 ± 6.47 %. For these mid-level CWs also specificity on average is lower (80.93 ± 8.05 %) than for level 1 (90.25 ± 1.50 %) and 5 (99.60 ± 0.22 %).

For the DT, an average accuracy of 52.9 ± 3.3 % was found (Figure 4.8a). This is comparable to the results of the NB classifier. Again, the maximum average sensitivity is found for level 1 (61.92 ± 3.94 %), while the lowest average sensitivity is found for level 5 (24.75 ± 1.06 %). Also, mean specificity considering the mid-level CW (2 to 4) is low 82.19 ± 4.71 % compared to level 1 (91.64 ± 0.94 %) and 5 (98.50 ± 0.62 %).

Considering the DT and the NB classifier, the remaining models, namely GP, SVM, and KNN, provide better results in terms of accuracy. With these models accuracies are 66.83 ± 1.14 %, 69.35 ± 2.17 % and 72.6 ± 19.7 %, respectively.

Consistent with the previous results [274] with DT and NB, the lowest sensitivity and specificity are found for class 5 (GP: 28.48 ± 13.30 %, SVM: 43.51 ± 11.45 %, KNN: 55.17 ± 5.17 %). This class is underrepresented in the data set. Also, the highest sensitivity and specificity are found for CW level 1 (all classifiers except GP).

With regard to the mid-level CWs (2 to 4), a centering of the sensitivity is observed. This is, that the sensitivity is maximal for level 3, whereas the classification of levels 2 and 4 is less sensitive. Hence, the confusion is distributed (and centered) around level 3, which is also reflected in the specificity.



(a) decision tree **(b)** k-nearest neighbor

Figure 4.8: Confusion matrix for the DT- (a) and KNN- (b) based CW estimation found with 10-fold CV. Last row (y-axis) contains sensitivity (TPR) and false-negative rate (FNR) (bracketed). Last column (x-axis) contains specificity (TNR) and false-positive rate (FPR) (bracketed).

Considering this, it can be seen that misclassifications (or inaccuracies) are mainly a result of the confusion in these CW-levels, namely 2, 3, and 4. This becomes visible by comparing the results of the DT and KNN (Figure 4.8). In contrast to DT, the classes in the KNN's confusion matrix are more clearly separated from each other. Thus the KNN's accuracy is higher. In this respect, the classifiers KNN (Figure 4.8b) and SVM show a similar overall picture regarding the confusion. The same applies to the GP-based model, which, however, is less sensitive with respect to CW level 1. Therefore, its accuracy is reduced. The accuracy of KNN and SVM is 3.2 % apart, which is assumed to be negligible with respect to the variability within the CV (1.97 % to 2.17 %).

Next, results from 10-fold CV are compared to those found with LOGO. Strikingly, the mean accuracy across all models, for the fine-grained task, is reduced by 27.37 % (which corresponds to a degradation of 56.17 %). Moreover, the ranking of the methods changes. The best estimates are now found with NB, whereas the lowest accuracy is found with KNN. Moreover, the inter-classifier variability in terms of accuracy is lower using the LOGO validation (3.18 % compared to 9.00 %). These observations suggest that without personal characteristics flowing into the validation partition, not all information in the feature set can be used. This is especially true for the inter-class confusion related to mid-level CW (2 to 4).

Still, not all uncertainties are covered by the evaluated models, which is due to the confusion between the classes 2 to 4. This can easily be seen by shrinking the classification task to a binary problem. In this case, self-reported CW level 1 is interpreted

as no CW. All remaining levels are taken as present CW. With this new binary target average accuracy in 10-fold CV across all classifiers is found to be $90.22 \pm 4.06\,\%$.

Considering the binary classification task, the ranking of the tested classifiers remains mainly unchanged. However, in the binary setting, GP outperforms SVM and KNN. Again, the difference between SVM and GP (0.7 %) is minimal with respect to the variability found within the CV (1.1 %). In contrast to the fine-grained tasks, the NB and DT-based models also provided acceptable classification results. The accuracy found with NB or DT is $85.10 \pm 1.44\,\%$ or $87.51 \pm 2.17\,\%$, respectively.

The absolute distance between the 2 lowest-ranked classifiers (NB, DT) and the 3 top-ranked classifiers (SVM, KNN, GP) is lower in the binary task (6.53 %) compared to the fine-grained task (16.41 %). This indicates that the mid-level CW variation dominates the complexity of the classification. This assumption is also supported, connecting it to the results found with the LOGO CV, where the mean accuracy is $82.21 \pm 1.25\,\%$.

The best results are found with the GP or SVM (absolute difference in accuracy is 0.05 %). Compared to the 10-fold CV, that is a difference of 8.01 % or a percentage deterioration of 9.29 %. Again, this highlights the assumption made on the confusion in the mid-level CW classes.

By relaxing the fine-grained constraint, and thus separate binary targets only, the estimation is improved and comparable to that found with 10-fold CV. This result also suggests that it is possible to generalization across a more diverse set of participants, even without using personalized characteristics.

4.3.4 Discussion

Within this work, the fine-grained estimation of CW was demonstrated using commercially graded body sensors. By focusing on a fine-grained estimation based on short-term signals, the complexity of the classification task was extended. Additionally, state-of-the-art results for the binary classification task were reproduced. For the binary task, it was shown that, in comparison to the fine-grained classification task, its overall accuracy is lower. This is explained by a lower sensitivity regarding the mid-level CWs.

One possible explanation for this observed variation could be the nature of the subjective self-report. Here, a difference between the subjective perception of CW and the objective bodily response cannot be excluded (e.g. a mismatch between sensation and self-perception). It was assumed that more detailed self-reports, based on NASA-TLX ([103], sec. 4.2), could clarify this issue. Therefore, in the 2nd test run of the experiment, participants were asked to answer the NASA-TLX score in addition to the uni-modal CW Likert scale (sec. 4.2).

The analysis of the NASA-TLX score revealed that the physical workload was rated the lowest for all phases of the experiment. This is consistent with the expectation since the participants didn't need to be physically active during the experiment. The effort, mental, and performance workload items were ranked the highest in this order. In all, no differences between these groups could be identified.

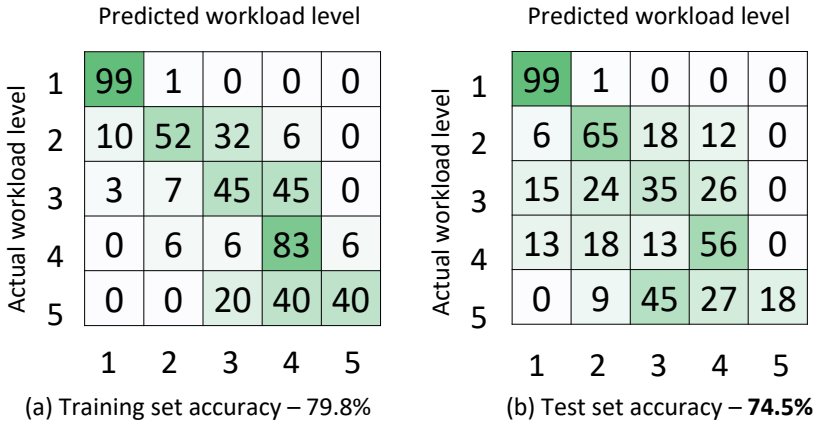


Figure 4.9: Comparison results of a 5-level CW estimation in an autonomous driving scenario, as reported by Manawadu et al. [155]. Similar accuracy and variability patterns in respect to confusion were found in this work. Reprinted with permission, ©2018, IEEE

Comparable rankings could be found for temporal demands and frustration. For both items, it was also found that the mixed memory and reaction phase was rated the highest. In general, the items of the NASA-TLX score were ranked equally for all phases, except for the Stroop test. With regard to effort and physical demand, it was given the highest rating, whereas, concerning all other items, it was given the lowest ranking.

It can be concluded that the Stroop test differs significantly from all other phases of the experiment, while the remaining phases of the experiment are interchangeable to some extent. In summary, no difference could be found between the uni-modal CW measure (Likert scale) and the median of all TLX-items. Thus, for further investigation, additional studies to examine or classify the different tasks would be interesting. Based on these insights, an experiment could then be controlled in more detail so that it concentrates on a single item (e.g. *frustration* or *effort*) only. Also, it could be helpful to challenge the participants more and thus allows to have a wider or more evenly distributed range of perceived difficulties. Besides the more detail view on the dimensions of CW, additional performance measures like error-rate or time-on-task could help to further clarify the variation in subjectively perceived CW.

Considering the observed mid-levels CW confusion, it can be inferred that both the (self-reported) target values and the predictors (EDA, HR) are affected by noise. Interestingly, there is no concluding answer whether there is noise in the (self-reported) target values in the predictors (EDA, HR) or both. Nevertheless, it can be seen that even most advanced methods suffer from this mid-level CW confusion.

An example is found in the work of Manawadu et al. [155]. They implemented a fine-grained estimation model for CW in the setting of semi-autonomous driving. The reported accuracy for the 5 leveled fine-grained estimation is similar to that presented in this chapter (74.5 %, Figure 4.9).

Strikingly, even with the deep ANN-based classification approach presented by Manawadu et al. [155], and additional sensory information based on EEG signals, the confusion for the mid-level CW (i.e. level 2-5) remains. Manawadu et al. [155] work around this "human error" by altering the ground truth CW labels and use a soft-threshold approach. With this approach misclassification in the range of ±1 level is allowed. As a result, accuracy is improved to 96.5 %. This again is in close agreement with the binary-classification result presented in this chapter (sec. 4.3.3.3).

Taking the assumption of noisy predictors and target values into account, GP was included in the comparison. This is because GPs are well-known to act as a linear smoother and, therefore, generally provide good performance in noisy settings [187]. Due to the soft-margin approach, SVM can also act as a linear-smoother, given appropriate regularization. Both methods are known to provide a good trade-off regarding the bias-variance dilemma.

Indeed, GP and SVM-based classification outperformed DT and NB classification. However, the best results were found using KNN. KNN-based classification, in turn, is known to be prone to noise (over-fitting), especially if k value is chosen very small. It was observed that without regularization by using CV, a neighborhood of $k = 1$ gave the best accuracy. In the final model, the neighbor value is still small ($k = 4$). It cannot be excluded that the high accuracy found is an effect of over-fitting to noise. This also agrees with the finding of a low accuracy using NB classification. This is because NB is known to be robust against noise by trading-off towards a bias error. Nevertheless, another reason why NB could have performed worst is the lack of independence of the features that are part of the feature set.

For this reason, it is of interest to have a more detailed view of the correlation between self-reported CW and objective measures, e.g. error rate or time-on-task. Furthermore, the question arises if using bodily reactions alone, a separation among different dimensions of CW (i.e. mental demands, effort, time pressure, frustration) is possible at all [91]. It remains an open challenge for future research to investigate and clarify on the distortion of self-reported CW toughly. To clarify this, at least a more controlled experimental setup is required. This is based on the observation that no significant difference could be found between the NASA-TLX and the simplistic uni-modal Likert scale. Hence it could be concluded that a setting that exclusively targets a single dimension, e.g. time pressure, is needed.

Nevertheless, according to the results for the 3 top-rated models, misclassification rarely exceeded more than one class (or level). Therefore, despite the lower overall accuracy, the fine-grained estimation should be favorable because it facilitates a detailed specification of the perceived CW. Although concerning the top 3 models (GP, SVM, and KNN), comparable accuracy was found, they differ in implementation details.

For instance, because SVM is more widespread compared to GP and computationally efficient implementations are commonly available, their use might be preferred. Another example is found with the KNN method, were memory or computational demands are quickly exceeded during inference compared to SVM and GP.

According to the feature analysis and classification results (sec. 4.3.3.2), sophisticated HRV features were found to be less important than simple measures (e.g. mean or minimum of HR). This could be explained due to a low internal sampling frequency of the HRM used (Polar H6, sec. 4.3.2.2). In general, a minimum sample rate of 500 Hz to 1000 Hz is suggested [154]. This is in order to guarantee that the deviation of the RR-intervals is as low as 1 ms to 2 ms. For the Polar H6, used in this experiment, sample frequency is unknown.

Additionally, most of the HRV features used are not invariant to translations (spectral features) or scaling (temporal features). This, however, can occur due to a lag in the QRS-detection (regarding the HRM used) or simply due to a shift of the mean HR. However, for earlier HRM-versions of the Polar H6 (from the same manufacturer), a sound agreement has already been reported regarding the RR-distances [92].

Moreover, features ranking revealed that the HR-based features contribute to the estimation the most. They are followed by the features based on EDA. As emphasized by Sun et al. [228], care has to be taken if HR is selected as a predictor because it is possibly influenced by means of physical activity.

Indeed, the detailed analysis based on NASA-TLX revealed significant differences in self-reported physical activity between the relaxation phase (task 1) and the more demanding tasks 3 to 5. It must be noted that the classification results could be disturbed due to this correlation. In this respect, the differences in physical activity could have caused similar effects as were already reported in relation to the touch features.

Yet, this effect would be limited to the binary classification task alone. This is because, in contrast to the touch features, with the TLX-item *physical demands*, no significant differences in-between the stages of 3 to 5 could be found. Thus, if spurious relations were learned in the models, these are most probably limited to the results of the binary classification task.

As a final remark on the comparison of HR and EDA features, it should be noted that the mobile EDA sensor differs in terms of absolute values compared to a laboratory device. Hence, sophisticated features like the peak-area or -amplitude could be distorted. This is even though the raw sensor readings were normalized participant-wise.

4.4 Conclusion and Future Work

Stress and mental health are subjects of current research worldwide. In this chapter, it was shown how physiological signals recorded by means of WBS can be used to estimate cognitive workload. The conducted experiment mimics the setting of a modern

working environment, in which the participants were asked to perform various tasks on a tablet computer.

Self-reported subjective CW was questioned directly after each task. These were used as ground truth information. In order to predict CW, EDA and HR were recorded. At first, it could be shown that the sensor readings are comparable to those taken from laboratory reference. In total, 42 features were calculated on the basis of the EDA and HR sensor. Most significant features and their ideal window sizes and overlaps were determined with an initial spot-resting approach based on 10-fold CV DTs. The identified sparse feature subset contains 10 features, which include 5 EDA-based, and 5 HR- or HRV-based features. The feature subset was then evaluated by comparing the accuracy of multiple well-established ML methods.

By employing the developed and applied models, it was shown that it is possible to distinguish between stress-free and stressful tasks. Furthermore, the fine-grained estimation with 5 levels (classes) is possible. In conclusion, a classification accuracy of 94.6 % for the binary CW estimation and an accuracy of 72.6 % for the fine-grained estimation was found.

Moreover, it was shown that these features can be calculated on a short-term basis. This is a mandatory requirement in order to set up an adaptive assistive system, which is capable of balancing the complexity of a given task accordingly to the user's cognitive capacity.

It was furthermore observed that the subjective estimation of CW, which serves as ground truth, is affected by uncertainty. Similar results were described independently to this work in later published studies [155]. This again highlights the necessity to compare subjective and objective markers such as the cortisol level in future work. In this regard, preliminary work on the integration of chemical sensors in WBS has already been presented [26, 161]. However, before these and comparable sensors can be reliably integrated into WBS, several challenges still have to be solved, which include long-term stability and device-skin adherence [161].

Apart from these potentials for future work, the added value of WBS is already evident today. In psychology, the so-called *experience sampling* method has been discussed for some years [70]. A similar concept is found with *ecological momentary assessments* [218]. Both aim to capture subjective impressions or emotions as directly as possible.

Today, questionnaires are often no longer designed on paper and retrospectives tools but are collected as directly as possible, e.g. using a smartphone (or as in this work a tablet computer). In this way, common biases e.g. a *recall bias* or *recency effect*, could be counteracted[11]. Although using a smartphone, a question (as part of a questionnaire) can be asked right after an event, the measurement of physiological parameters takes

[11] These effects are given here for the sake of illustration. With *recall bias*, false memory reports are denoted in general. The *recency effect* describes a phenomena where the last remembered information is more present in the short-time memory compared to an information which was remembered earlier. The list of memory or cognitive biases in psychology is long and an independent branch of research. An overview can be found in [60].

place immediately, which is parallel to an event. Using WBS thus allows gathering insight into the temporal dynamics of emotions and subjective perception in particular.

The use of WBS forms the basis for new research methods in this field, the applicability, and effectiveness of which was demonstrated in this chapter. In view of decreasing costs for mobile computing and wearable sensors, a further spread can be expected. For future work, this means that WBS will open up new research questions. It is moreover conceivable to use the sensors as indirect markers. In this sense, spontaneous changes of activity, such as the increase of HR or sudden onset of movement, could be used as triggers to initiate user surveys. These triggers could then be useful to look at future psychological questions more closely and from a new perspective, i.e. under the view of dynamics in activity behavior.

Table 4.5: Selected methods used for feature extraction.

Signal	Category	Function	Definition
Heart rate (HR)	time	average	$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$
		standard deviation	$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2}$
Heart rate variability (HRV)	time	average	$1/n \sum_{i=1}^{n} a_i$
	temporal	skew	$1/n \sum_{i=1}^{n}(x_i-\mu)^3$
		kurtosis	$1/n \sum_{i=1}^{n}(x_i-\mu)^4$
	heart rate variability	NN50	$\sum_{i=1}^{n-1}(x_i - x_{i+1} > .05)$
		RMSSD	$\sqrt{1/n \sum_{i=1}^{n}(x_i - x_{i+1})^2}$
		SDSD	$\sigma((x_1 - x_2)\ldots(x_{n-1} - x_n))$
		SD1	$\sqrt{.5 \cdot SDSD^2}$
		SD2	$\sqrt{(2 \cdot SDSD^2) - (.5 \cdot \sigma^2(x))}$
		SD12	$SD1/SD2$
	spectral	VLF	energy 0.00 Hz to 0.04 Hz
		LF	energy 0.04 Hz to 0.15 Hz
		HF	energy 0.15 Hz to 0.40 Hz
		nLF	normalized energy ($LF/LF+HF$)
		nHF	normalized energy ($HF/LF+HF$)
		LF/HF	LF/HF
	geometric	TRI	relative frequency of mode
		TINN	triangular fit to histogram
		rrHRV	median Euclidean distance of successive RR towards their mean
Electrodermal-activity (EDA, SCR, SCL)	geometric (peak)	count	number of peaks
		amplitude	amplitude of peak (max)
		duration	distance between the two minimums surrounding a peak
		area	integral between the two minimums surrounding a peak

5 Monitoring Physical Activity

In the course of this chapter, the topic of physical activity (PA) estimation utilizing WBS is addressed. The focus is on the application in a professional context, i.e. firefighting.

After a brief motivation (sec. 5.1), existing reference and alternative methods to measure PA (or EE) are outlined (sec. 5.2). Thereafter, recent multi-modal approaches to estimate EE utilizing WBS are introduced (sec. 5.3). Additionally, aspects of the calibration of WBS are addressed (sec. 5.4). Parts of these chapters were previously published in [272, 275]. In the following, a multi-modal and robust model to estimate the PA of firefighters is introduced (sec. 5.5).

5.1 Background and Motivation

Quantifying PA is practiced both in private and in professional contexts [76]. Professional examples can be found with soldiers, athletes [140], or ergonomists [179] (sec. 2.2.1). Today, fitness trackers or activity trackers are used in private for self-measurement (an extreme can be found in the quantified-self movement) [5].

While private individuals and athletes may initially have an interest in recording their activities or performance out of a selfish interest, the relevance for the general public results from the positive health effects of PA. This is because PA and cardio-respiratory fitness are regarded as key factors in avoiding numerous diseases of affluence [179, 222, 237]. These include obesity, type 2 diabetes, cardiovascular diseases, or mental health [247]. An active way of life thus promotes the health of each individual and therefore is of major relevance towards public health [240].

The relevance of PA goes beyond private life and is just as important for other dimensions of life, such as work. However, different user groups also have different expectations and requirements. For instance, validity and cost, but also comfort and functionality, are dimensions that conflict with each other [119]. Here lies the basic idea of the design and application of WBS and other wearable devices, which is to construct sensors, technical systems, and algorithms in such a way that these conflicting expectations are optimized in parallel. Regarding costs and time, WBSs already have an advantage over laboratory methods. Specifying and maximizing validity, on the other hand, is the topic of current research work [9, 78].

Estimating PA utilizing WBS is a widespread technique with a long history (sec. 2.2.1). The approaches used have evolved from simple linear methods based solely on accelerometer data [167] or heart rate [220] towards more sophisticated methods fusing multiple sensor data using advanced non-linear ML methods [72]. Furthermore, recent

work engaged architectures that use several activity-specific models rather than develop one single stand-alone model [12, 72].

Since then, quantifying one's own PA has emerged into everyday life, having fitness trackers or specific *Apps* for smartphones and smartwatches available on the market [185]. However, for applications apart from being a leisure activity, the validity of the predicted outcome is more crucial and still a relevant research topic [172].

5.2 Operationalization of Physical Activity

The operationalization of physical activity (PA) can be subdivided into the branches of human activity recognition (HAR) and estimating energy expenditure (EE).

HAR has its focus on behavior, and in this understanding, PA is a sum of different activities and their duration. Often, the task is to detect so-called activities of daily living (ADL), e.g. walking stairs, dancing, or sweeping the floor. The intensity of the same can also be of interest. Indeed, HAR can be seen as an abstraction of EE. This is because having HAR combined with information on duration and intensity can be used to estimate EE. In fact, recording activity diaries, which are created manually through observation, is a commonly used method to quantify PA and EE.

A comprehensive guideline of specific activity codes and corresponding values for EE can be found in [4]. It contains intensity information on 821 different activities of daily living (ADL). It is structured broadly into activities such as cycling, dancing, or housework. For many activities, there are also more graduated subdivisions. For example, different speeds are distinguished for cycling or different details for housework such as wiping, sweeping, making the bed, or watering flowers.

In this work, however, the focus is on accurate EE estimation utilizing WBS. Therefore, at first, a brief overview of reference and secondary methods is given (sec. 5.2.1 - 5.2.2). A more comprehensive overview can be found in [113, 137]. In the subsequent, metrics used to quantify EE are compared (sec. 5.2.3).

5.2.1 Reference Methods

In the following, reference methods to obtain ground truth data for EE estimation are briefly discussed. These presented methods can be subdivided into direct (sec. 5.2.1.1) and indirect methods (sec. 5.2.1.2 - 5.2.1.3).

5.2.1.1 Direct Calorimetry

Metabolic energy is converted to external work (e.g. kinetic or potential energy transfer), or energy is stored in the organism (e.g. heat or chemically bound energy) and metabolic end-products (e.g. feces). Otherwise, all energy entering the organism is

finally transferred into heat (sec. 2.1.1). This includes the internal work (e.g. heart muscle activity), as well as all metabolic processes. [111]

With direct calorimetry (DC), this very amount of heat loss of the body through conduction, convection, and thermal radiation (evaporative heat loss is measured separately) is measured. It is considered the most exact method to measure EE. Also, the temporal resolution of modern systems is good. The difficulties lie in the practical implementation. [160, p. 179]

One difficulty of DC is to construct a thermally isolated yet observable space. Having an isolated environment is the only possible way to measure the emitted heat accurately. However, complete isolation is not possible since, for example, the respiratory gases in the closed chamber must be continuously renewed. It is even more challenging to design mobile systems that meet the insulation criterion. [125]

Since the participants have to be enclosed in a chamber (whole-room calorimeters) or suit, the practical implementation is also more difficult. Its application involves a great deal of controlling technology and is expensive. For this reason, it is rarely used for practical experiments, e.g. in sports science. [125]

5.2.1.2 Doubly Labeled Water

Instead of observing heat transfer directly, observation of the metabolism can be used to measure EE. The rationale behind this is that all produced heat is eventually leaving the body. Thus heat loss can be measured indirectly by observing or estimating the heat production (instead of heat loss sec. 5.2.1.1).

Because oxidation contributes the most to energy metabolism, the oxygen uptake rate is a suitable parameter to measure EE indirectly. One such indirect method is doubly labeled water (DLW) [160, p. 185]. With the DLW method, isotopic labeling of water is used, which is taken orally by the participant. The isotopes used are ^{18}O (normal oxygen: ^{16}O) and ^{2}H (deuterium; normal hydrogen: ^{1}H). Neither of the two isotopes is radioactive. However, they can be easily distinguished from other isotopes by other technical means.

In application, the water drunk by the participant mixes with the normal body water and spreads evenly throughout the body. While ^{18}O can be re-integrated into both water and carbon dioxide, ^{2}H remains in water only and is excreted in the urine. This ^{2}H excretion will always be higher than that of ^{18}O. The difference is a measure of CO_2 formation and thus provides information about the exhaled air. Because urine samples have to be collected over multiple days, no short-term changes in EE can be measured, but only means across multiple days. Also, DLW is associated with high costs.

5.2.1.3 Indirect Calorimetry

Similar to the DLW method, indirect calorimetry (IC) is used to observe metabolic heat production. With IC, however, respiratory gas exchange is measured. More precisely,

the volume of oxygen uptake (VO_2) and produced carbon dioxide (VCO_2) is measured. The quotient of the two measures (VCO_2/VO_2) is known as the respiratory quotient (RQ), which can be used to conclude the metabolic activity (sec. 2.1.2), [160, p. 180]

For instance, the metabolism of fatty acids produces significantly less carbon dioxide than the metabolism of carbohydrates. In turn, the observed RQ is low (0.7). In the opposite case of predominant oxidation of carbohydrates, the RQ is higher (1.0).

If EE is covered by oxidation of carbohydrates only, the caloric equivalent of oxygen is 21.1 kJ/LO_2. The Weir-equation (eq. 5.1, [241]) can be used to calculate EE[1].

$$EE = (1.106 \cdot RQ + 3.941) \cdot VCO_2 \cdot 4.1868 \qquad (\text{kJ}) \qquad (5.1)$$

Technical implementations of IC are often found as spirometry devices (breathing mask). They are available as mobile devices, consisting of a breathing mask, airflow sensor, and gas analyzer. The participants breathe through the mask, and samples of the inhaled and exhaled air are taken breath-by-breath (also known as BxB). Also, mixing-chamber systems exist, which are known to be more accurate compared to BxB-systems [28]. With those systems, the complete exhaled air is sampled. Because of the relatively good validity of IC, its mobility, and low costs compared to DC and DLW, the method is often used in practice to measure EE. [105]

5.2.2 Secondary Methods

Another starting point to measure EE is the use of WBSs. They can be used to unobtrusively measure bodily functions (e.g. respiration rate or heart rate). These can serve as an abstract view on the EE, i.e. they are used as a proxy (sec. 2.1.2).

Moreover, using accelerometer signals allows to (at least partly) observe external work, i.e. movement of the body or muscle work. Both techniques can be used as secondary methods to estimate EE and are presented in the subsequent.

5.2.2.1 Heart Rate

HR is a prominent predictor for EE [1] because HR correlates with oxygen consumption (sec. 2.1.3). However, inter-individual differences (minimum or maximum HR) must be compensated [139]. In addition, the relationship between HR and EE is not strictly linear but, given very low or extreme PA, non-linear relation is observed.

It is also known that HR reacts differently to different activities [21]. This is, e.g. because of changes in stroke volume due to different postures [56], the influence of external heat [6], or because different sized muscles are used in certain activities (e.g. comparing leg and arm work [225]). Moreover, non-linear response of HR, e.g. in

[1] Regarding IC, it is typically assumed that no proteins are metabolized, although corrections based on urinary measurements are available.

sedentary situations due to psychological strain, must be taken into account [56, 89] (chapter 4).

A first method to estimate the TEE by using HR was presented by Spurr et al. [220]. They made use of a linear model (eq. 5.2, $r = 0.87$, SEE $= 0.91\,\mathrm{MJ}$)[2]. To overcome certain shortcomings (i.e. inter-participant variation and nonlinearities) Ceesay et al. [56] proposed the "FLEX"-HR method. The "FLEX" value is an average of the highest HR measured while the participant is standing and the lowest HR while the participant is walking. With this approach, HR below the "FLEX" value is assumed to equal the RMR, while a HR above is used to estimate EE (e.g. by regression). To increase robustness for individual differences, Keytel et al. [126] proposed a model that included anthropometric features as well (eq. 5.3 - eq. 5.4).

In all, HR can only provide an abstraction of EE, because it acts as a proxy (estimating blood flow and thus oxygen transport to the muscles).

$$EE = (0.92 \cdot h + 1) \cdot 1000 \cdot 1440^{-1} \tag{5.2}$$

$$EE_m = -95.7735 + (0.271 \cdot a + 0.394 \cdot m + 0.404 \cdot o + 0.634 \cdot h) \tag{5.3}$$

$$EE_w = -59.3954 + (0.274 \cdot a + 0.103 \cdot m + 0.380 \cdot o + 0.450 \cdot h) \tag{5.4}$$

EE	-	energy expenditure	kJ/min	h	-	heart rate	beats per minute
EE_m	-	energy expenditure, male	kJ/min	a	-	age	year
EE_w	-	energy expenditure, female	kJ/min	m	-	body mass	kg
				o	-	VO_2max	mL/kg

5.2.2.2 Other Physiological

Besides HR, other physiological parameters allow determining EE, as well. For instance, respiratory rate (RR) can be used as an alternative proxy for oxygen consumption. However, it is less accurate than HR [12, 275]. This is explained by the wider variety of modulating respiratory minute volume, compared to the modulation of the heart's stroke volume (sec. 2.1.3). Also, skin temperature [133], skin humidity [12], or heat-flux (in respect to the skin and the environment) [67, 149] can be used as a proxy towards the direct observation of heat loss (sec. 5.2.1.1). However, those methods show a lower correlation compared to HR as a predictor for EE [12].

5.2.2.3 Accelerometry

Up to this point, all EE-estimation methods described relied on the measurement of body heat loss either directly or indirectly through the observation of the metabolism or

[2]Here, r is the correlation coefficient and SEE is the standard error of estimates [220]. No further specifications are provided.

other bodily functions. An alternative approach towards the estimation of EE is based on the measurement of physical work, e.g. the contraction of muscles or movements of the body. An objective approach is the use of accelerometers attached to the body.

Similar to pedometers (sec. 2.2.1), initial accelerometry approaches to estimate EE made use of so-called *counts*. Both, linear (eq. 5.5, [88]) and non-linear methods were proposed (eq. 5.6, [58]). To calculate a *count*, typically, the number of peaks in the accelerometer signal or the integral (cumulative sum) is calculated. It was soon found that those models underestimate EE of ADL and at the same time, overestimate low-intensity activities. To obtain improved estimates, linear combinations of regression models were proposed for different intensity classes (eq. 5.7, [69])[3].

$$EE_k = (0.00094 \cdot c_c + 0.1346 \cdot m - 7.37418) \cdot 4.184 \tag{5.5}$$

$$\begin{aligned} EE_j &= a \cdot \left(\sqrt{(c_x^2 + c_y^2)}\right)^{p_1} + b \cdot c_z^{p_2} \\ a &= 12.81 \cdot m + 843.22 \cdot 1000^{-1} \\ b &= 38.90 \cdot m - 682.44 \cdot g + 692.5 \cdot 1000^{-1} \\ p_1 &= 2.66 \cdot m + 146.72 \cdot 1000^{-1} \\ p_2 &= 3.85 \cdot m + 968.28 \cdot 1000^{-1} \end{aligned} \tag{5.6}$$

$$EE_k = \begin{cases} 1.0 & : c_a \leq 10 \\ 2.550956 \cdot \exp(0.000137466 \cdot c_a) & : c_a > 10 \wedge Z(c_a) \leq 13\,\% \\ 1.466072 + 0.2107550 \cdot \ln c_a & \\ \quad - 0.0595362 \cdot (\ln c_a)^2 & : c_a > 10 \wedge Z(c_a) > 13\,\% \\ \quad + 0.0157002 \cdot (\ln c)^3 & \end{cases} \tag{5.7}$$

EE_j	-	energy expenditure	kJ/min
EE_k	-	energy expenditure	kcal/min
c_c	-	accelerometer counts from “CSA Model 5032” [88]	1/min
$c_{x\|y\|z}$	-	accelerometer counts from “Tritrac” [58] (horizontal)	1/min
c_a	-	accelerometer counts from “Actical” [69]	1/min
Z	-	std. divided by mean from 4 consecutive 15 s epochs of c_s	%
m	-	body mass	kg
g	-	sex	1 - men, 2 - woman

It was eventually understood that the direct mapping of physical work (whole-body accelerometry) to EE is misleading [148]. One reason for this is that no uniform *efficiency coefficient* between physical work and EE exists. Also, not all muscles are

[3] In the original publications [69, 88] EE is given as metabolic equivalent (MET) (eq. 5.5 and 5.7). Here it is assumed that 1 MET equals1 kcal.

observed. Instead, measurements are based on a single sensor (or a limited set of sensors) only. This leads to an oversimplified view of the body as a single mass.

Furthermore, even using multiple accelerometers on all joints of the body would not guarantee to result in a reasonable abstraction of muscle work. This is because the accelerometers are usually loosely mounted on to the skin, which is a soft-tissue. Therefore, damping in the resonance of the measured acceleration signal is present. Thus, the real acceleration of the human body cannot be measured. [175, 245].

To compensate for such errors, several models based on sophisticated regression techniques have been presented in the literature, e.g. SVM, DT, ANN, or ensemble methods [124, 144]. Besides different inference methods, positioning of the sensor [83] or the use of distributed sensor systems is discussed [14].

5.2.3 Metrics

With reference methods (sec. 5.2.1), the EE is typically given as joule (sec. 2.1.1). Depending on the method used, other metrics are also used [113, 208]. An example is found with the abstraction of EE as *counts* derived from acceleration data (sec. 5.2.2.3).

In the case of using IC (sec. 5.2.1.3), often, MET values for oxygen consumption are given as a measure of EE (eq. 5.8). The MET corresponds to the amount of $3.5\,\mathrm{mL}/(\mathrm{kg\,min})$ or $3.15\,\mathrm{mL}/(\mathrm{kg\,min})$ of oxygen consumed for male or female participants, respectively. As a simple approximation, 1 MET is often equated with 1 kcal.

Using MET instead of calculating EE (sec. 5.2.1.3) furthermore avoids possible error propagation. Besides, also a normalization takes place since the oxygen consumption is weighted according to participants' body mass and sex. In this way, EE is normalized and becomes more comparable among different participants.

Instead of normalizing using body mass, EE can be normalized by means of BMR. In this way, EE is given as a multiple of the BMR (eq. 5.9). This is referred to as the physical activity ratio (PAR)[4]. Instead of using BMR, normalization using RMR is also found in the literature. This is because RMR is easier to measure.

$$\mathrm{MET} = \begin{cases} \mathrm{OUR} \cdot (3.5\,\mathrm{mL} \cdot m)^{-1} & \text{(in relation to oxygen)} \\ \mathrm{EE} \quad \cdot (1\,\mathrm{kcal} \ \cdot m)^{-1} & \text{(in relation to heat)} \end{cases} \tag{5.8}$$

$$\mathrm{PAR} = \mathrm{EE} \cdot \mathrm{BMR}^{-1} \tag{5.9}$$

EE	-	energy expenditure	kJ/min	BMR	-	basal metabolic rate	kJ
PAR	-	physical activity level	1/min	OUR	-	oxygen uptake rate	mL/min
MET	-	metabolic equivalent	1/(min kg)	m	-	body mass	kg

PAR and MET are both suitable metrics to reflect EE because they allow for easy comparison among different participants. In addition, it must be taken into account

[4] Besides physical activity ratio (PAR) the expression physical activity level (PAL) is used, if the total energy expenditure (TEE) for a period of 24 h is referred

that using secondary methods without normalization otherwise addresses 2 problems at the same time: Firstly, the measurement of BMR and secondly the measurement of PAEE. Since the BMR is exclusively determined by anthropometric properties (e.g. height, weight, sex), and secondary-methods cannot measure these properties, the BMR should be removed.

Nevertheless, EE can still be calculated from the PAR if the BMR is known. Even if the real BMR is unknown, methods exist to estimate it. One standard approach, still used today, was already proposed in 1918 by Harris et al. [102] (Harris-Benedict equation: eq. 5.10 - 5.11). It is based on body mass, height, age, and sex.

A revised model, using the very same anthropometric features, but fitted on a broader sample, was proposed by Mifflin et al. [165] (Mifflin-St Jeor Equation: eq. 5.12). Likewise, estimation for the RMR in terms of oxygen consumption has been proposed [46]. These can be used to calculate EE from MET. Obtaining closer estimates is possible, but requires additional data, some of which is difficult to collect (e.g. fat-free mass) [165].

$$\mathrm{BMR_m} = 57.36 \cdot w + 20.93 \cdot h - 28.47 \cdot a + 278.42 \tag{5.10}$$

$$\mathrm{BMR_w} = 40.19 \cdot w + 7.54 \cdot h - 19.68 \cdot a + 2742.35 \tag{5.11}$$

$$\mathrm{BMR} = 10.0 \cdot w + 6.25 \cdot h - 5.0 \cdot a + 5 - 166 \cdot g \tag{5.12}$$

$\mathrm{BMR_m}$	-	BMR, male	kJ	$\mathrm{BMR_w}$	-	BMR, female	year
a	-	age	year	h	-	body height	cm
w	-	body mass	kg	g	-	sex	0 - male, 1 - female

5.3 State of the Art: Multi-Modal Models

Uni-modal methods to estimate EE, based on HR (sec. 5.2.2.3) or accelerometer data (sec. 5.2.2.1) have certain drawbacks in the application. On the one hand, HR tends to overestimate EE in sedentary settings. On the other hand, accelerometer-based methods fail to generalize for different types of motion or activities. This applies in particular for activities that were not part of the sample used to build the corresponding model. Also, they are incapable of observing motion intensity (e.g. carrying loads) [148]. For both reasons, they tend to underestimate EE. To overcome these shortcomings, modern approaches use both acceleration and physiological markers in combination.

In the work of Strath et al. [225], a pedometer was used to distinguish between leg- and arm-work. For each scenario, a different regression (each based on HR) was used to estimate EE. Combining the pedometer and the HRM, the agreement of EE estimation compared to IC could be improved from $R^2 = 0.54$ (pedometer alone) or $R^2 = 0.67$ (HR alone) to $R^2 = 0.81$. The data set includes data from 30 participants (16 male, 14 female), who were measured for 15 min each. In proceeding work, it was shown that a combination of personalized HR and movement features further improves

the estimation [43]. The agreement reported is up to $R^2 = 0.96$. Here, walking and running activities were examined, from a total of 20 participants (11 male, 9 female).

In more recent work, ML methods are used instead of classical regression. For instance, Lin et al. [142] presented an ANN-based model, which used HR features and data from 3 accelerometers, which were placed on the wrist, waist, and ankle. Activities in their data set contained, sedentary (e.g. lying, sitting, standing), moderate (e.g. walking, bicycling), and vigorous (e.g. walking stairs, running) intensities. A total of 26 participants (20 male, 6 female) participated in the experiment. They classified these activities using DT-based models and trained different ANNs-based regressions using HR and acceleration features for each activity. They report high agreement (10-fold CV) for sedentary ($R^2 = 0.93$), or moderate and vigorous activities ($R^2 = 1.00$).

The effectiveness of combining HR, RR, and accelerometer data was again examined by Altini et al. [12]. Also, other physiological markers were investigated, such as humidity, temperature, or EDA. In their experiment, a total of 16 participants were involved (12 male, 4 female). They summarized that HR is the best physiological marker to estimate EE ($R^2 = 0.93$). Regarding other physiological markers, weak correlations are reported (RR: $R^2 = 0.76$; EDA: $R^2 = 0.72$; skin humidity: $R^2 = 0.70$). However, they work well in combination with the acceleration sensor data because they enable separation of sedentary or active behavior. In general, Altini et al. [12] recommend combining acceleration sensor data with physiological predictors because they are "complementary" to each other. In addition, it is argued to position the acceleration sensor close to the body's center of mass because this position is most suitable for a generalized estimation of EE.

In preliminary work [275], also a nearly perfect agreement was found using accelerometer, HR, and respiration features ($R^2 = 0.97$) with a total of 15 participants (10 male, 5 female). That is for inference based on a MARS-based model trained on sedentary, walking, and running activities. Furthermore, the presented model was capable of precisely predict EE even if not all data is available. More precisely, this was demonstrated for the case that HR is missing due to a corrupted ECG signal because of motion artifacts, which often occur when the participant is running fast.

However, the problem of estimating EE is not yet completely solved. This is because, if the validation rules are tightened, it can be seen that the models do not generalize across all participants nor all activities. With the example of own preliminary work [275], the agreement decreases from $R^2 = 0.93$ to $R^2 = 0.82$, if the model is validated against new unseen data. The lesson learned from this experiment was that the challenge in PA estimation lies in the use of specific features, which allow generalizing across different individuals and activities.

One common method to cross out inter-individual differences concerning HR is the FLEX-HR method described in (sec. 5.2.2.1). In recent work, methods comparable to the FLEX-HR approach were proposed, which act as an automated calibration procedure. Therefore, running speed (based on acceleration) and HR are sampled in free-living conditions and later on used to build up individual calibration models

[15]. Besides, anthropometric features are often embedded into the models, e.g. body weight, maximal HR, or cardiorespiratory fitness (CRF). Regarding CRF, Altini et al. [16] showed that it could further improve the estimation of PA up to 18.2 %.

In the work of Altini et al. [16], not only inter-individual differences were considered but also activity-specific models were used. They furthermore showed that the error in estimating EE could successfully be reduced using activity-specific models (e.g. one regression model per activity) [13]. In this way, RMSE was effectively reduced by 57 % (1.51 kcal/min to 0.86 kcal/min).

This idea was later adopted by Gjoreski et al. [95]. Their data set is available to the public and contains data from 10 participants. They made use of a special partition (Standard Deviation Reduction) to construct ensembles of base learners fitted to sub-partitions of training data. In this way, 49 models were trained not only on different activities but also on different intensities of these activities. The best ensembles used ANN as base learners (RMSE = 0.850 MET, MAE = 0.613 MET).

Based on the very same data set, similar results (MAE = 0.52 MET) are reported even with simplified split criteria, sorting activities into sedentary, and moderate to vigorous activities only [72]. For the base learners, they used 2 multiple-regression models for each of these groups. This result is in agreement with the suggestions of Altini et al. [14], who recommend a partition into sedentary (static values for each: lying, sitting, and standing) and multiple activity-related regression models (whole-body motion, walking, biking and running). Even without any specific grouping method, similar results are reported using boosted DT (RMSE = 0.970 MET, MAE = 0.709 MET) [55].

Constructing such non-group-specific models is essential because the success of the partitioning methods ([72, 95]) directly depends on the number and diversity of different activities within the data set. The detection of individual activities thereby poses a separate challenge. Consequentially, the intermediate step increases the risk of systematic measurement deviations (due to misclassification) and over-fitting of the models (too fine partitioning, no generalization)[5].

At this point, it could be concluded that sophisticated ML methods are needed to successfully predict EE, as demonstrated in the literature [16, 55, 72]. However, recent work [146, 169] again highlighted the effectiveness of more straightforward approaches. For instance, in the work of Lu et al. [146], a small ANN using 5 neurons in the hidden layer, and 4 features (HR, RR, wrist and thigh acceleration) was presented. They also reported a high degree of agreement ($R^2 = 0.92$), however, on a self-created data set.

5.4 Compliance and Calibration

In multi-modal EE estimation, advantages of physiological predictors are combined with objective observation of motion (e.g. acceleration, sec. 5.3). However, compliance

[5] Indeed, handling misclassification of unknown activities and transitions in-between certain activities is an active branch of HAR research [193]

and calibration issues remain, which was shown in [272]. It was highlighted that valid measurement of vital signs and reliable measurement of motion depends on prior calibration and appropriate use of the WBSs. This involves the correct attachment and application of a WBS, as well as the pre-execution of calibration or validation steps.

Here, a summary of the signal disturbances and countermeasures presented in [272] is given because they were reused for the experiment presented in sec. 5.5.

5.4.1 Signals and Disturbances

To obtain reliable sensor readings, a variety of disturbances have to be considered. Taking the functional principle of the sensors into consideration, the necessary steps to ensure accuracy and precision can be determined. In the subsequent, an overview of common disturbances affecting the BG-V4.2 is outlined.

5.4.1.1 Electrocardiogram

Typical disturbances affecting the ECG include electrical interference, poor electrode contact (and skin conductance) as well as the electrode's positioning [90].

Examples concerning electrical interference are found with the line noise (AC noise 50 or 60 Hz) or high-frequency noise from the electrical activity of the skeletal muscles (EMG noise). In order to remove such electrical interference, analog filters can be used [90]. In BG-V4.2 and BG-V5, an analog active low pass filter of 4th order is used (cut-off frequency 33.86 Hz). To remove the baseline wander, an additional high pass filter is applied (cut-off frequency 1.56 Hz).

Dry electrodes or insufficient skin contact are other common reasons for measurement errors. Regarding textile electrodes, which are commonly used in chest straps, proper moistening is required. Additionally, a minimal tension of the chest strap is needed to ensure good skin contact. Setting the chest strap's tension is also relevant in terms of motion artifacts, which often occur during vigorous physical activity [173]. Obviously, the positioning of the electrodes or their contact cannot be corrected by the device itself. Therefore, these prerequisites must be checked by the experimenter or the wearer (user) before an experiment is carried out.

5.4.1.2 Accelerometer

Considering motion sensors, the initial calibration by the manufacturer guarantees precise measurements. Also, random and systematic errors are relatively small. For instance, the random observational error of the *LIS331HH*[6] accelerometer integrated into the BG-V4.2 is in the order of magnitude of 10^{-3} g. Likewise, systematic observational errors are bounded to approximately 10^{-4} g (e.g. caused by temperature

[6] STMicroelectronics, LIS331HH `https://www.st.com/resource/en/datasheet/lis331hh.pdf`

fluctuation). However, these systematic observational errors are typically dominated by an orientation offset. This offset results from a misalignment between the sensor's frame and the body's frame of the wearer. The latter depends on how and where the chest strap was applied to the wearer's body [239, 246].

5.4.2 Calibration and Noise Detection

The impacts of noise or other disturbances are widely addressed in recent research. Likewise, methods to detect or minimize the effects of disturbances are proposed:

Concerning HR estimation, the effect of noise in the ECG-signal was investigated by Friesen et al. [90]. Within their work, they demonstrated the vulnerability of common algorithms used to estimate HR, in case the underlying ECG-signal is affected by different types of noise. Furthermore, concerning mobile HR acquisition, Nikolic-Popovic et al. [173] explain the effect of motion artifacts for HR variability estimation. In order to detect noisy signals, and thus to prevent false alarms or misleading information, several methods are known to detect the quality of the ECG-signal [195].

Likewise, the calibration of accelerometers used for physical activity monitoring is important. In this respect, Wang et al. [246] examined the impact of the orientation error of an accelerometer. They showed that an orientation error greater than 3° adversely affects the PA estimation. Similarly, Alinia et al. [10] examine a scenario in which the position of a WBS was interchanged. They demonstrated that without knowing the real position of the WBS, accurate estimation of physical activity was impossible. This was again confirmed on the example of HAR by Yurtman et al. [265], who found deviation in accuracy up to 18.8 %. An overview of different approaches to calibrate accelerometers is given by Won et al. [259].

Taking these examples, it can be argued that without a precedent check-up and calibration of a WBS, reliable data is unobtainable. As exemplary pictured above, this applies to accelerometers and ECG recordings. Nevertheless, other sensory elements can be affected as well (e.g. respiration sensor [269]).

5.4.2.1 Heart Rate Validation

Various methods exist to assess the quality of an ECG recording [73]. However, these methods are often developed for clinical investigations and are not optimized for efficiency. Aiming towards an implementation for a WBS, which offers limited resources, less complex solutions are preferable. In addition, the use-cases associated with WBSs application typically require valid HR estimation only. A medical examination or quality criterion is therefore not necessary.

In order to obtain the ECG's quality in real-time, the solution presented in [138] was adopted and tuned. In [138], multiple weak metrics are combined into one strong predictor applicable for clinical usage. Here, this approach was partly reused, but restricted to use only one of the weak predictors, which is the signal's kurtosis.

Because kurtosis is a measure of the probability distribution's *spikiness*, it is well suited to distinguish a valid ECG from white noise. Yet, it is not applicable to detect scattered spikes, which occur due to motion artifacts (chest strap temporally loses contact with the skin). Therefore, an additional rule was added, which marks the signal as invalid if its range exceeds 75 % of the total measuring range (3072 LSB; 12-Bit resolution). The final quality measure combines the 2 decision rules (eq. 5.13).

$$\text{Quality}_{\text{ECG}} = \text{K}\left(\text{ECG}\right) \geq 5.4 \;\wedge\; \text{R}\left(\text{ECG}\right) \leq 3072 \tag{5.13}$$

$$\begin{aligned} \text{Kurtosis:} \quad & \text{K}(X) = \quad \tfrac{1}{n}\sum_{i=1}^{n}\left(\tfrac{x_i-\bar{x}}{\sigma(X)}\right)^4 \\ \text{Range:} \quad & \text{R}(X) = \quad \max(X) - \min(X) \end{aligned}$$

To test its feasibility, it is validated against the *PhysioNet MIT-BIH Noise Stress Test Database*[7] [96]. Each time series in the data set contains 50 % of noise-free and 50 % of noisy data. As a result, it is found that the quality measure is suitable to detect large disturbances with a signal-to-noise ratio (SNR) lower than 6 dB (Table 5.1). In the presence of almost undisturbed signals (SNR $\geq$ 18 dB), no disturbances are detected. Thus, the signal is marked as valid. Correspondingly, the accuracy and the positive predictive value considerably drop. However, given a SNR greater than 6 dB, accuracy is at least 96 %.

It becomes clear that only prominent disturbances are detected with the presented approach. Still, the number of false alarms (due to false negatives) is effectively limited. Therefore, non-acceptable ECG records can be detected and falsely calculation of the HR can be prevented. Moreover, the metric is a valuable tool for the experimenter and wearer to evaluate the ECG recording in advance to an experiment.

Table 5.1: Accuracy (ACC), positive (PPV), and negative predictive value (NPV) of the ECG-quality prediction, evaluated against PhysioNet MIT-BIH Noise Stress Test Database.

	data set: 118			data set: 119		
SNR	ACC	PPV	NPV	ACC	PPV	NPV
No noise	95 %	100 %	95 %	100 %	100 %	100 %
−6 dB	96 %	100 %	92 %	100 %	100 %	100 %
0 dB	96 %	100 %	92 %	100 %	100 %	100 %
6 dB	96 %	100 %	92 %	97 %	93 %	100 %
12 dB	93 %	95 %	92 %	64 %	28 %	100 %
18 dB	76 %	60 %	92 %	50 %	0 %	100 %
24 dB	51 %	10 %	92 %	50 %	0 %	100 %

[7] PhysioNet MIT-BIH Noise Stress Test Database, `https://physionet.org/content/nstdb/1.0.0/`

5.4.2.2 Acceleration and body posture calibration

Regarding acceleration readings, the sensor's frame (Figure 5.1) and the user's body frame need to be aligned. To correct the alignment, the participants are asked to adjust the sensors' position manually. Next, acceleration data is recorded while the user is holding a reference position. Therefore, the user is asked to keep its back straight while standing against a wall.

Here, the assumption is made that the deviation between the body's frame and the sensor's frame is high and depends on how the participant applied the sensor. In contrast, the deviations based on possible variations in the body posture are assumed to be small. Apparently, with this approach, no definite statement can be made for the *real* alignment. Nevertheless, taking a reference posture to distinguish the users' body frame can improve inter-individual accuracy. Hence, making it more reasonable to compare data among various participants.

To obtain the rotation matrix between the sensor and the user's body frame, up to 10 s of sensor readings are recorded during which the participant is asked to keep the reference posture. If the mean absolute deviation is small ($\leq 10^{-6}$ g), the sensor readings are mapped to a reference vector $v_r = [0, 0, -1]\,\mathrm{g}$. Therefore, the method proposed by [101] based on Rodrigues' rotation formula is used. An implementation is found in [178] (eq. 5.14).

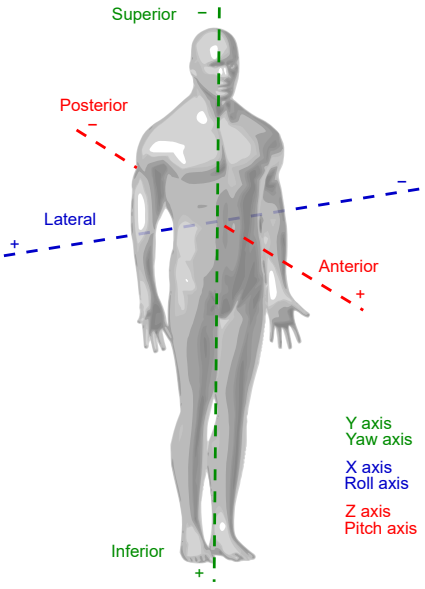


Figure 5.1: Orientation of the BG-V4.2[8].

$$R(a,b) = \tag{5.14}$$
$$I_3 + v_x + v_x^2 \cdot \frac{1-(a\cdot b)}{\|v^2\|}$$

$$v = a \times b$$
$$v_x = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}$$

R	Rotation matrix
I	Identity matrix
a, b	acceleration vectors
v_x	skew-symmetric cross-product

5.5 Experimental Estimation of Physical Activity in Firefighters

Commercially available fitness trackers are mainly designed for entertainment or wellness purpose. Yet, WBS and the estimation of EE are indeed interesting for occupational healthcare use. Occupational groups with high physical demands, such as police officers, military personnel, or firefighters, are a good example. For instance, typical tasks of firefighters include the recovery and rescue of people and animals and fire protection. These tasks require a certain level of physical and psychological capacity. This is because it is necessary to react flexibly, quickly, and situation-specific in operations. Only a rested, alert, and physically fit person can meet all these requirements.

For these reasons, the operating times and training units of firefighters must be well-structured [256]. However, the real individual demands (e.g. the number of actions and their intensity) vary from shift to shift. Here, WBSs can contribute by objectively recording the PA (e.g. as EE) at work. Based on this, e.g. shift schedules could be planned individually in order to prevent excessive workload.

In this chapter, the estimation of EE utilizing WBS for the select target group of firefighters is chosen. Recent scientific results regarding WBS use for EE estimation were already introduced (sec. 5.3). However, the presented results so far, are solely based on laboratory tests and focus on ADL.

Here, the focus is on the feasibility and possible pitfalls of using WBS in a professional context. Primarily, the difference of EE with and without wearing personal protective equipment (PPE) is explicitly addressed.

At first, related work is presented (sec. 5.5.1), before the experimental and computational methods are outlined (sec. 5.5.2). Subsequently, the results are presented (sec. 5.5.3) and validated against a publicly available data set (sec. 5.5.4). Finally, an excursus on condition monitoring for PPE is given (sec. 5.5.5), and the experimental results are discussed in summary (sec. 5.5.6).

5.5.1 Related Work

The use of commercially available WBS to collect objective EE information is broadly discussed in the literature [63, 76]. Considering the reported absolute results in terms of accuracy, it appears that they are often comparable. Interestingly, however, the conclusions differ widely. For instance, El-Amrawy et al. [79] reported the accuracy of step counting for various commercial fitness trackers to be in the range of 79.8 % to 99.1 %. This result is in agreement with what was reported in the meta-study provided by Feehan et al. [85]. However, El-Amrawy et al. [79] conclude that such devices will thus give sound output for EE estimation, while Feehan et al. [85] state the contrary.

[8] Based on: YassineMrabet, CC BY 3.0. Planes removed; annotations added.
Original: `https://de.wikipedia.org/wiki/Datei:Human_anatomy_planes.svg`

Likewise, other researchers summarize these findings. For instance, Nelson et al. [172] state that fitness trackers provide valid step counts, however, fail to estimate EE accurately. Likewise, Edwardson et al. [78] argue that WBSs offer "reliability and validity" if the specific limitations of WBS are appropriately considered when designing an experiment (here they refer to the use of the *activPAL* device (sec. 3.1.2.1) for sedentary behavior research). It can be said that the application's use-case determines the question of applicability.

By taking the experiences and knowledge from related work, it can be assumed that counting steps is oversimplified and misleading for the EE estimation of firefighters. At the same time, the validity requirements are less strict compared to diagnostic medicine or performance measurement in sport science. This is making the use of consumer-grade devices an option to consider.

An overview of the accuracy of fitness or activity trackers (sec. 3.1.1.1) can be found in the work of Chowdhury et al. [63]. They investigated the deviations of EE estimation comparing 4 consumer-grade, and 2 devices used in research. They reported MAPEs in the range of 27 % to 40 % for the consumer-grade devices. The best agreement was found with the *Actiheart* tracker (sec. 3.1.2.1), which showed a MAPE of 20 ± 15 %.

The applicability of WBSs for firefighter investigations in general and specific WBSs embedded into PPE have recently been presented [151, 219]. Also, algorithms to classify activities of firefighters (and emergency responders) have been presented [71, 134, 206]. Yet, solutions that provide information and support decision-making for operations management are still in its beginnings.

Moreover, the estimation of EE for firefighters, in particular, has rarely been addressed so far. Existing work often focuses on specific measures, e.g. heat stress [209]. Other methods combine readily accessible external information and use them as a proxy towards EE. More precisely, they make use of the HR, the air depletion rate from the respiratory protective equipment (RPE), and the time needed to finish an exercise [256]. In this respect, it was presented that it is possible to compare firefighters' performance in training situations objectively.

Furthermore, related work exists, which reports positive effects using WBS [110]. However, validation of the measurements is missing since no laboratory reference of EE was recorded. Other, more recent work again relies on subjective ratings of perceived exertion only [132]. Thus, with this work, it is intended to demonstrate the use of unobtrusive WBS to continuously keep track of firefighters' EE in order to assess individual physiological capacity. State-of-the-art methods to estimate EE (sec. 5.3) are used, and difficulties that need to be deliberated when the participants wear PPE are highlighted.

5.5.2 Methods

In the following, the experiment's design is outlined. This comprises the experimental protocol (sec. 5.5.2.1), the participants' characteristics (sec. 5.5.2.2), and the utilized

Table 5.2: Experimental protocol with approximate duration or distance covered in each station. Station 8 was part of the 1st trial only.

Station	Duration	Distance	Description
1	<10 min	-	resting (siting)
2	<5 min	-	resting (standing)
3		20 m	endless ladder
4		100 m	treadmill (6 km/h, 10 %)
5	1 min	-	bicycle ergometer (175 W))
6	<10 min	51 m	training gallery
7		150 m	treadmill (6 km/h, 10 %)
8*	<12 min	-	treadmill ramp (6 km/h to 12 km/h, 10 %)
9	<15 min	-	resting (siting)

hardware (sec. 5.5.2.3). After that, the pre-processing steps, the calculation of ground truth, and the extraction of the features are presented (sec. 5.5.2.4). Finally, methods used to analyze the data set (sec. 5.5.2.5) and the resulting models for EE estimation are summarized (sec. 5.5.2.6).

5.5.2.1 Experimental Protocol

The experiment's design closely follows the guidelines of the regular physical ability test for persons wearing heavy RPE in Germany (G26.3). The protocol includes climbing an endless leader, running on a treadmill, cycling on a bicycle ergometer, and completing a run through a training gallery (or maze), which includes climbing various obstacles on different levels or crawling through tubes (Table 5.2).

In addition, data were recorded while the participants were resting on a chair 10 min before the first and after the last exercise. In contrast to the real physical ability test, the test chamber was not darkened nor fogged. Neither distracting noises were used. This was done to avoid additional psychological strain.

To examine the effect of wearing PPE and RPE, each participant completed the experiment twice. During the 1st trial, the participants were wearing casual sports clothing, while during the 2nd trial, they were equipped with PPE and RPE.

The PPE consisted of a helmet, boots, pants, a coat, and gloves. The RPE consisted of the breathing apparatus only. The face mask was not worn due to the IC that was worn instead. The additional weight of PPE and RPE was approximately 25 kg for each participant. During all stations of the experiment, the RPE was worn on the back. While completing the training gallery, RPE was removed and had to be carried along. The exact execution (pushing, carrying, or pulling the RPE) was not specified.

Table 5.3: Mean and standard deviation of the participants' characteristics, the BORG-scale, and environmental data. Data for the 1st and 2nd trial is marked with 1 or 2, respectively. (BMI - Body Mass Index; Exp. - Experience with G26.3, Temp. - Temperature, Hum. - Humidity).

Parameter	Mean ± std	Unit	Parameter	Mean ± std	Unit
Age	37.4 ± 8.4	a	BORG,1	11.3 ± 1.4	–
Height	1.85 ± 0.06	m	BORG,2	13.6 ± 1.9	–
Mass	84.8 ± 12.3	kg	Hum.,1	55 ± 7	%
BMI	24.8 ± 3.0	kg/m^2	Hum.,2	53 ± 5	%
Exp.	16.2 ± 8.1	a	Temp.,1	17.8 ± 0.6	°C
Smokers	3	-	Temp.,2	17.6 ± 0.4	°C

Moreover, additional data was recorded (Table 5.3). That includes a BORG-scale[9] [38] rating to obtain the subjectively perceived exertion. Also, temperature and humidity were recorded at the beginning of each experiment. Comparing both trials of the experiment, neither for temperature ($p = 0.38$) nor for humidity ($p = 0.26$) a significant difference was found (paired t-test).

5.5.2.2 Participants

All participants had completed the regular physical ability test for firefighters wearing heavy RPE in Germany (G26.3) prior to the experiment. Thus, all participants were familiar with the different stations of the experimental setup, which closely followed the regular physical ability test.

In total, 13 male, experienced firefighters volunteered to participate in the experiment (Table 5.3). All participants were informed about the experiment's purpose and procedure, and the usage of their personal data. They gave their written consent to participate and agreed to scientific evaluation and publication of the anonymized data. Also, they were asked not to eat nor to drink anything except water 6 h before the experiment.

5.5.2.3 Hardware Setup

In order to obtain reference values (ground truth) of the participants' EE, a mobile IC, the *Cortex MetaMax 3B (R2)*[10] was used. It allows measuring oxygen uptake rate (VO_2 mL/min) and carbon dioxide production (VCO_2 mL/min) breath-by-breath (sec. 5.2.1.3). Calibration of the device was done before each run.

[9] The BORG-scale is a subjective, thus relative measure of exertion. Here, the original scale reaching from level 6 (no exertion) up to 20 (very, very hard) was used [38].

[10] Cortex Biophysik GmbH, `www.cortex-medical.de/`

Furthermore, RR (1/min) and respiratory minute volume (L/min) are recorded. Taking these measurements, EE is calculated in the unit kJ/min (sec. 5.2.1.3, eq. 5.1), PAR, and MET (sec. 5.2.1.3, eq. 5.8).

The BG-V4.2 was used during the experiment (sec. 3.1.3), to record the participants' HR and body acceleration. Additionally, body-near temperature and humidity were recorded. Therefore, an additional sensor (BME280, sec. 3.3.1.3) was placed next to the BG-V4.2. The ECG, acceleration, and environmental data were sampled with a frequency of 200 Hz, 100 Hz, and 1 Hz, respectively.

5.5.2.4 Pre-Processing and Feature Extraction

Target values (ground truth data) is obtained from IC measurements. To compensate for individual differences, the measured EE was normalized based on the participants' RMR (sec. 5.2.3). The resulting value is the PAR. To estimate BMR the participants RMR is divided by the factor 1.2. RMR, in turn, is calculated from the median EE while the participants were resting on a chair during the first phase of the experiment. Furthermore, MET was calculated, which was normalized by the participants' weight, assuming a constant oxygen consumption of 3.5 mL/(min kg) in resting conditions.

Regarding HR and RR, mean, minimum, maximum, range, variance, and slope of both raw and normalized data (HR, HR_n and RR, RR_n) was calculated. Normalization was done based on the 1 % or 99 % percentile for the lowest or highest value (eq. 5.15), respectively. To calculate slope, HR and RR were first low-pass filtered using a 10-second moving mean window. The slope was then calculated as the difference between the first and last values within a time frame. Artifacts and outliers were removed from the HR by automatically examining ECG quality (sec. 5.4.2.1).

$$N(X) = \frac{X - X_1}{X_{99} - X_1} \tag{5.15}$$

Multiple features were extracted from the accelerometer data on the X- (G_X, lateral) and Z-acceleration (G_Z, anterior-posterior) as well as on the summarized total (G_C, eq. 5.18), horizontal (G_H, eq. 5.17) and vertical acceleration (G_V, superior-inferior, eq. 5.16) and also on the pitch (G_P, eq. 5.19) and role (G_R, eq. 5.20) [182]. Features on all these signals in the time domain include: mean, median, minimum, maximum, variance, kurtosis, skewness, range, and root-mean-squared.

$$G_V = G_Y \tag{5.16}$$

$$G_H = \sqrt{G_X^2 + G_Z^2} \tag{5.17}$$

$$G_C = \sqrt{G_X^2 + G_Y^2 + G_Z^2} \tag{5.18}$$

$$G_P = \operatorname{atan2}\left(\frac{G_Z}{\operatorname{sgn}(G_Y) \cdot \sqrt{G_Y^2 + G_X^2}} \right) \tag{5.19}$$

$$G_R = \operatorname{atan2}\left(\frac{-G_X}{\sqrt{G_Z^2 + G_Y^2}} \right) \tag{5.20}$$

For all axis, except pitch and roll, additionally features based on the spectrum were calculated. These were: mean frequency, median frequency, energy in the frequency bands from 0 Hz to 5 Hz up to 45 Hz to 50 Hz in steps of 5 Hz. Also, all features were calculated on the 1st derivative of each signal.

Pre-processing of the accelerometer data was done as in previous work [275]. In brief, the following steps were applied: Accelerometer axis misalignment was compensation by de-rotation. Therefore, reference position data were recorded while the participants were asked to align themselves on a wall keeping their back straight (sec. 5.4.2.2). To reduce noise in the raw accelerometer data, a low-pass filter (Butterworth, 5. order, cut-off-frequency 20 Hz) was applied.

Regarding the environmental signals, mean humidity, mean temperature, as well as 2 indices derived from these values, were calculated. These are the "Heatindex" [221] (eq. 5.21, [33]) and the "Humidex" or humidity index [159]. Here a simplified version neglecting effects of radiation or wind is used (eq. 5.22).

These indices are intended to serve as an estimate of the skin-to-air temperature, e.g. the micro-climate[11]. They do account for the fact that heat loss (as a product of convection, conduction, and radiation) from the body is limited concerning relative humidity.

Here, it must be considered that the sensor readings are affected by the environmental temperature, and the heat production (loss) from the body. Moreover, the movement of the body eventually causes fluctuations in the sensor readings due to changes in convection. To remove those spurious readings, before further processing, the signal is low pass filtered (Butterworth, 5. order, cut-off-frequency 0.3 mHz). This is done under the assumption that environmental temperature changes are by orders of magnitude slower than the changes due to heat loss from the body (EE).

$$\begin{aligned} H_{\text{heatidx}} = \quad & -8.784695 + 1.61139411 \cdot t + 2.338549 \cdot r - 0.14611605 \cdot t \cdot r \\ & -1.2308094 \cdot 10^{-2} \cdot t^2 - 1.6424828 \cdot 10^{-2} \cdot r^2 \\ & +2.211732 \cdot 10^{-3} \cdot t^2 \cdot r + 7.2546 \cdot 10^{-4} \cdot t \cdot r^2 \\ & -3.582 \cdot 10^{-6} \cdot t^2 \cdot r^2 \end{aligned} \tag{5.21}$$

$$H_{\text{humidx}} = \quad t + 0.348 \cdot \left(\frac{r}{100} \cdot 6.105 \cdot \exp\left(17.27 \cdot \frac{t}{237.7 + t} \right) \right) - 4.25 \tag{5.22}$$

t - temperature / °C r - relative humidity / %

Additionally, the participants' VO_2max is taken into consideration. It is intended to serve as a coefficient of CRF. It is used in order to compensate individual differences regarding the HR to exercise intensity response. VO_2max is calculated as the maximal oxygen consumption observed during the treadmill ramp test.

[11] An overview of thermal discomfort induces is given by Havenith et al. [106] and Blazejczyk et al. [33]

5.5.2.5 Segmentation and feature selection

In order to select a suitable window size for the data segmentation, the MAE of different DT-based models is considered. The DT-based models are trained on feature sets calculated on sliding windows with different combinations of size and overlap.

All evaluated sliding windows are within the range of 4 s to 10 s (step size 2 s), 10 s to 60 s (step size 10 s), and 60 s to 120 s (step size 30 s). For each sliding window, also different overlaps are evaluated. These are in the range of 0 % to 90 % (step size 12.5 %). To prevent over-fitting and ensure comparability, the maximum number of splits of the DTs was limited to 10. During training, a LOGO CV was applied. The reported accuracy is the average across all splits.

In order to reduce the dimension of the feature vector and to ensure that only important features are selected, the feature space is evaluated in more detail. Therefore, features are ranked using *RReliefF* method described in [197]. The method has proven to be reliable in selecting relevant features, especially if interactions among the features are expected. It is based on the KNN idea, searching for pairs of target values, which are close to each other. If now a feature vector is also close, it is rewarded (ranked higher), while it is ranked lower if the feature differs in the distance.

Additionally, strongly correlated features are removed to avoid redundancies (Pearson's linear correlation is used). With this process, relevant features are identified. At the same time, less complex features are favored. To take one example, *range* and *variance* are correlated strongly. The calculation of range, however, is less complex, hence the range feature is chosen in favor.

All results were obtained using algorithms implemented with Curve Fitting Toolbox and Statistics and Machine Learning Toolbox provided by *MATLAB*[12].

5.5.2.6 Regression methods

To set up a regression model, ML algorithms from the *scikit-learn*[13] library were used. The methods chosen for evaluation are: Multiple linear regression (Elastic Net-, Ridge-, Lasso-Lars- and Bayesian-regression), k-nearest neighbor (KNN), decision tree (DT), artificial neural network (ANN), support vector machine (SVM) and ensemble methods based on DTs, namely Gradient Boosting and Bagging.

In addition to these methods provided by the *scikit-learn* library, MARS-models were considered. Therefore, the *py-earth*[14] package was used, which is a contribution to the *scikit-learn* library. The hyper-parameters of all models were optimized using Baysian-optimization based on the *scikit-optimize*[15] library.

[12] The MathWorks, Inc., Version 2018b, `https://www.mathworks.com/products/matlab.html`
[13] scikit-learn, Version 0.20.1, BSD Licence, `http://scikit-learn.org`
[14] py-earth, version 0.5.2, `https://github.com/scikit-optimize`
[15] scikit-optimize, version 0.1.0, `https://github.com/scikit-learn-contrib/py-earth`

Different metrics are considered to evaluate the predictive performance of the regression models that are used to estimate EE. These are the root mean squared error (RMSE), the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the coefficient of determination (R^2).

5.5.3 Results

At the beginning of this section, the overall experimental results are introduced (sec. 5.5.3.1). Subsequently, analytical results are presented. At first, the segmentation of the data (sec. 5.5.3.2) and the feature selection (sec. 5.5.3.3) is outlined. Next, the EE regression models are compared (sec. 5.5.3.4), and the best models are analyzed in detail (sec. 5.5.3.4). Finally, the model is re-validated to highlight common pitfalls setting up EE estimation models on a small sample (sec. 5.5.3.6).

5.5.3.1 Experimental

The entire data set includes 12.48 h of experimental data. Of these, 7.08 h were spent on the 1st trial and the remaining 5.40 h hours on the 2nd trial. The 1st trial was longer due to the additional ramp test performed by all participants. As neither the ramp test nor the run through the training gallery was timed, the completion times of the participants are different. On average, the participants completed the 1st and 2nd trial of the experiment in 0.54 ± 0.05 h or 0.42 ± 0.06 h, respectively.

Due to the malfunction of the environmental sensor, no temperature nor humidity data are available for 2 of the 26 runs. The malfunction resulted from physical damage to the sensor. For the later trials with PPE, thus the position of the sensor was changed, and it was attached to the jackets inside to avoid further data loss. However, this lead to lower temperature readings as the distance to the participants' skin was higher. For this reason, only scaled and normalized temperature readings can be used for comparison.

Evaluation of the EE at first clearly points out the differences between the rest or recovery phase (sedentary), and the active phases of the experiment. EE increases rapidly with the beginning of the first active phase (climbing ladders). Subsequently, between the individual phases, EE remains between a medium or high level and reaches its maximum at the end of the treadmill test (ramp). In the final recovery phase, the EE slowly decreases and approximately regains the level at rest.

The EE-curve shows apparent differences among the participants (Figure 5.2). It is found that the EE for one participant can be about twice as high as for another participant. Normalizing the absolute EE by converting it to PAR or MET (sec. 5.2.3), these distances decrease. The weight-related normalization of the OUR, i.e. as MET, reduces this distance already (coefficient of variation reduces from 0.159 to 0.118). If normalization is performed using the BMR, i.e. in terms of physical activity level (PAL), this further reduces the distance with respect to the resting phase (by definition).

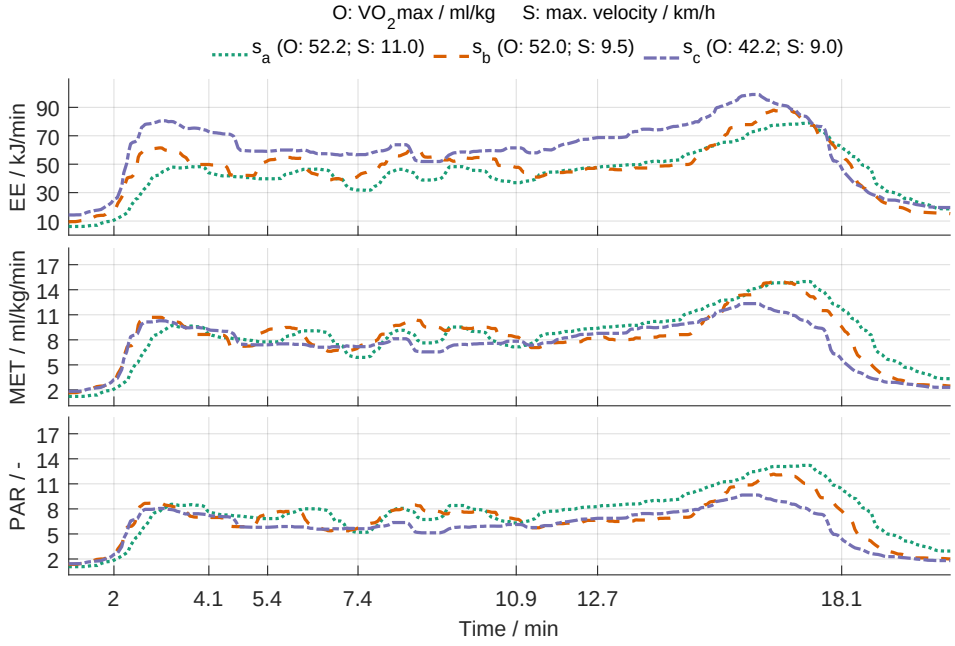


Figure 5.2: EE, PAR, and MET for 3 different participants during the experiment without wearing PPE. The time interval corresponds to the mean starting point of the different phases, which, however, are not exactly equal for all participants.

Considering the remaining course of the experiment, the variability increases (coefficient of variation 0.236). One possible reason for this is that the RMR was not determined precisely. In addition, the participants had different final speeds on the treadmill. Thus the individual demands were different. As a consequence, some participants experienced a RQ greater than 1. This implies that anaerobic energy supply occurred (sec. 5.2.1.3). As a consequence thereof, systematic errors are likely, if the calculation of EE is based on the basis of the RQ. At least this applies to those participants with anaerobe energy supply. For this reason, MET is used for the following analysis in favor of the PAR.

In the subsequent, the distribution of the MET values found during the different stations and trials of the experiment are inspected in more detail (Figure 5.3).

The highest average MET was found with the treadmill exercise. This is for both trials (1st trial: 8.24 ± 0.73 MET; 2nd trial: 10.55 ± 0.92 MET). With the beginning of the first exercise (ladder climbing), it can be noticed that MET is rising fast (due to the warm-up). This phase thus shows the highest variance of EE (± 2.64 MET).

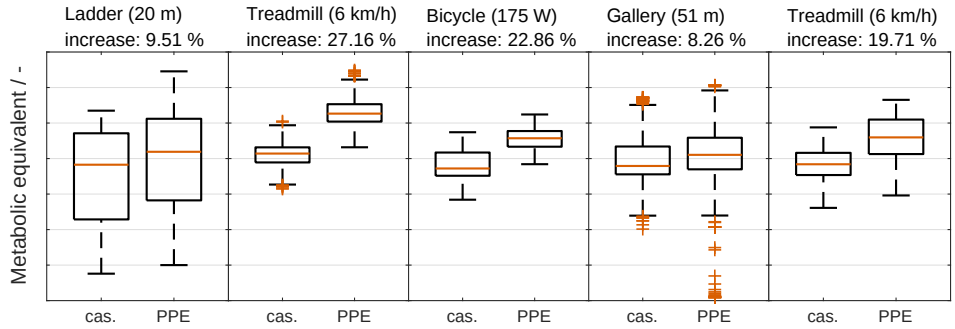


Figure 5.3: MET during the different stages of the experiment for both trials, with PPE and without (cas.). MET is noticeably increased when the participants are wearing PPE in all stages but the adaptation and training gallery.

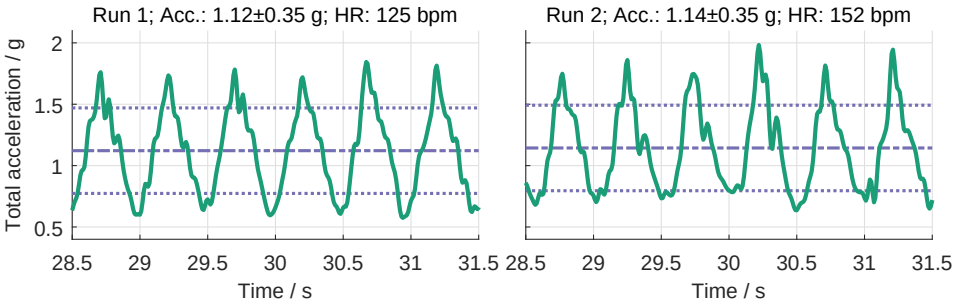


Figure 5.4: Comparison of the raw acceleration data and the median HR between both trials. The values depicted are from a single participant during the treadmill station.

In general, the average MET significantly increases between both trials of the experiment (Wilcoxon signed-rank test, $p < 0.001$). The average increase is $17.5 \pm 8.3\,\%$ across all stations except resting (stations 2 - 6). An increase in MET is found for all stations. The smallest increase is found in the training gallery station. The reason is that in contrast to all other stations, the run through the training gallery is self-paced, and time to finish is not limited. It is observed that, on average, the participants need $76 \pm 27\,\mathrm{s}$ (44 %) longer to finish the task with PPE equipped. Consequently, the EE in this exercise is higher when the participants are wearing PPE, although average momentarily MET is only slightly increased.

In accordance with MET, HR is found to be increased by $14.8 \pm 5.7\,\%$. Similar, RR increases by $23.8 \pm 6.3\,\%$. The mean energy in the raw acceleration data, however, is decreased by $4.8 \pm 7.3\,\%$.

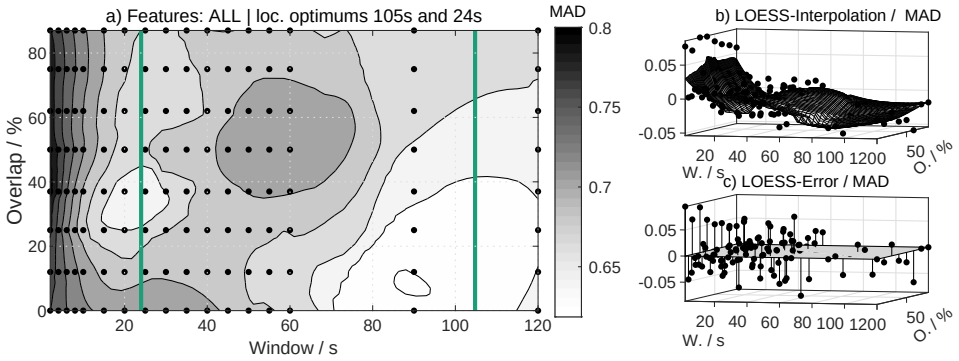


Figure 5.5: MAD found with DTs (maximum number of splits 10) on feature sets extracted on sliding windows with different sizes and overlap. Evaluated results were interpolated using locally weighted scatter-plot smoothing. Local optima are marked with a green line.

A detailed analysis showed that the acceleration patterns are only marginally affected when the participants wear PPE. No variation in raw acceleration is found for the bicycle station. Manually reviewing the raw data, only the merest variation in the gait patterns for some participants can be identified during treadmill run and ladder climbing (Figure 5.4). Since the participants carried RPE and had to put the device down and up again several times (e.g. in front of obstacles due to confined space), the movement patterns differ between both trials.

5.5.3.2 Segmentation

The pre-test carried out to identify a feasible segmentation of the raw data (based on DTs), revealed that different local optima exist. That is for features that were derived from the accelerometer (26 s), were related to HR (120 s) or the environmental sensors (104 s). These differences are again reflected in the results found with the entire feature set, which contains the features from all sensors (Figure 5.5). It is observed that differences among the optima cancel each other out. Thus, it can be concluded that no single best configuration exists.

Nevertheless, models with high accuracy can be found likewise for long and short windows. Yet, substantial distinctions exist for short and long windows. Regarding the longest segmentation, it is evident that only a few observations of EE remain. This results in a low variance in both the target and feature space. In the extreme, with a window size of 120 s, on average only 14.4 (11 to 19) observations remain. These observations are then further smoothed out if overlapping features are used. In consequence, resulting estimations are oversimplified and are no longer reflecting any variability of the EE.

It should be noted that choosing a small segmentation size leads to a high range of variability in the feature space, which must be justified appropriately as well. The rationale is that high-frequency variability in oxygen consumption reported by mobile IC most likely reflects measurement noise rather than real physiological interrelations [196].

These observations highlight, despite the fact that different local optima exist, that the choice of a suitable window size is not indifferent. As a compromise, for the subsequent analysis of the data set, a 20 s-window with an overlap of 50 % is chosen. With this choice, a new estimate is available every 10 s. This window size is also in agreement with considerations on the smoothing of EE estimates from IC and thus physiologically feasible.

5.5.3.3 Feature Selection

Before individual features are evaluated, the predictive performance of the feature (sub)-sets from all sensors is tested by spot-checking (Figure 5.6). For this purpose, again, DTs are trained on the feature sets derived for each sensor. Additionally, the activity is used as a predictor. That is the station the participant is currently completing.

All independent feature sets reveal to contain relevant information for EE estimation. Using LOGO CV, the best fit is found using the accelerometer features (RMSE = 1.36), followed by HR (RMSE = 1.66), and RR (RMSE = 1.87). The lowest agreement is found for the environmental features (RMSE = 3.17). Strikingly, these results are almost reversed when the 10-fold CV is used as the validation strategy instead of LOGO (Table 5.4). In that case, the best agreement is found with the environmental features (RMSE = 1.08), followed by HR (RMSE = 1.14) and acceleration (RMSE = 1.16).

At this point, the limited amount of participants in the data set has to be taken into account. As a consequence thereof, it cannot be excluded that with the 10-fold CV spurious relations are learned. This effect can clearly be seen when the station (or activity) is considered as a feature. In that case, the MET values can simply be mapped to their actual mean value found in the corresponding station of the experiment. In this way, a fixed mean value for the EE is learned for each station. This is only weakly affected by the choice of a validation strategy.

Moreover, assuming such a bias, it can be explained why HR falls behind acceleration data if the LOGO CV strategy is applied. With the LOGO validation, no individual HR response for the remaining participant under consideration can be learned from the data. For acceleration data, in contrast, there is no such inter-participant variability (The same applies to the activity or station features). In summary, it can be said that utilizing a LOGO validation pattern is a mandatory requirement to reflect the predictive performance correctly.

In this respect, the temperature or humidity and the specific mean EE of a participant have to be critically reviewed again. In-depth analysis reveals correlations of the independent trials of the experiment. That applies to the trials with and without PPE.

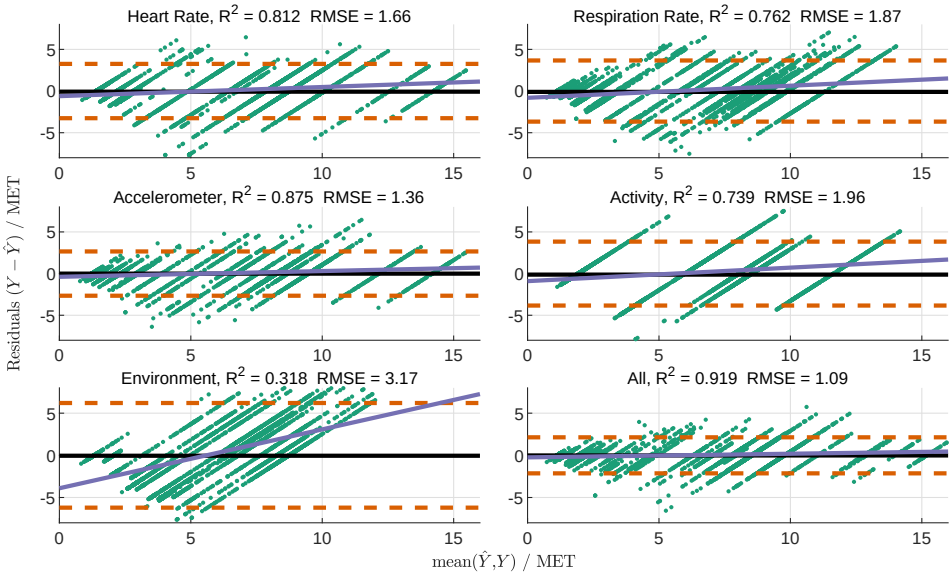


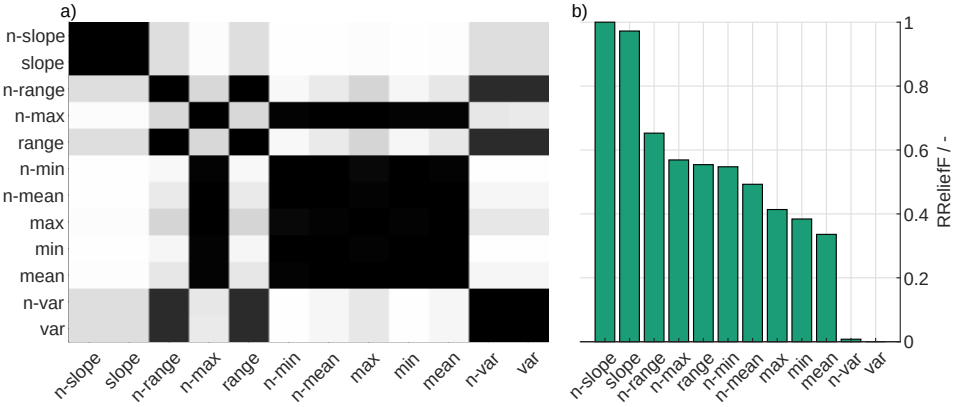
Figure 5.6: Bland-Altman diagrams for the prediction results found with the full feature-set and subsets containing all features from a single sensor or using the activity (or station) information as a standalone feature.

There is a lack of convection, while the participants are wearing PPE. Consequently, the patterns in the time series are different for these trials. Although humidity and temperature data are normalized, these spurious correlations with EE remain. The data is excluded from further analysis because of this reason.

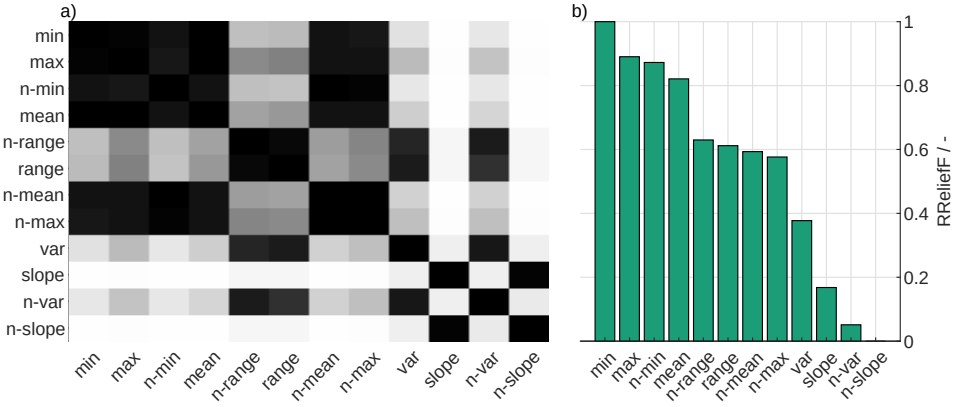
In the following, relevant features are selected from the different subsets (Figure 5.7). Regarding HR, the top-rated feature is the slope followed by the range. Minimum, maximum, and mean HR are equally weighted, where the distance of the mean value is slightly higher than the distance between minimum and maximum. The variance is never selected, thus does not seem to contain meaningful information. In general, normalized features are chosen in favor of non-normalized features. This was about to be expected, as the non-normalized features do not take into account inter-participant variations in HR (sec. 5.2.2.1).

Similar to HR, for RR, minimum, maximum, and mean values are identified as relevant features. Other features are found to be less relevant. Also, normalization is found to be less relevant. Indeed, that matches the fact that inter-participant variation regarding RR is lower compared to HR.

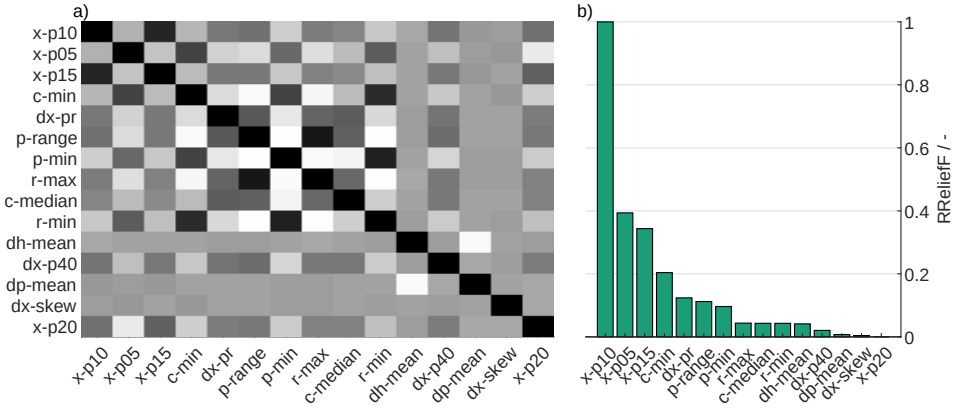
Considering the accelerometer data, the top-ranked features are found to be from the frequency spectrum. More specifically, band power in the 0.5 Hz to 1.0 Hz, 0.0 Hz



(a) Heart Rate



(b) Respiration Rate



(c) Accelerometer

Figure 5.7: Correlation and ranking (RReliefF) of different features sorted by the sensors.

Table 5.4: Comparison of the classification results of all sensors and their corresponding feature subsets. The RMSE and the R^2 are presented, which were found using a LOGO or 10-fold CV (kF) strategy.

Sensor or Feature	RMSE (MET)			R^2 (%)		
	kF	LOGO	diff.	kF	LOGO	diff.
Heart Rate	1.14	1.66	−0.52	90.9	81.2	9.7
Respiration Rate	1.62	1.87	−0.25	81.5	76.2	5.3
Environment	1.08	3.17	−2.09	91.9	31.8	60.1
Accelerometer	1.16	1.36	−0.20	90.6	87.5	3.1
Activity (Station)	1.90	1.96	−0.06	74.6	73.9	0.7
All	0.65	1.09	−0.44	91.7	91.9	−0.2

to 0.5 Hz and 1.0 Hz to 1.5 Hz range on the X-axis (lateral movement) are ranked highest. Regarding the time domain, minimum, and median (calculated on the total acceleration vector) are also selected frequently. This observation is in agreement with other reported results regarding HAR problems [199, 206].

This is interesting for two reasons. On the one hand, because it is known that HAR can be used to estimate EE indirectly. Here, this fact has already been pointed out, by the use of the station (i.e. activity) as a feature at the beginning of this section. On the other hand, the use of these features comes with the risk of over-fitting to the specific data set.

In the present case, only a limited number of different activities are present, and a clear separation between different activities is not always possible (e.g. in the training gallery). For this reason, and in order to reduce the risk of over-fitting, only a spare set of features is selected. In this respect, another group of features is noticeable, namely the range, and the minimum and maximum value of the pitch and roll. These values are selected because it is expected that these offer a more abstract view of the participants' movement.

For the final feature set, physiologically relevant features are selected. Regarding HR, this is the minimum value, as it is less affected by outliers. Still, the dynamics in HR are captured by additionally selecting the range feature. In addition, the slope feature is selected, which reflects the temporal dynamic of HR and EE (sec. 2.1.1). HR features are calculated on the normalized HR. In this way, inter-participant variations are taken into account correctly.

Regarding the RR, features are calculated in the same way as it is done for the HR. Concerning the accelerometer data, the position information is used, namely pitch and roll calculated from the accelerometer data. Hereby, the absolute position is represented as mean value, the motion dynamics by the range feature.

Table 5.5: Comparison of regression methods in terms of minimum RMSE found with Bayesian optimization given normalized (n) and non-normalized (-) input. For the linear regression (LR) method, the features were also polynomial transformed (squaring of individual and multiplying independent features).

Method	RMSE (n)	RMSE (-)	Method	RMSE (n)	RMSE (-)
ANN	1.41	1.61	SVM	1.77	1.98
MARS	1.53	1.59	LR	1.93	8.40
Boosting-DT	1.84	1.65	KNN	2.02	2.01
Bagging-DT	1.70	1.68	DT	2.26	2.50

To improve the SNR of the slope and motion features, they are grouped by means of discretization. In this respect, the slope features (calculated on HR and RR) and the range of pitch and roll are mapped to 5 distinct values. The body orientation (mean pitch and roll) is mapped to 12 distinct values. In addition to the features discussed, the VO_2max value is included as a measure of cardio-respiratory fitness (sec. 5.5.2.4).

5.5.3.4 Model Selection

Initially, spot-testing is carried out to select appropriate ML methods, given the previously defined feature set (sec. 5.5.3.3). In this process, also the effect of further expanding the feature space by use of polynomial features and scaling the feature space by standardization and normalization is inspected.

In total, 510 different models are evaluated. Mean RMSE is 3.54 among all tested configurations. The overall variation is considerably high (RMSE $= \pm 2.11$). In most cases, this can directly be derived from the characteristics of the corresponding methods. E.g. linear methods take advantage of polynomial features, whereas non-linear methods intrinsically map inter-dependencies among different features. Also, methods that are non-invariant to scaling, e.g. SVM or ANN benefit more if standardization is applied (e.g. in contrast to DT).

The top-performing models found during spot-testing were then selected, and their hyper-parameters were fine-tuned using Bayesian optimization (Table 5.5).

5.5.3.5 Model Analysis

The 2 most accurate models were found using ANN and MARS. The ANN-based model uses a logistic activation function and is based on a single layer with 4 neurons (and 11 inputs). The MARS-based model consists of a total of 36 linear basis functions. Hence, both can be considered light-weight in terms of memory and computational demands.

Both models are comparable in terms of accuracy. That is, regardless of the metric under consideration. The ANN, however, is slightly better in terms of R^2 and RMSE

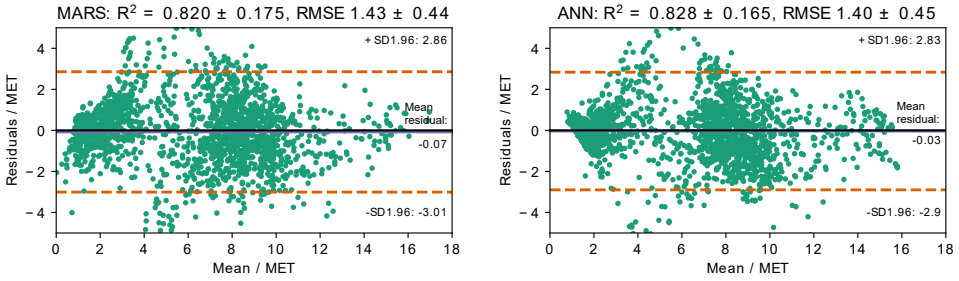


Figure 5.8: Bland-Altman diagram showing residuals and mean of the true MET (y) and estimate MET values ($\hat{y}$). Estimates were generated using the optimized MARS-, and ANN-based model (Bayesian optimization). Residuals are distributed comparably. Both models show low bias and variance.

(ANN: $R^2 = 0.828$, RMSE $= 1.40$; MARS: $R^2 = 0.820$, RMSE $= 1.43$). The biggest difference is found regarding the MAE (ANN: MAE $= 0.65$; MARS: MAE $= 0.78$).

With respect to EE estimation, it could be argued that the latter is more important. That is because a low precision (e.g. short-term deviations) is less critical, considering EE on a daily scale. In the best case, it is averaged out entirely. In contrast, the lower accuracy in terms of MAE indicates a systematic error (bias), which eventually affects the mean of the estimated EE.

A more detailed view of the error distribution using Bland-Altman representation (Figure 5.8) emphasized that the mean residual in both models is also comparable. No systematical error can be found. That is, the models neither systematically under- nor over-estimate MET. What is striking, however, are the clusters around 1 MET to 2 MET (sedentary) and 8 MET (high-intensity or sub-maximal effort). These clusters reflect the distribution of MET in the data set.

It becomes clear that the data set as a whole has only limited variability concerning different intensities. This is a consequence of the experiment's design. In particular, light-intensity tasks are underrepresented. As a result, a specific over-adaptation to the surrounding boundary areas becomes visible. Accordingly, the maximum deviation is found at about 4 MET to 6 MET.

The sole consideration of the error metrics and error distribution does not allow to draw a clear distinction of both models. The most persuasive argument to select the ANN-based model instead of the MARS-based model is the slightly improved error so far. That, however, could also be an effect, which is limited to the particular data set. This is because it only represents a limited sample of a much more diverse population.

This very problem is likewise stated in related work as well (sec. 5.5.1). As a consequence, evaluating the error metrics is only helpful to a limited extent. Although visualizations of the error distribution (e.g. Bland-Altman) help to characterize the errors, they do not provide detailed insights into the models' response. For this reason,

in the subsequent, the model's response for certain inputs will be analyzed as an evidence-based check in order to clarify the relationships found.

The final model uses HR (minimum, range, and slope), and RR (minimum, range, and slope) in combination with body orientation (mean and range of pitch and roll), and CRF. As a first result, it is observed that with both methods (MARS and ANN), similar relationships have been learned.

Strikingly, a virtually linear relationship between HR and MET, is represented (Figure 5.9, upper left), which matches prior expectations given physiological knowledge. Regarding the MARS-based model, an almost linear representation is found, while the sigmoid shape (s-curved) becomes clearly visible in the ANN-based model's response. Furthermore, interaction with the slope feature is found (which is not present in the MARS-based model). In this sense, the output is positively offset if lower HR values are present that also show a sharp positive slope. Likewise, a negative slope does the same in the case of mid-level HR value.

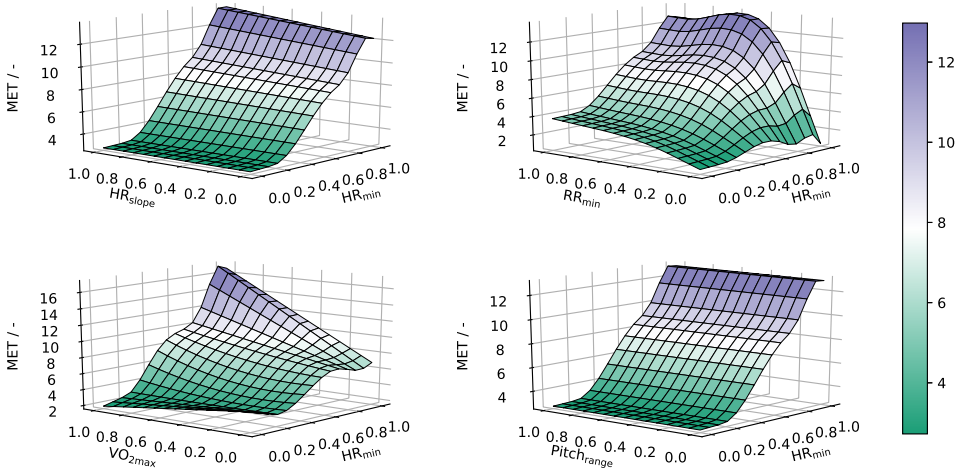
Considering the CRF feature (Figure 5.9, lower left), another offset regarding the HR mapping is found. With maximal HR, CRF acts as a limiting or moderating factor. In this regard, high MET will only be estimated if CRF is high as well. The mapping found with MARS oscillates, while the ANN (again) shows a smoothed, sigmoid shape. Thus, considering the ANN-based model, MET is rated higher across the full-scale range of HR (no such interaction is found with RR).

Next, the ANN's response given motion dynamic (range of pitch; Figure 5.9, lower right) is inspected. The pitch and range features add to the MET estimate in non-static conditions. That is for above low-level HR values. In the opposite case, if no motion occurs but HR rises, the response curve stays within a lower level.

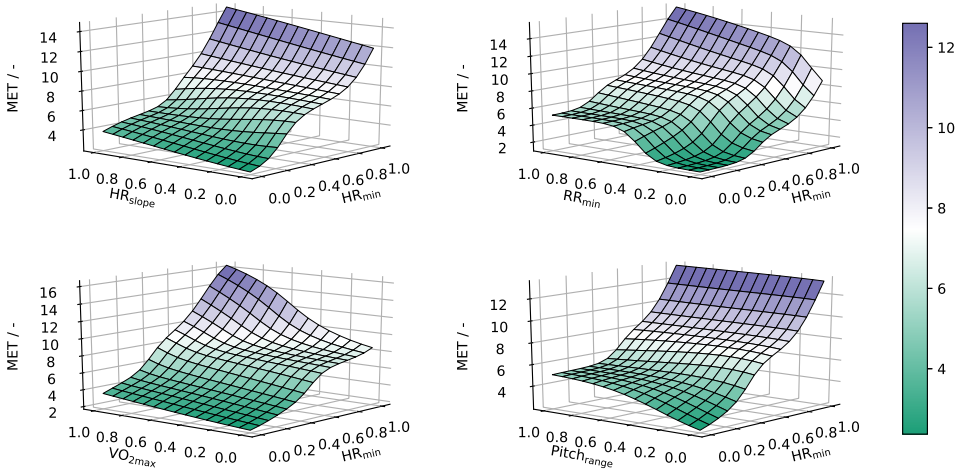
That, however, is for HR in the lower-levels only. Otherwise, HR dominates the estimation. A similar effect is found with RR (Figure 5.9, upper right). It moderates the model's response, which leads to a lower estimate of MET, especially if HR and RR diverge. Just like the effect of motion, it offsets the estimate, given lower HR values. If both HR and RR rise evenly, the response equals the case described with the interaction of CRF and HR.

Considering now the MARS-based model, no interaction between motion and HR can be found. Furthermore, the moderating effect of RR (fixing MET at a low level), is more extreme. Compared to the ANN-based model, the MARS-based model's response is either more linear (no interaction) or more oscillating (e.g. CRF versus HR).

This observation can be explained, considering the smoothing nature of the sigmoid activation-function used by the ANN-based model. That is, the model's response equals a *simple* linear combination of all 4 sigmoid responses in the hidden layer (all weights can be found in sec. C). Not only this smooths the response, but it also limits the output to fixed boundaries of 1.24 MET to 17.8 MET (including the intercepts). Both findings are good reasons to choose the ANN-based model instead of the MARS-based model, although the error (R^2 and RMSE), as well as the residual distribution found with Bland-Altman analysis (Figure 5.8), are comparable.



(a) Response of the MARS-model.



(b) Response of the ANN-model.

Figure 5.9: Multivariate analysis of the EE models' response of MARS (a) and ANN (b) for HR against 1. HR_{slope}, 2. RR, 3. CRF, and 4. $Pitch_{mean}$. Remaining features were fixed at their corresponding mean value for each analysis ($HR_{min} = 0.46$, $HR_{range} = 0.07$, $HR_{slope} = 0.50$, $RR_{min} = 0.39$, $RR_{range} = 0.10$, $RR_{slope} = 0.50$, $Pitch_{mean} = 0.19$, $Pitch_{range} = 0.12$, $Roll_{mean} = 0.59$, $Roll_{range} = 0.50$, $CRF = 0.62$).

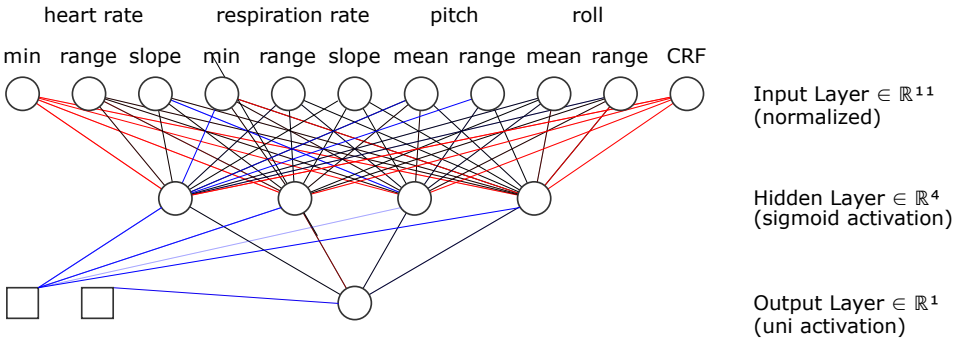


Figure 5.10: Graphical representation of the ANN-based model's weights from LOGO validation. Each neuron is represented as a circle. The intercept vectors are represented as squares. Each line represents a weight, where red or blue lines indicate positive or negative values, respectively. The lines' transparency was mapped linearly to the corresponding normalized weights.

A closer look at the ANN-based model's weights (Figure 5.10) confirms the observations made from visual inspection of the model's response. It can furthermore be seen that HR, RR, pitch, and CRF dominate the model. In this respect, the pitch is the only factor contributing to a negative offset towards the estimates. In direct comparison, the slope and range information only contribute little to the estimate. In particular, it can be seen that the RR's slope is almost weighted zero. By considering this, the model could be further shrunken by omitting the RR's slope.

All models were constructed to provide a MET estimate every 10 s. Depending on the application, not only the short-term deviations but also the overall mean deviation is of interest. With Bland-Altman analysis, the mean of all deviations is −0.03 MET, which corresponds to an error of 0.5 % given a mean of 5.59 MET in the data set.

This result, however, can be misleading because it does not attribute the variability among the different participants and trials of the experiment. Therefore, the mean MET estimated for each participant (also separated by trial) is inspected. The MAE ranges from 0.50 MET to 2.35 MET or 0.62 MET to 2.58 MET for the ANN-based, or MARS-based model, respectively. The corresponding mean MAE is 0.99 ± 0.45 MET, or 1.08 ± 0.45 MET, respectively. The mean MAPE for the ANN-based model and the MARS-based model is 11.00 ± 9.54 % (range 0.00 % to 46.00 %) and 11.86 ± 9.87 % (range 0.10 % to 49.00 %), respectively.

5.5.3.6 Trial Comparison

The firefighter user group has special requirements for PA estimation. Wearing PPE and carrying the additional equipment causes extra weight load. In addition, the micro-climate under the PPE is different compared to casual clothing. This means that

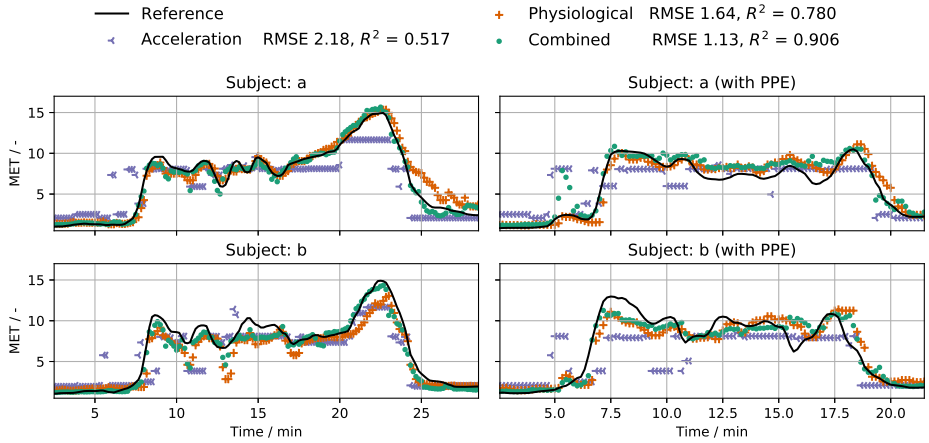


Figure 5.11: Exemplary response plot of the ANN-based model estimating MET for 2 participants. The response of the uni-modal (HR or acceleration) and the multi-modal model are compared. Both models were trained on the data of the 1st trial without PPE only.

the very same activity differs in its intensity. This was already shown, by reflecting the EE increase comparing both experimental trials (sec. 5.5.3.1).

As a closing test for the ANN-based EE estimation model, the 2 trials are compared to each other (Figure 5.11). The model is trained on the full feature set and 2 different sub-spaces of the feature space containing physiological (HR, RR, and VO_2max only) or acceleration data, only.

At first, it is noticed that the error in the combined model is lower, compared to previous results. This is explained by the missing LOGO CV because here, the 2 trials are validated against each other. Thus, data from each participant is available in every set, and inter-individual differences can, therefore, better be balanced.

Comparing the results of the accelerometer- and physiology-based model, strikingly, there is a difference of 28.3 %. It can also be seen that the accelerometer-based model lacks certain dynamics. Its estimates show a more stepped pattern, compared to the physiological model. Also, these steps are identical for both trials (with and without wearing PPE). Furthermore, in resting conditions, at the end of the experiment, the accelerometer-based model underestimates EE. In contrast, HR reflects these dynamics better, however, it tends to overestimate EE during resting conditions. These well-known drawbacks (sec. 5.3) are successfully avoided by combining the features.

Regarding the previously reported results, it must be noted that the hyperparameters were optimized for the selected feature-space and validation scheme (LOGO CV). Hence, the comparison of these solutions is not objective. In order to further support the findings, the models are re-trained and compared again. However, now, a 2-fold CV (between the experiment's trials) is used.

Table 5.6: Comparison of regression results using physiological or acceleration features only. Models are optimized using Bayesian optimization given normalized inputs using a 2-fold CV based on the 1st and 2nd trial of the experiment.

Method	Physiological		Acceleration		Difference / %	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
ANN	1.57	0.828	2.15	0.670	31.3	21.1
MARS	1.81	0.772	2.29	0.636	23.4	19.3
Boosting (DT)	1.90	0.748	2.26	0.644	17.3	14.9
Bagging (DT)	1.96	0.734	2.31	0.629	16.4	15.4

The top-4 algorithms, namely ANN, MARS, Boosting, and Bagging (both with DT as base learners), are selected. The models are trained separately using the physiological features (minimum, maximum, range, and slope of HR and RR, and CRF), and the acceleration features only. To furthermore increase comparability, the top-11 features from the acceleration sensor are used (sec. 5.5.3.3, Figure 5.7c). Hence, both feature spaces are comparable in terms of dimension.

In summary, the RMSE for the physiological and accelerometer-based model errors found are RMSE = 1.57 to 1.96 against RMSE = 2.15 to 2.31 and R^2 = 0.828 to 0.734 against R^2 = 0.670 to 0.629 (Table 5.6). In all tested configurations, using the physiological features leads to better estimates for both RMSE and R^2. However, it can be seen that the re-optimization improved the R^2 of the accelerometer-based model compared to the results reported earlier (Figure 5.11). Still, the physiological-based model outperforms the accelerometer-based model by at least 14.9 % (based on R^2).

It can be concluded that the intensity-information from the physiological data becomes more critical if external factors are influencing the activity's intensity. The comparison has shown that in these cases, it is advantageous to weight the physiological markers higher than body acceleration. Still, a combined model is mandatory to avoid overestimation during sedentary behavior.

5.5.4 Model Benchmark and Comparison

Models to estimate EE, are based on experimental data, the acquisition of which is associated with a high experimental cost. In current studies, the number of test persons is therefore low. Typically, it is in the order of 10 to 30 participants. The number of participants in this study is also low (N=13). In order to check whether the findings made in this work are reliable, the previously analyzed ANN-based model (sec. 5.5.3.5) is re-evaluated on a publicly available data set provided by Gjoreski et al. [95].

5.5.4.1 Benchmark Data Set

The data set contains a total of 16.55 h of data from 10 participants. In contrast to the here presented experiment, boundary conditions (e.g. different clothing, weight load) play no role in this data set. The focus is on method selection. The EE is given as MET values, recorded by means of a Cosmed K4b2 IC (sec. 5.2.1.3).

Secondary sensors were used to recorded HR, RR, skin-, and body temperature (sec. 5.3). Also, body acceleration was recorded on the participants' chest and thigh. From the acceleration data, peak counts are calculated. For all modalities, pre-processed data is available in windows of 10 s. The raw sensor data itself, however, is not available.

Because the raw data is not available, the ANN-based model is re-trained on the given feature sub-set for comparison reasons. However, only HR, RR, acceleration peak count, and activity features are used. Furthermore, both physiological features are smoothed using a 2nd order Savitzky-Golay filter using 13 data points. This is done to remove outliers. In addition, a slope feature is calculated for RR and HR, which is based on the difference of two successive feature values. These differences are mapped onto 3 discrete values, indicating a negative, positive, or no slope. Similar, the acceleration peak count is mapped to 5 discrete values.

5.5.4.2 Target value pre-processing

For the following comparison, 3 different target variables are used. First, the unmodified target (y MET) as originally given in the data set. Secondly, two smoothed variants of the target variable are used (y_b and y_s). In order to obtain these, the target variable is filtered using a 5th order Butterworth filter (y_b) with a cut-off frequency of 0.04 Hz. Additionally, a 2-nd order Savitzky-Golay filter (y_s) using 13 data points is used for comparison. Both are applied because of the assumption that the high-frequency changes in the physiological EE signal are a result of measurement noise. This pre-processing is done in accordance with the suggestions given by Robergs et al. [196].

The rationale to expect inevitable noise in the measurement data from IC is based on previously reported observation of the so-called *BxB-noise* (from breath-by-breath) [99, p. 290]. The *BxB-noise* is explained as a result of the error propagation in respiratory flow analysis. In this respect, the error is due to the inaccuracy of the volume sensor and occurs in all open-circuit ICs.

Following the recommendations of Robergs et al. [196], the assumption is made that high-frequency variability in the IC signal can solely be described as measurement noise. This is a strong assumption, which, however, is physiologically evident. Following the physiological principles of EE, its variability must be a reflection of the changes in the blood flow through the muscles or the oxygen uptake of the cells (sec. 2.1.2). This relationship can also be directly deduced from Fick's principle[16]. In turn, if a

[16] With Fick's principle the cardiac output (c_o) is modeled as a function of oxygen consumption (vO_2) and oxygen difference between arterial and venous blood (a_{diff}). Accordingly, oxygen consumption can be

spontaneous increase of the EE is observed, an equivalent response of heart rate should be observed as well.

Considering now the original data set, the mean percentage change (of two successive values) of the HR is $0.6 \pm 1.1\,\%$. The corresponding percentage change of EE (y), however, is $3.9 \pm 0.3\,\%$. Filtering the original data, and thus removing high-frequency fluctuations, this variation is reduced to $0.7 \pm 1.3\,\%$ (y_b) or $0.3 \pm 0.5\,\%$ (y_s). It can be seen that the smoothed target variables match the variability found with HR.

However, the noise signals ($y - y_f$ and $y - y_s$) are found to be non-Gaussian distributed, which contradicts the assumption of measurement noise. Instead, the noise distributions are leptokurtic (Kurtosis$\left(y - y_f\right) = 7.74$, Kurtosis$\left(y - y_s\right) = 8.81$), and their distribution follows a more Laplacian or Cauchy-like shape. In detail revision of the difference between the smoothed and original target shows a mean deviation of 0.00 MET and a standard deviation of 0.45 MET or 0.35 MET for the Savitzky-Golay or Butterworth filter.

5.5.4.3 Evaluation

The originally reported accuracy for the data set presented by Gjoreski et al. [95] is RMSE $= 0.850$ or MAE $= 0.613$. It was also used by Catal et al. [55], who reported RMSE $= 0.757$ or MAE $= 0.709$. Using, the 4-neuron ANN-based model, it results in RMSE $= 0.942 \pm 0.289$ or MAE $= 0.719 \pm 0.250$. These results are obtained using the data set without body or skin temperature. All reported results are mean values (and standard deviation) from the individual results of the LOGO validation.

Strikingly, when the ANN-based model is compared to Catal et al., the deviation in terms of the RMSE (27.9 %) is high, while it is as low as 1.4 % considering the MAE. However, considering the results presented by Gjoreski et al., the deviation is high for both error metrics (RMSE: 13.9 %, MAE: 17.3 %). These results show that the model is not capable of reflecting all variability within the data set. That is, considering the original target variable (y), only.

Comparing the error metrics of the ANN-based model given the smoothed (y_s) or original signal, these differences change. The model's accuracy is improved by 13.45 % or 13.04 %, considering the RMSE or the MAE, respectively. Comparing these new results to Gjoreski et al. [95] and Catal et al. [55], the deviations in terms of the RMSE are $-1.66\,\%$ or $9.9\,\%$, respectively. Making the same comparison, but considering the MAE, the deviations are $1.9\,\%$ or $-12.6\,\%$, respectively. Thus, depending on which metric is considered, the ANN-based model performs better or worse than the comparison values. In any case, it can be concluded that by using the smoothed target values, the ANN-based model provides comparable results.

Whether the interpretation of the EE fluctuations as measurement noise is feasible, cannot be conclusively answered at this point. In order to clarify this issue, the

given as $VO_2 = c_o \cdot a_v O_2 diff$. According to this assumption, differences in oxygen consumption directly follow differences in either c_o or a_{diff}.

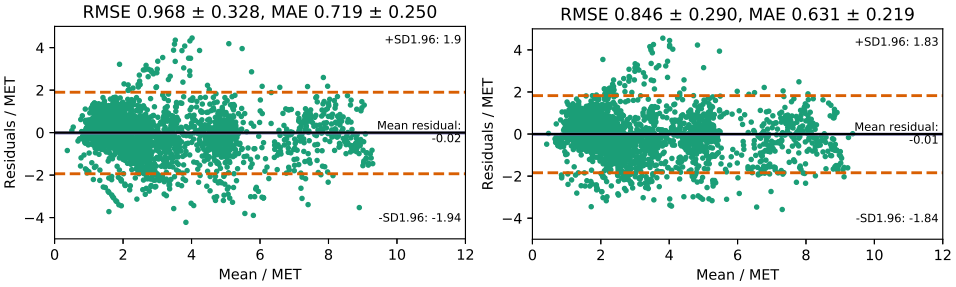


Figure 5.12: Bland-Altman Diagram of the ANN-based model's response trained on the data set provided by Gjoreski et al. [95]. On the left side, the original target is used, while on the right side, it was smoothed using a Savitzky-Golay filter.

smoothed target values are superimposed with a synthetic noise signal and tested again (Figure 5.12). Therefore, 2 different noise signals are used.

The noise signals are sampled from a normal distribution with zero mean and a standard deviation in accordance with the standard deviation found in the corresponding difference signals (n_f: 0.35, n_s: 0.45). The model's accuracy is then re-evaluated with respect to these synthetic-noise targets. The original signal and the synthetic signal are compared by calculating the difference between the respective accuracy metrics[17].

Re-evaluating the model's accuracy, it is found that the results are similar, regardless of using the original target or the synthetic variant, which is superimposed with white noise. Considering the Butterworth filtered target, the difference is limited to the second or third decimal place (RMSE: 0.027, MAE $= 0.004$). Moreover, with the Savitzky-Golay filtered target, there is no difference at all (RMSE: 0.000). Most interestingly, even a negative deviation is found, i.e. the accuracy has improved (MAE: -0.022). From these results, it can be learned that the model is invariant to any high-frequency components in the target values.

At this point, the constricted complexity of the proposed model becomes obvious. Its entropic capacity is relatively low, as only 4 neurons with 6 different features are used[18]. As a consequence, the RMSE using the original unfiltered data is high.

By smoothing the target data in a physiologically feasible way and thus removing high-frequency fluctuations, it can be seen that the ANN-based model provides reasonable estimations. In this sense, specific variations of the EE, which are assumed to be not physiologically evident, are removed from the model's response. Indeed, this can be seen as an advantage of the model since an over-fitting towards any noise is virtually impossible (of course, out-of-test errors are still possible due to over-fitting to the participant-dependent characteristics present in the sample under consideration.).

[17] That is, given the metric M (RMSE or MAE): $\mathrm{M}(\hat{y}, y_o) - \mathrm{M}(\hat{y}, y_f + n_f)$, or $\mathrm{M}(\hat{y}, y_o) - \mathrm{M}(\hat{y}, y_s + n_s)$

[18] In sec. 5.5.3.5 the same model with the same hyper-parameters is used, but with 11 features.

Besides, these findings pinpoint the need for an in-depth analysis of short-term fluctuations in EE. In this understanding, it is interesting to further elaborate short-term fluctuations of EE and identify other correlations, e.g. with the accelerometer signals. Such an analysis requires access to all components of the IC, including the air-flow meter, O_2- and CO_2-sensors. However, this is outside the scope of this work.

In summary, the proposed model can be considered a lightweight, low-variance model for EE estimation. Moreover, the remaining bias affects short-term fluctuations only. In fact, the mean bias across all predictions is reduced to 0.02 MET (Figure 5.12). This implies that if the mean PA, across more extended time periods, is of interest, those missing dynamics do not affect the final estimation results.

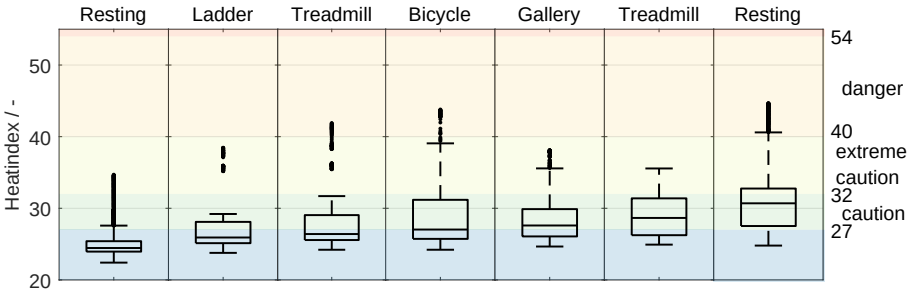
5.5.5 Environmental Condition Monitoring

Up to this point, the experimental data (sec. 5.5.2.3) was used to estimate PA only (sec. 5.5.3). Moreover, the sensor data can also directly be used to provide feedback for other means of improving occupational health and safety as well. Regarding firefighter activities, heat stress and heat-related injuries are a major concern. Detecting such heat problems early is thus critical in order to prevent impending dehydration or heat strokes. Gaining insights on these issues is also interesting because it can help to maintain optimal physical and mental performance ability in action. Therefore, the environmental sensor data is of interest.

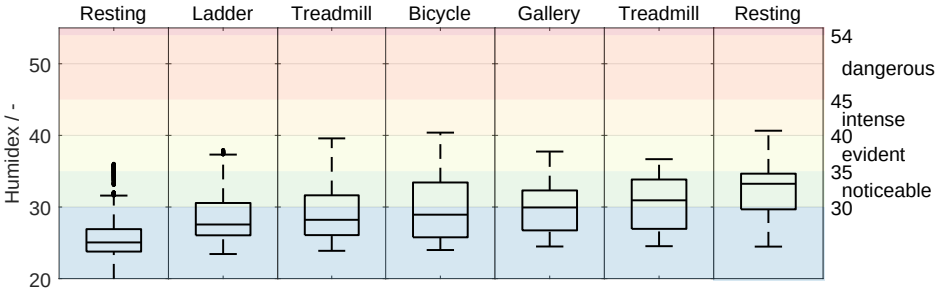
In fact, heat is a major problem for firefighters. In this regard, direct external heat exposure is less of a problem because firefighters wear insulated clothing (PPE). This acts as an insulating layer and provides proper protection against the external heat. Because of the very same insulating property, however, the physiologically generated excess heat (EE as a result of PA), cannot be effectively dissipated any longer. For this reason, convection and conduction, in particular, are limited, which affect heat loss and evaporation of sweat. That is for any activity that involves wearing PPE and, therefore, also applies to operation at non-fire scenes. As a result, the micro-climate under the PPE clothing changes. This effect was also observed in this experiment and will be explained in more detail below as an excursus.

Both humidity and temperature were monitored during the experiment (sec. 5.5.2.3). The micro-climate under clothing (PPE) is a function of heat and sweat production or loss. Due to the insulating PPE, temperature and humidity loss are limited. Thus, they are rising slowly in time (Figure 5.13). The increase in humidity is explained with the effect of the accumulation of moisture due to sweating.

The mean environmental temperature and humidity, measured before the experiment started (2nd trail only), was $17.6 \pm 0.4\,^{\circ}\mathrm{C}$ and $53 \pm 5\,\%$. Mean temperature and humidity under the PPE was 27.3 ± 1.5 and $74 \pm 15\,\%$. Maximal measurements were $30.1\,^{\circ}\mathrm{C}$ and $100\,\%$.

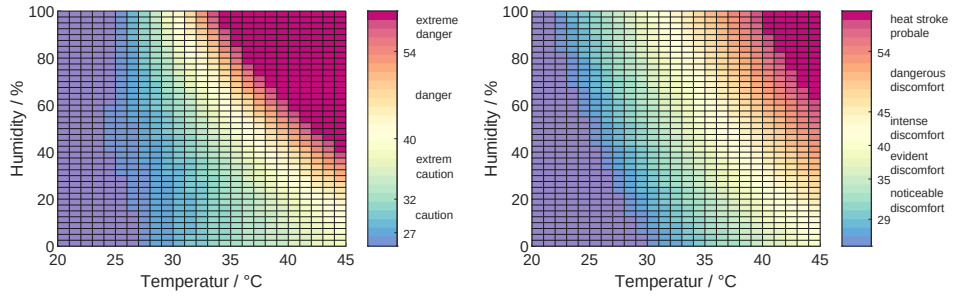


(a) Heatindex [221]



(b) Humidex [159]

Figure 5.13: Boxplots for the Heatindex (a) and Humidex (b) measured during the different stages of the 2nd trial of the experiment.



(a) Heatindex [221] **(b)** Humidex [159]

Figure 5.14: Heatindex (a) and Humidex (b) are combined measures of temperature and humidity to reflect heat stress or thermal discomfort.

Noticeably, an excessive accumulation of both heat and moisture is observed during the experiment. Both can be depicted in combination using the Heatindex or Humidex representation (sec. 5.5.2.4, Figure 5.14)[19].

Initially, these indices were designed to provide a better temperature estimation for weather forecasting. In this sense, they were used to estimate an apparent air temperature. The rationale behind this idea is that the same physical temperature *feels* different in dry or humid conditions. Nevertheless, these and other indices are used in epidemiologic and public health research. [106]

Overall, Heatindex is barley found to be in the *dangerous* region and never reaches *extreme danger* levels. What is striking, however, is that immediately with the onset of PA (ladder exercise), the uncritical area is crossed. The mean Heatindex remains at the level of *caution* after the bicycle station. This zone is never left. Not even at the end of the exercise, while the participants are resting. At this point, the index is instead rising again, after it had temporarily started to decrease lightly (training gallery).

The same trend is found looking at the Humidex index. Again, the uncritical area is firstly crossed within the first exercise. Also, in the later stages, the mean Humidex stays in the "noticeable" area. The *intense* range is reached only sporadically and briefly. The measurements never touch *dangerous* or *heat stroke* levels. In direct comparison, the limits of the Humidex are higher than those of the Heatindex. For this reason, areas marked as *cautiousness*, *intense* or *dangerous* are passed later or never at all.

In the work of He et al. [108], temperature and humidity under PPE were recorded for 6 firefighters similar to how it was done in this work. They furthermore tested the influence of altered external conditions, e.g. hot and dry, hot and humid but also cold settings. Regarding heat and humidity accumulation, the curve of both indices observed is in close agreement with the findings of [108].

The correlation of micro-climate changes and PA was already shown (sec. 5.5.3.3). At first glance, body-near temperature and humidity appear to be appropriate predictors for PA. However, spurious relation cannot be neglected. The effects of different clothing and their permeability to sweat and heat are one aspect.

Besides temperature and humidity, environmental conditions need to be considered likewise. That is e.g. heat radiation from sunlight or wind. All these conditions are not present in the laboratory test. It is therefore inferred that micro-climate measurements do not generalize across different participants, types of clothing, or external climatic conditions. To be applicable in practice, these influences need to be removed beforehand, e.g. by using an appropriate normalization factor provided by an additional external sensor.

Although it was found that the micro-climate (environmental) data is of limited use for PA estimation, its consideration as a separate index is apparent. For instance, He et al. [108] reported that the participants' overall sensation (that is the perceived strain

[19] The Heatindex representation is only valid for temperatures over 26 °C. Also, a simplified version of Humidex was used, which is calculated on basis of the humidity instead of the dew point temperature (sec. 5.5.2.4, eq. 5.22)

during an exercise) strongly depends on the micro-climate. Moreover, early detection of critical temperature ranges is a useful metric for command and control in training and the field.

5.5.6 Discussion

With this study, a model to estimate PA using MET representation is presented. The participants in the study are firefighters. They represent a select interest group for whom PA estimation as a professional application is of practical relevance, e.g. to promote occupational safety. The experimental setting highlights differences in metabolic EE caused by wearing everyday clothing or PPE during the same exercise or activity.

Concerning the experimental setup, it should be noted that no normalization of clothing nor environmental factors was done. This means that the participants chose their own clothes for the experiment and also used their own PPE. One example is that some test persons wore a balaclava (woolen cap) under their helmets, while others did not.

Moreover, temperature and humidity in the test room were not exactly the same for all tests. Thus, inter-participant differences are high. This weakens the results from the comparison of the 2 trials, although it was found that the PA was significantly different. A reliable conclusion on the additional strain caused by wearing the PPE or the additional weight load cannot be drawn due to the non-normalized conditions. This, however, does not affect the assessment and estimation of the PA entirely. The interference of spurious correlations was counteracted by removing individual predictors, such as temperature and humidity information, from the estimation model.

With regard to the conduct of the experiment, it should also be noted that no randomization of the activities was carried out. In addition, the equipment exercises, i.e. cycling and climbing ladders, are very short. Overall, they make up about only 7 % of the entire data set. For both reasons, no steady-state of EE was reached. Thus, no conclusions can be made on the real EE for these activities. This is critical because it is well-known that cycling is a challenging activity to estimate since only little body movement can be measured [169]. If cycling would account for a larger part of the data set, it could easily be that the estimation of the resting phases would be less accurate.

Nevertheless, the data set recorded has substantial distinctions due to the range of various types of movement. That is walking, crawling, crouching, and climbing. In this respect, it was noticed that, given the full feature-set, acceleration features on the X-Axis (lateral, left, and right) were ranked the highest among all other accelerometer features. This observation partly contradicts previous findings, in which all axes were found to add information to the model [275]. In this data set, however, it can be seen that the Z- (anterior-posterior, back and forth) and Y-Axis (vertical, up and down), are continuously rotated against each other. This happens every time the participants are switching from walking to crawling and vice versa.

In other work, for similar reasons, only total acceleration (eq. 5.18) is used [146, 169]. In this work, body orientation (pitch and roll), was used. Both mean and variance of pitch and roll contribute in deciding if the participant is sitting, lying, or moving. It could be assumed that this information is also useful to distinguish various other activities. Consequently, in future work, the use of orientation corrected acceleration measures should be investigated. This could help to find features that generalize better because it contains more information than the flattened total acceleration signal.

The experiment also revealed a systematic error if the PA estimation is based on acceleration data alone. Comparing both trials of the experiment, it was found that EE differed by $17.5 \pm 8.3\,\%$. Likewise, HR and RR followed this trend ($14.8 \pm 5.7\,\%$ or $23.8 \pm 6.3\,\%$, respectively). In other words, it is found that the physiological markers reflect the increase in metabolic EE. In contrast, the acceleration signals are not different from each other. As a consequence, the percentage difference in RMSE comparing models based on accelerometer data or physiological information alone is 16.4 % to 31.3 %. This highlights missing information on the intensity given accelerometer data.

In summary, the results underline the necessity of physiological markers to reflect the dynamics or intensities of PA. Still, prediction relies on having motion information to distinguish between sedentary and active activities, as it was previously suggested by Altini et al. [14].

Furthermore, it is noticed that normalization for inter-individual differences is crucial. That is normalization of physiological markers such as HR, which is partly determined by genetic factors or individual CRF. Consequently, CRF should be added to every PA model because otherwise, no predictions of the maximum metabolic EE can be made.

Moreover, slope features turned out to improve the estimations. In the model's response, it could be seen that the slope acts as an offset to the estimation with respect to HR. That is, if the HR rises (positive slope), the EE is also estimated higher. If HR decreases, EE estimation is slightly lower. That is in agreement with the temporal characteristics of metabolic EE, which shows certain inertia (sec. 2.1.3). Also, the model's response shows a certain plateau with decreasing of the HR. This is likely a reflection of the EPOC effect taking place during the cool-down phase at the end of the experiment.

Having a closer look at the feature space, it can be seen that due to the normalization layer, the use or adaption of input features becomes more flexible. This can be illustrated by the examples of HR and RR. If one of the two values is missing, they can simply be replaced (equated) by each other. This eliminates the compensating effect of having both inputs, but the model itself remains functional. Similarly, this applies to pitch, roll, and slope features. If these are fixed at their mean value (0.5), the prediction becomes less accurate, but no extreme outliers are to be expected.

Another vital aspect of the presented model arises from the nature of the utilized activation function. The ANN-based model uses sigmoid activation functions in the hidden layer (output layer is not activated, i.e. identity function is used). Because the sigmoid function is bounded, its use guarantees a fixed limit for the model's response.

For this reason, (even though the learned sample is small and thus prone to over-fitting) the model's estimates can never drift towards an extreme. Given the weights learned for the data set used in this work, the estimation will always be in the range of 1.24 MET to 17.8 MET.

The model presented in this chapter shows close agreement ($R^2 = 0.82$, RMSE = 1.4; sec. 5.5.3.5) to the reference gold standard (IC). Detailed analysis of the residuals, however, also revealed a certain variance. Therefore, it can be inferred that the model is not able to fully reveal all dynamics of PA.

Nevertheless, the error across all participants, in terms of mean deviation, is small −0.03 MET. Given the mean PA during the entire experiment (5.59 MET), this corresponds to an error of 0.5 % across all participants. At this point, however, it must be noted that this result applies to this very data set only (i.e. the sample under consideration), and does not necessarily hold for another sample population. Also, with this particular error metric (mean error), negative and positive errors cancel out each other.

Although the model's response is close to a perfect agreement, with respect to the sample under consideration, MAPE across all participants is 11.0 ± 9.5 %. This comparison of the different participants is a better estimate for its out-of-test error than the Bland-Altman analysis. It must be acknowledged that this error compares in magnitude to that previously reported in other work (20 ± 15 % [63]).

Here, it must furthermore be noted that it is not advisable to use the model in another context. The overall sample is too small and comes from a limited set of participants. Most importantly, no woman participated in the experiment. The model is thus expected to have a bias error if it would be re-applied onto a new sample without re-training the weights. Nevertheless, it could be shown that the ANN-based model (and its hyper-parameters) is a qualified candidate for PA estimation in general. This is because the model provides accurate estimations for the mean PA in the long term (averaged across all participants).

Re-Using a data set that was made available to the public by [95], the model was re-trained. It is found that the presented model does not cover all the dynamics in the ground truth data. Its variability is limited. Consequently, an inevitable bias error is observed. Nevertheless, on average, the estimated METs closely match the gold standards ground truth from IC.

Here, it remains unclear if the variation in the ground truth data comes from real metabolic changes or instead represent measurement uncertainties. It can be argued that from a physiological point of view, the dynamics are exceptional. This is in agreement with considerations for the use of IC data [196]. There is no formal proof that variability recorded with IC is not reflecting real physiological differences in EE. However, there is strong evidence that the variability is mainly measurement uncertainty. In order to trace this assumption of noisy ground truth, it was smoothed using digital filters. This way, certain dynamics are removed, and the new deviation found is close to that reported by Gjoreski et al. [95] (RMSE: −0.004, MAE: 0.018) or Catal et al. [55] (RMSE: 0.089, MAE: −0.078).

Moreover, that result is achieved, even without using the full feature set provided but using HR, RR, and acceleration-peak count only. Again, as a matter of the limited sample size, the model cannot be used in a generalized context. Yet, it becomes clear that the model's entropic capacity is a good fit because it is matching the task's complexity. This is also consistent with recent findings presented by Lu et al. [146] (sec. 5.3).

5.6 Conclusion and Future Work

In this chapter, the challenges of mobile and unobtrusive PA estimation using WBSs were outlined. On the basis of data obtained from a self-conducted experiment with 13 professional firefighters, feasibility and limitations of the use of WBS for PA estimation in a professional context were highlighted. It is concluded that WBSs are well suited for the prospected use-case, which is to estimate PA. With the extensive study of the data obtained, common pitfalls in the use of WBSs were identified. This includes uncovering spurious relations, e.g. the use of environmental data such as temperature or humidity, as well as the use of motion features that are prone to out-of-test errors.

As a result, a lightweight ANN is presented that uses heart rate (minimum, slope and range), respiratory rate (minimum, slope and range), body orientation (mean and range of the pitch and roll), and individual cardiorespiratory fitness in its input. The 11-dimensional input is forwarded to 4 neurons with a sigmoid activation function. Its architecture is similar to that presented in [146] and thus supports their finding on the suitability of light-weight estimation models (concerning the entropic capacity). Other, far more complex models under consideration in this chapter did not contribute to improving the estimates any further (i.e. larger network architectures or ensemble methods like boosting). Contrary to this, there are indications that these represent a distorted *view* or mapping of the input features and the EE.

The model presented in this chapter is physiologically grounded, and thus reflects commonly known relationships between exercise intensity, physiological responses, and individual cardiorespiratory fitness. This is achieved by using prior knowledge, e.g. an evident normalization of the physiological input parameters was applied. Apart from that, the physiological relations are learned automatically by the ANN. The agreement is in the order of magnitude compared to the state-of-the-art. This is an important finding since the data set poses a challenge because it contains a wide variety of activities and intensities.

However, it is noticed that the model does not reflect all fluctuations. In other words, its variability is limited, and for this reason, it provides smoothed MET estimates only. Nevertheless, by taking the knowledge of human energy metabolism into account (sec. 2.1.3), this can also be seen as an advantage of the model. That is because it is known that high-frequency fluctuations of PA are not reflected by the HR or the RR (sec. 5.5.4). As a consequence thereof, this implies that models with a low entropic capacity (or variability) already match the entropy (or information content) of these

physiological features. Moreover, they are not prone to over-fitting. In contrast, other more complex models could (in the worst-case), be prone to over-fitting to certain imperfections of the measurement system (e.g. *BxB-noise*, sec. 5.5.4).

It can be concluded that, given the predictors HR, RR, and acceleration, the presented model and similar light-weight approaches [12, 146, 169] belong to a *near to an optimal* family of solution. Following the principle of parsimony, these models should also be preferred, as they are less prone to over-fitting.

Yet, the problem of PA estimation cannot be considered fully solved. From sports science and physiology, other attributes and effects are known that could be used to estimate PA more precisely.

Firstly, that concerns the use of other sensors than those used in this work. For instance, the use of sensors embedded into a wearer's shoes could be used to capture ground-reaction forces [170]. These could complement whole-body acceleration (as it was used in this work) to provide load-weight (intensity) information. Furthermore, heat-flux sensors [149] are an interesting branch to measure *pseudo*-direct calorimetry values, which are more promising than environmental measurements used in this work (micro-climate). Also, biochemical sensors could be interesting to gain insights into non-aerobic components of EE. Such sensors, furthermore, could be used to derive information about sweat production or dehydration [26].

Secondly, changes in EE involve more temporal effects than can be coped with utilizing simple slope features, auto-regressive methods, or lag-features. In this work, CRF was used as a *moderating* factor. The average level of PA, however, changes the dynamics of EE as well [186]. From their insights, also, more general limitations concerning PA can be drawn. For example, it is known that in the long term, EE cannot rise above a limit of approximately 2.5 PAR on average [236].

In summary, despite the limitations discussed in this work, the estimation of PA by means of WBS can be considered a valuable tool for research and other professional contexts. Moreover, their use is not limited to these fields but concerns everyday life as well. Indeed, the positive effects of utilizing WBSs were recently reported [44]. In this respect, Brickwood et al. [44] state that the use of fitness-trackers is a motivating factor, which helps to increase physical activity and likewise decrease sedentary behavior. This might be considered the most import contribution of WBSs (apart from being easy-to-use, yet accuracy tools for PA estimation).

6 Summary and Outlook

This thesis has addressed and answered several questions on the design and application of wireless body sensors (WBSs): What are WBSs, and how are they distinguished from medical equipment? What are the possible applications and fields of use? Which methods are suitable and which measures need to be placed in order to reliably predict abstract concepts such as cognitive workload or physical activity.

In chapter 3, the requirements, and the design aspects of a WBS, the *BI-Vital* (BG-V5), were outlined. Its level of maturity matches that of a pilot line product. It was validated against its predecessor, and a small batch was manufactured, which now is readily available for upcoming lectures or students' work.

Due to the scalability of the *BI-Vital*, it is possible to cover different use-cases. This starts from serving as a heart rate monitor utilizing the inter-operable Bluetooth Low Energy protocol, continues with acting as a high-resolution wireless data-logger to transmit an electrocardiogram (ECG) recording and motion data, and reaches all the way to the use as a prototyping platform for advanced embedded algorithms.

As an exemplary use-case for embedded inference, the on-line classification of cardiac arrhythmia through a convolutional neural network was presented and analyzed energetically. As a result, it is highlighted that an accurate and continuous on-line inference is achievable for up to 24 h before the device needs to be recharged.

Additional application examples can be found in supervised work. For instance, embedded inference solutions for running-speeds were compared [285], which is an import topic in view of the goal of implementing real-world applications. Also, the use of the BG-V5's inertial measurement unit was evaluated to support a vision-based online indoor-tracking system [290]. This example is a first step towards the use of the *BI-Vital* within a larger system infrastructure.

Moreover, this thesis provides new experimental results to the scientific community. In this regard, two select real-world problems taken from the scope of occupational health and safety were presented:

Within the scope of chapter 4, the concept of cognitive workload was examined, its operationalization discussed, and measures to estimate cognitive workload (CW) were presented. To obtain these results, an experiment was conducted to induce CW and measure psycho-physiological responses, which were then used to predict CW.

Succeeding a statistical inspection, multiple state-of-the-art machine learning methods were used and compared to each other in order to identify a valid representation of CW. As a result, models were identified that allow estimating CW in a fine-grained manner (5 classes) with high temporal resolution (15 s) and an accuracy of up to 72.6 %.

It was, however, noticed that these methods miss inter-participant generalization. This highlights the need to clarify this issue in future research.

Considering the results, the question arises to what extent psycho-physiological responses are suitable to act as a proxy towards the activity of the autonomic nervous system. More precisely, it is interesting to identify normalization factors (e.g. cardio-respiratory fitness) and other physiological markers (e.g. salivary cortisol) that could be measured by WBSs and help to improve the inter-participant agreement or consistency.

Nevertheless, the applicability of WBS for the unobtrusive acquisition of physiological measures was demonstrated. These can serve as additional *real-time* markers (recorded simultaneously to any test) in order to augment the interpretation of experimental data in psychology. This data is usually not accessible with other methods, e.g. by retrospectively executed questionnaires.

In contrast, chapter 5 focuses on the estimation of physical activity (PA) (in terms of energy expenditure), which is a topic that has already become widespread. In this thesis, it was questioned in which respect WBS can be of practical use in a professional context. Thereof, an experiment to estimate PA in firefighters was conducted using state-of-the-art methods, e.g. whole-body accelerometry, heart rate monitoring, and respiration rate monitoring compared to indirect calorimetry reference measurements. Thereby, otherwise easily overseen pitfalls were highlighted. These, for instance, arise from different weight loads due to the use of respiratory protective equipment, which results in different PA for the same activity. Because these weight loads are invisible to the acceleration sensor, they can only be overcome by using combined measures.

Having an accurate model to estimate PA, as it was presented in this thesis, the investigation of long-term effects can be considered for future work. In this way, temporal effects and methods to re- or auto-calibrate estimation models in order to match individual characteristics can be addressed.

The model presented in this thesis already provides a simple solution for re-calibration via the fitness feature, but for future work, more sophisticated (or automated) solutions are preferable. Besides, the validation of additional sensory information is to be addressed. This includes but is not limited to the use of heat-flux sensors or sweat sensors.

The results obtained here moreover support to clarify the so-far inconsistently answered question on the complexity of machine learning models for estimating energy expenditure. It was found that light-weight models readily provide accurate and physiologically evident estimates. This was uncovered through an extensive analysis of the final model and an additional validation, which is based on publicly available data.

In summary, the model presented in this thesis is robust due to the combined use of multiple measures, is light-weight and thus suitable for the use by resource-constrained WBSs, and physiologically evident, which was shown through the detailed analysis of the model.

By summarizing both application examples, the potential future use of WBSs for practical implementation was demonstrated. Moreover, trough the summary of the history of WBS and the outline of the development of the BI-Vital, it is learned that WBSs (and wearable devices in general) have changed in perception regarding non-engineering scientific disciplines. For instance, the frequent use of step counters or heart rate monitors in public health and sport science were discussed (sec. 5.1).

Moreover, the application of WBS reaches even further beyond the boundaries of scientific work. Fitness trackers and heart rate monitors have been used as a hobby and in sports for many decades. In fact, the first battery-operated wireless and ECG-based device for heart rate calculation via finger measurement was introduced as early as 1977 (sec. 2.2.1). What is remarkable now is that confidence in technologies and algorithms has increased, especially among the non-scientific public.

Smartphones or smartwatches, which are used as step-counters or tools to assess sleep-quality, are examples for this development [18, 19, 179]. Even beyond this, first digital health-related devices find their way onto the market. An example is the Apple Watch ECG-app[1], which allows detecting cardiac arrhythmia (sec. 2.2.1 and sec. 3.5). It is not the first device of its kind, but it has a vast area of influence, with an estimated 13 million units sold in 2016 [50]. Considering these signs of progress, it becomes evident that WBSs are leaving the field of specialized applications in research and are becoming more widespread instead.

The continuing proliferation of WBSs is a development to be welcomed because it makes the specific advantages of body sensors available to the public. The application examples discussed in this thesis demonstrated some of these. Nevertheless, the limitations have also become apparent, and this work concludes with the hint that the application of body sensors is still too strongly based on closed-world assumptions. Therefore, the specific results of many applications presented in the literature are often hard to reproduce outside the lab. The numerous validation studies, also on commercially available devices widely used, confirm this.

Moreover, ethical questions also arise. On the one hand, this concerns privacy, but on the other hand, it also concerns the representation of the data to the wearer, which is sometimes overly simplified. In this respect, also, the lack of professional support in interpreting the data is a concern. [166]

Research and development around body sensors, therefore, require more considerable attention to aspects of validation outside the lab. This, however, does primarily affect the application of consumer products. Besides, open research questions can also be found within the scientific communities itself. Regarding machine learning, topics like online learning or learning from sparse data are of particular interest concerning the data obtained by WBSs [177]. With respect to material science, the research on new sensors, which are sensitive to physiological metabolic markers (like lactate in the sweat), is up to come [26, 161]. Furthermore, miniaturization of electronics

[1] Apple Inc., `https://www.apple.com/healthcare/apple-watch/`

and the design of flexible hardware are promising advancements regarding the future development of WBS [29].

Research is already underway on all of these topics, but work is still in its beginnings. In this respect, this thesis concludes with the hope that the knowledge accumulated here (and elsewhere) about the design and application of wireless body sensors can kindle their future development.

List of Abbreviations

ADC	analog-digital converter
ADL	activities of daily living
ADP	adenosine diphosphate
AF	atrial fibrillation
AMP	adenosine monophosphate
ANN	artificial neural network
ANS	autonomic nervous system
ANT	ANT™is a proprietary wireless protocol marketed by Garmin Canada
ATP	adenosine triphosphate
BG-V4.2	BI-Vital, Bielefeld-Vitalmonitor, version 4.2
BG-V5	BI-Vital, Bielefeld-Vitalmonitor, version 5.0
BLE	Bluetooth Low Energy
BMR	basal metabolic rate
CNN	convolutional neural network
CNS	central nervous system
CPS	cyber-physical system
CPU	central processing unit
CRC	cyclic redundancy check
CRF	cardiorespiratory fitness
CV	cross-validation
CW	cognitive workload
DAC	digital-analog converter
DC	direct calorimetry
DLW	doubly labeled water
DPS	degree per second (1 DPS $= \pi/180$ rad/s)
DSP	digital signal processing
DT	decision tree
ECG	electrocardiogram
EDA	electrodermal activity
EE	energy expenditure
EEG	electroencephalography
EMG	electromyography
EPOC	excess post-exercise oxygen consumption
ES	embedded system

FFT	fast Fourier transform
FN	false negative
FNR	false-negative rate
FP	false positive
FPR	false-positive rate
FPU	floating-point unit
GP	Gaussian process
HAL	hardware abstraction library
HAR	human activity recognition
HR	heart rate
HRM	heart rate monitor
HRV	heart rate variability
IC	indirect calorimetry
IEC	International Electrotechnical Commission
IMU	inertial measurement unit
InAmp	instrumentation amplifier
IoT	internet of things
ISM	industrial, scientific, and medical radio band
ITU	International Telecommunication Union
KNN	k-nearest neighbor
LDO	low-dropout
LED	light-emitting diode
Li-Pol	lithium polymer battery
LOGO	leave-one-group-out
LR	linear regression
LSB	least significant bit
MAD	median absolute deviation
MAE	mean absolute error
MAPE	mean absolute percentage error
MARS	multivariate adaptive regression splines
MCU	microcontroller unit
MEMS	micro-electromechanical system
MET	metabolic equivalent
ML	machine learning
MSE	mean squared error

NAND	flash memory based on not-and gate architecture
NB	naive Bayes
NOR	flash memory based on not-or gate architecture
OpAmp	operational amplifier
OUR	oxygen uptake rate
PA	physical activity
PAEE	physical activity-related energy expenditure
PAL	physical activity level
PAR	physical activity ratio
PCB	printed circuit board
PCr	phosphorylated creatine
PDM	pulse-density modulation
PLA	polylactic acid
PNS	parasympathetic nervous system
PPE	personal protective equipment
PPG	photoplethysmography
PPV	positive-predictive value
PVA	polyvinyl alcohol
PWM	pulse-width modulation
QRS	characteristic part of the ECG waveform
QSPI	quad serial peripheral interface
R^2	coefficient of determination
RAM	random-access memory
RMR	resting metabolic rate
RMSE	root mean squared error
ROM	read-only memory
RPE	respiratory protective equipment
RQ	respiratory quotient
RR	respiratory rate
RTC	real-time clock
SB	ShockBurst™is a proprietary wireless protocol marketed by Nordic Semiconductors
SCL	skin conductance level
SCR	skin conductance response
SDK	software development kit

SIMD	single instruction multiple data
SLIP	serial line internet protocol
SNR	signal-to-noise ratio
SNS	sympathetic nervous system
SoC	system-on-chip
SPI	serial peripheral interface
SPL	sound pressure level
SRAM	static random-access memory
SVM	support vector machine
TEE	total energy expenditure
TEF	thermic effect of food
TN	true negative
TNR	true-negative rate
TP	true positive
TPR	true-positive rate
UART	universal asynchronous receiver transmitter
USB	universal serial bus
UWB	ultra-wideband
VO_2max	aerobic capacity
WBAN	wireless body area network
WBS	wireless body sensor
WBSN	wireless body sensor network
WPAN	wireless personal area network
WSN	wireless sensor node

List of Figures

List of Tables

Bibliography

[1] J. Achten and A. E. Jeukendrup. “Heart rate monitoring: applications and limitations”. In: *Sports Medicine* 33.7 (2003), pp. 517–538. DOI: 10.2165/00007256-200333070-00004.

[2] *AD8235; Datasheet*. Analog Devices, Inc., 2016. URL: https://www.analog.com/media/en/technical-documentation/data-sheets/ad8235.pdf.

[3] *ADA4505-1/ADA4505-2/ADA4505-4; Datasheet*. Analog Devices, Inc., 2017. URL: https://www.analog.com/media/en/technical-documentation/data-sheets/ada4505-1_4505-2_4505-4.pdf.

[4] B. E. Ainsworth, W. L. Haskell, S. D. Herrmann, N. Meckes, D. R. Basserr, C. Tudor-Locke, J. L. Greer, J. Vezina, M. C. Whitt-Glover, and A. S. Leon. “2011 Compendium of Physical Activities: a second update of codes and MET values”. In: *Medicine & Science in Sports & Exercise* 43.8 (Aug. 2011), pp. 1575–1581. DOI: 10.1249/MSS.0b013e31821ece12.

[5] B. Ajana. “Digital health and the biopolitics of the Quantified Self”. In: *Digital health* 3 (2017), p. 2055207616689509. DOI: 10.1177/2055207616689509.

[6] M. H. Al-Haboubi. “Energy expenditure during moderate work at various climates”. In: *International Journal of Industrial Ergonomics* 17.5 (1996), pp. 379–388. DOI: 10.1016/0169-8141(95)00002-X.

[7] M. M. Al Rahhal, Y. Bazi, M. Al Zuair, E. Othman, and B. BenJdira. “Convolutional Neural Networks for Electrocardiogram Classification”. In: *Journal of Medical and Biological Engineering* 38.6 (2018), pp. 1014–1025. DOI: 10.1007/s40846-018-0389-7.

[8] A. Alberdi, A. Aztiria, and A. Basarab. “Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review: A review”. In: *Journal of biomedical informatics* 59 (2016), pp. 49–75. DOI: 10.1016/j.jbi.2015.11.007.

[9] F.-P. Alberto, M. Nathanael, B. Mathew, and B. E. Ainsworth. “Wearable monitors criterion validity for energy expenditure in sedentary and light activities”. In: *Journal of sport and health science* 6.1 (2017), pp. 103–110. DOI: 10.1016/j.jshs.2016.10.005.

[10] P. Alinia, R. Saeedi, B. Mortazavi, A. Rokni, and H. Ghasemzadeh. “Impact of sensor misplacement on estimating metabolic equivalent of task with wearables”. In: *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2015, pp. 1–6. DOI: 10.1109/BSN.2015.7299385.

[11] E. Alpaydın. *Introduction to machine learning*. Adaptive computation and machine learning. MIT Press, 2004. ISBN: 9780262012119.

[12] M. Altini, J. Penders, R. Vullers, and O. Amft. “Combining wearable accelerometer and physiological data for activity and energy expenditure estimation”. In: *Proceedings of the 4th Conference on Wireless Health - WH ’13*. Ed. by R. Jafari and W. Nilsen. ACM Press, 2013, pp. 1–8. DOI: 10.1145/2534088.2534106.

[13] M. Altini, J. Penders, and O. Amft. “Energy expenditure estimation using wearable sensors”. In: *Proceedings of the conference on Wireless Health - WH ’12*. Ed. by W. Kaiser and R. McCray. ACM Press, 2012, pp. 1–8. DOI: 10.1145/2448096.2448097.

[14] M. Altini, J. Penders, R. Vullers, and O. Amft. “Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning”. In: *IEEE journal of biomedical and health informatics* 19.1 (2015), pp. 219–226. DOI: 10.1109/JBHI.2014.2313039.

[15] M. Altini, P. Casale, J. F. Penders, and O. Amft. “Personalization of Energy Expenditure Estimation in Free Living Using Topic Models”. In: *IEEE journal of biomedical and health informatics* 19.5 (2015), pp. 1577–1586. DOI: 10.1109/JBHI.2015.2418256.

[16] M. Altini, P. Casale, J. Penders, and O. Amft. “Personalized cardiorespiratory fitness and energy expenditure estimation using hierarchical Bayesian models”. In: *Journal of biomedical informatics* 56 (2015), pp. 195–204. DOI: 10.1016/j.jbi.2015.06.008.

[17] S. Amalan, A. Shyam, A. S. Anusha, S. P. Preejith, A. Tony, J. Jayaraj, and S. Mohanasankar. “Electrodermal Activity based Classification of Induced Stress in a Controlled Setting”. In: *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, 11.06.2018 - 13.06.2018, pp. 1–6. DOI: 10.1109/MeMeA.2018.8438703.

[18] O. Amft. “How Wearable Computing Is Shaping Digital Health”. In: *IEEE Pervasive Computing* 17.1 (2018), pp. 92–98. DOI: 10.1109/MPRV.2018.011591067.

[19] O. Amft and K. van Laerhoven. “What Will We Wear After Smartphones?” In: *IEEE Pervasive Computing* 16.4 (2017), pp. 80–85. DOI: 10.1109/MPRV.2017.3971124.

[20] *AN4426; Application note: Tutorial for MEMS microphones*. STMicroelectronics, 2015. URL: https://www.st.com/resource/en/application_note/dm00103199.pdf.

[21] R. B. Andrews. “Indices of heart rate as substitutes for respiratory calorimetry”. In: *American Industrial Hygiene Association journal* 27.6 (1966), pp. 526–532. DOI: 10.1080/00028896609342466.

[22] *Apollo MCU; Datasheet: DA-A1-1p00*. Ambiq Micro, 2017. URL: https://www.fujitsu.com/uk/Images/Apollo_MCU_Data_Sheet_DS-A1-1p00.pdf.

[23] A. Apvrille. *Geek usages for your Fitbit Flex tracker*. Oct. 21, 2015. URL: http://archive.hack.lu/2015/fitbit-hacklu-slides.pdf.

[24] F. de Arriba-Pérez, M. Caeiro-Rodríguez, and J. M. Santos-Gago. “Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios”. In: *Sensors* 16.9 (2016). DOI: 10.3390/s16091538.

[25] K. Ashton. “That ‘internet of things’ thing”. In: *RFID journal* 22.7 (2009), pp. 97–114. URL: http://www.itrco.jp/libraries/RFIDjournal-ThatInternetofThingsThing.pdf.

[26] M. Bariya, H. Y. Y. Nyein, and A. Javey. “Wearable sweat sensors”. In: *Nature Electronics* 1.3 (2018), pp. 160–171. DOI: 10.1038/s41928-018-0043-y.

[27] E. J. Bassey, H. M. Dallosso, P. H. Fentem, J. M. Irving, and J. M. Patrick. “Validation of a simple mechanical accelerometer (pedometer) for the estimation of walking activity”. In: *European Journal of Applied Physiology and Occupational Physiology* 56.3 (1987), pp. 323–330. DOI: 10.1007/BF00690900.

[28] C. Beijst, G. Schep, E. van Breda, P. F. F. Wijn, and C. van Pul. “Accuracy and precision of CPET equipment: A comparison of breath-by-breath and mixing chamber systems”. In: *Journal of medical engineering & technology* 37.1 (2013), pp. 35–42. DOI: 10.3109/03091902.2012.733057.

[29] G. Bhat, R. Deb, and U. Y. Ogras. “OpenHealth: Open Source Platform for Wearable Health Monitoring”. In: *IEEE Design & Test* (2019), p. 1. DOI: 10.1109/MDAT.2019.2906110.

[30] T. Bhowmik, J. Dey, and V. N. Tiwari. “A novel method for accurate estimation of HRV from smartwatch PPG signals”. In: *IEEE Engineering in Medicine and Biology Society* 2017 (2017), pp. 109–112. DOI: 10.1109/EMBC.2017.8036774.

[31] H.-K. Biesalski, P. Grimm, and S. Nowitzki-Grimm. *Taschenatlas der Ernährung*. 5th ed. Thieme, 2011. ISBN: 9783131153555.

[32] Bill Riedel. “A Surface-Micromachined, Monolithic Accelerometer: ADXL50: the first commercially available surface-micromachined device. All signal-conditioning circuitry is on-chip”. In: *Analog Dialogue* 27.2 (1993), pp. 3–7. ISSN: 0161-3626.

[33] K. Blazejczyk, Y. EPSTEIN, G. Jendritzky, H. Staiger, and B. Tinz. “Comparison of UTCI to selected thermal indices”. In: *International journal of biometeorology* 56.3 (2012), pp. 515–535. DOI: 10.1007/s00484-011-0453-2.

[34] P. D. Bliese, J. R. Edwards, and S. Sonnentag. “Stress and well-being at work: A century of empirical trends reflecting theoretical and societal influences”. In: *The Journal of applied psychology* 102.3 (2017), pp. 389–402. DOI: 10.1037/apl0000109.

[35] *BME280; Datasheet: Combined humidity and pressure sensor*. Bosch Sensortec, 2018. URL: https://ae-bst.resource.bosch.com/media/_tech/media/datasheets/BST-BME280-DS002.pdf.

[36] *BNO055; Datasheet: Intelligent 9-axis absolute orientation sensor*. Bosch Sensortec, 2016. URL: https://ae-bst.resource.bosch.com/media/_tech/media/datasheets/BST-BNO055-DS000.pdf.

[37] M. E. Bodner and E. C. Rhodes. “A Review of the Concept of the Heart Rate Deflection Point”. In: *Sports Medicine* 30.1 (2000), pp. 31–46. DOI: 10.2165/00007256-200030010-00004.

[38] G. A. Borg. “Psychophysical bases of perceived exertion”. In: *Medicine & Science in Sports & Exercise* 14.5 (May 1982), pp. 377–381. DOI: 10.1249/00005768-198205000-00012.

[39] R. Boushel. “Muscle metaboreflex control of the circulation during exercise”. In: *Acta physiologica (Oxford, England)* 199.4 (2010), pp. 367–383. DOI: 10.1111/j.1748-1716.2010.02133.x.

[40] N. A. Bowling and C. Kirkendall. “Workload: A Review of Causes, Consequences, and Potential Interventions”. In: *Contemporary occupational health psychology*. Ed. by J. Houdmont, S. Leka, and R. R. Sinclair. WiBlackwell, 2012, pp. 221–238. DOI: 10.1002/9781119942849.ch13.

[41] A. Bozkurt. “Wearable technology: The current state of the art”. In: *EDULEARN17 Proceedings*. Ed. by L. Gómez Chova, A. López Martínez, and I. Candel Torres. EDULEARN proceedings. IATED, 2017, pp. 8704–8711. DOI: 10.21125/edulearn.2017.0631.

[42] *BQ25120A; Datasheet: Low IQ Highly Integrated Battery Charge Management Solution for Wearables and IoT*. Texas Instruments, 2016. URL: http://www.ti.com/lit/ds/symlink/bq25120a.pdf.

[43] S. Brage, N. Brage, P. W. Franks, U. Ekelund, and N. J. Wareham. “Reliability and validity of the combined heart rate and movement sensor Actiheart”. In: *European journal of clinical nutrition* 59.4 (2005), pp. 561–570. DOI: 10.1038/sj.ejcn.1602118.

[44] K.-J. Brickwood, G. Watson, J. O'Brien, and A. D. Williams. “Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis”. In: *JMIR mHealth and uHealth* 7.4 (Apr. 2019). DOI: 10.2196/11819.

[45] G. E. Burch and N. P. DePasquale. *A history of electrocardiography*. Reprint. Norman Publishing, 1990. ISBN: 978-0930405212.

[46] N. M. Byrne, A. P. Hills, G. R. Hunter, R. L. Weinsier, and Y. Schutz. “Metabolic equivalent: one size does not fit all”. In: *Journal of applied physiology (Bethesda, Md.: 1985)* 99.3 (2005), pp. 1112–1119. DOI: 10.1152/japplphysiol.00023.2004.

[47] J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson. *Handbook of Psychophysiology*. Ed. by J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson. 3rd ed. Cambridge University Press, Dec. 2016. DOI: 10.1017/9781107415782.

[48] B. Cain. *A review of the mental workload literature*. Ed. by Defence Research and Development Canada Toronto. North Atlantic Treaty Organization, Research and Technology Organization, 2007. URL: https://apps.dtic.mil/dtic/tr/fulltext/u2/a474193.pdf.

[49] C. Campbell and P. Tarasewich. “What Can You Say with Only Three Pixels?” In: *Mobile Human-Computer Interaction - MobileHCI 2004*. Ed. by D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, S. Brewster, and M. Dunlop. Vol. 3160. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, pp. 1–12. DOI: 10.1007/978-3-540-28637-0_1.

[50] Canalys, ed. *Media alert: Apple Watch has its best quarter and takes nearly 80% of total smartwatch revenue in Q4: Press release 2017/1631: Press release 2017/1631*. July 2017. URL: https://www.canalys.com/static/press_release/2017/media-alert-070217-apple-watch-has-its-best-quarter-and-takes-nearly-80-total-smartwatch-revenue-q4.pdf.

[51] D. P. Cardinali. *Autonomic Nervous System*. Springer International Publishing, 2018. DOI: 10.1007/978-3-319-57571-1.

[52] A. J. Casson. “Opportunities and challenges for ultra low power signal processing in wearable healthcare”. In: *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 424–428. DOI: 10.1109/eusipco.2015.7362418.

[53] P. Castiglioni, P. Meriggi, A. Faini, and M. Di Rienzo. "Cepstral based approach for online quantification of ECG quality in freely moving subjects". In: *Computing in Cardiology (CinC)*. Ed. by P. Castiglioni, P. Meriggi, A. Faini, and M. Di Rienzo. 2011. ISBN: 9781457706127.

[54] R. Castrillón, J. J. Pérez, and H. Andrade-Caicedo. "Electrical performance of PEDOT:PSS-based textile electrodes for wearable ECG monitoring: a comparative study". In: *Biomedical engineering online* 17.1 (2018), p. 38. DOI: 10.1186/s12938-018-0469-5.

[55] C. Catal and A. Akbulut. "Automatic energy expenditure measurement for health science". In: *Computer methods and programs in biomedicine* 157 (2018), pp. 31–37. DOI: 10.1016/j.cmpb.2018.01.015.

[56] S. M. Ceesay, A. M. Prentice, K. C. Day, P. R. Murgatroyd, G. R. Goldberg, W. Scott, and G. B. Spurr. "The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry". In: *British Journal of Nutrition* 61.02 (1989), p. 175. DOI: 10.1079/BJN19890107.

[57] M. Chan, D. Estève, J.-Y. Fourniols, C. Escriba, and E. Campo. "Smart wearable systems: current status and future challenges". In: *Artificial intelligence in medicine* 56.3 (2012), pp. 137–156. DOI: 10.1016/j.artmed.2012.09.003.

[58] K. Y. Chen and M. Sun. "Improving energy expenditure estimation by using a triaxial accelerometer". In: *Journal of Applied Physiology* 83.6 (Dec. 1997), pp. 2112–2122. DOI: 10.1152/jappl.1997.83.6.2112.

[59] D. Chisholm, K. Sweeny, P. Sheehan, B. Rasmussen, F. Smit, P. Cuijpers, and S. Saxena. "Scaling-up treatment of depression and anxiety: A global return on investment analysis". In: *The Lancet Psychiatry* 3.5 (2016), pp. 415–424. DOI: 10.1016/S2215-0366(16)30024-4.

[60] B. C. K. Choi and A. W. P. Pak. "A catalog of biases in questionnaires". In: *Preventing chronic disease* 2.1 (2005). URL: https://www.cdc.gov/pcd/issues/2005/jan/pdf/04_0050.pdf.

[61] J. Choi, B. Ahmed, and R. Gutierrez-Osuna. "Development and Evaluation of an Ambulatory Stress Monitor Based on Wearable Sensors". In: *IEEE Transactions on Information Technology in Biomedicine* 16.2 (2012), pp. 279–286. DOI: 10.1109/TITB.2011.2169804.

[62] J. Choi and R. Gutierrez-Osuna. "Using Heart Rate Monitors to Detect Mental Stress". In: *Sixth International Workshop on Wearable and Implantable Body Sensor Networks*. 2009, pp. 219–223. DOI: 10.1109/BSN.2009.13.

[63] E. A. Chowdhury, M. J. Western, T. E. Nightingale, O. J. Peacock, and D. Thompson. "Assessment of laboratory and daily energy expenditure estimates from consumer multisensor physical activity monitors". In: *PLOS ONE* 12.2 (Feb. 2017). Ed. by L. P. Ardigò. DOI: 10.1371/journal.pone.0171720.

[64] P. Christ. "Detektion und Analyse physiologischer und biokinematischer Parameter mit Körpersensoren". Dissertation. Bielefeld University, 2016.

[65] P. Christ, G. Sievers, J. Einhaus, T. Jungeblut, M. Porrmann, and U. Ruckert. "Pareto-optimal signal processing on low-power microprocessors". In: *2013 IEEE SENSORS*. IEEE, 2013, pp. 1–4. DOI: 10.1109/ICSENS.2013.6688593.

[66] G. D. Clifford, C. Liu, B. Moody, L.-W. H. Lehman, I. Silva, Q. Li, A. E. Johnson, and R. G. Mark. "AF Classification from a Short Single Lead ECG Recording: the PhysioNet/-Computing in Cardiology Challenge 2017". In: *Computing in cardiology* 44 (2017). DOI: 10.22489/CinC.2017.065-469.

[67] W. H. Close, M. J. Dauncey, and D. L. Ingram. "Heat loss from humans measured with a direct calorimeter and heat-flow meters". In: *British Journal of Nutrition* 43.1 (Jan. 1980), pp. 87–93. DOI: 10.1079/bjn19800067.

[68] F. S. Cotton and D. B. Dill. "ON THE RELATION BETWEEN THE HEART RATE DURING EXERCISE AND THAT OF THE IMMEDIATE POST-EXERCISE PERIOD". In: *American Journal of Physiology-Legacy Content* 111.3 (1935), pp. 554–556. DOI: 10.1152/ajplegacy.1935.111.3.554.

[69] S. E. Crouter and D. R. Bassett. "A new 2-regression model for the Actical accelerometer". In: *British journal of sports medicine* 42.3 (2008), pp. 217–224. DOI: 10.1136/bjsm.2006.033399.

[70] M. Csikszentmihalyi and R. Larson. "Validity and Reliability of the Experience-Sampling Method". In: *Flow and the Foundations of Positive Psychology*. Ed. by M. Csikszentmihalyi. Springer Netherlands, 2014, pp. 35–54. DOI: 10.1007/978-94-017-9088-8_3.

[71] D. Curone, A. Tognetti, E. L. Secco, G. Anania, N. Carbonaro, D. D. Rossi, and G. Magenes. "Heart Rate and Accelerometer Data Fusion for Activity Assessment of Rescuers During Emergency Interventions". In: *IEEE Transactions on Information Technology in Biomedicine* 14.3 (May 2010), pp. 702–710. DOI: 10.1109/titb.2010.2047727.

[72] B. Cvetkovic, R. Milic, and M. Lustrek. "Estimating Energy Expenditure With Multiple Models Using Different Wearable Sensors". In: *IEEE journal of biomedical and health informatics* 20.4 (2016), pp. 1081–1087. DOI: 10.1109/JBHI.2015.2432911.

[73] C. Daluwatte, L. Johannesen, L. Galeotti, J. Vicente, D. G. Strauss, and C. G. Scully. "Assessing ECG signal quality indices to discriminate ECGs with artefacts from pathologically different arrhythmic ECGs". In: *Physiological measurement* 37.8 (2016), pp. 1370–1382. DOI: 10.1088/0967-3334/37/8/1370.

[74] B. Di Martino, K.-C. Li, L. T. Yang, and A. Esposito. *Internet of Everything*. Springer Singapore, 2018. DOI: 10.1007/978-981-10-5861-5.

[75] D. M. Diamond, A. M. Campbell, C. R. Park, J. Halonen, and P. R. Zoladz. "The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law". In: *Neural plasticity* 2007 (2007), p. 60803. DOI: 10.1155/2007/60803.

[76] K. P. Dowd, R. Szeklicki, M. A. Minetto, M. H. Murphy, A. Polito, E. Ghigo, H. van der Ploeg, U. Ekelund, J. Maciaszek, R. Stemplewski, M. Tomczak, and A. E. Donnelly. "A systematic literature review of reviews on techniques for physical activity measurement in adults: A DEDIPAC study". In: *The international journal of behavioral nutrition and physical activity* 15.1 (2018), p. 15. DOI: 10.1186/s12966-017-0636-2.

[77] R. C. Drew. “Baroreflex and neurovascular responses to skeletal muscle mechanoreflex activation in humans: an exercise in integrative physiology”. In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 313.6 (Dec. 2017). DOI: 10.1152/ajpregu.00242.2017.

[78] C. L. Edwardson, E. A. H. Winkler, D. H. Bodicoat, T. Yates, M. J. Davies, D. W. Dunstan, and G. N. Healy. “Considerations when using the activPAL monitor in field-based research with adult populations”. In: *Journal of sport and health science* 6.2 (2017), pp. 162–178. DOI: 10.1016/j.jshs.2016.02.002.

[79] F. El-Amrawy and M. I. Nounou. “Are Currently Available Wearable Devices for Activity Tracking and Heart Rate Monitoring Accurate, Precise, and Medically Beneficial?” In: *Healthcare informatics research* 21.4 (2015), pp. 315–320. DOI: 10.4258/hir.2015.21.4.315.

[80] P. Electro. *Polar - Innovations - Who we are @ONLINE*. 2018. URL: https://www.polar.com/us-en/about_polar/who_we_are/innovations (visited on 08/24/2018).

[81] M. Elgendi, M. Meo, and D. Abbott. “A Proof-of-Concept Study: Simple and Effective Detection of P and T Waves in Arrhythmic ECG Signals”. In: *Bioengineering (Basel, Switzerland)* 3.4 (2016). DOI: 10.3390/bioengineering3040026.

[82] M. Elgendi, B. Eskofier, S. Dokos, and D. Abbott. “Revisiting QRS Detection Methodologies for Portable, Wearable, Battery-Operated, and Wireless ECG Systems”. In: *PLoS ONE* 9.1 (Jan. 2014). Ed. by L. A. N. Amaral. DOI: 10.1371/journal.pone.0084018.

[83] K. Ellis, J. Kerr, S. Godbole, G. Lanckriet, D. Wing, and S. Marshall. “A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers”. In: *Physiological measurement* 35.11 (2014), pp. 2191–2203. DOI: 10.1088/0967-3334/35/11/2191.

[84] X. Fafoutis, L. Marchegiani, A. Elsts, J. Pope, R. Piechocki, and I. Craddock. “Extending the battery lifetime of wearable sensors with embedded machine learning”. In: *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*. IEEE, 2018, pp. 269–274. DOI: 10.1109/wf-iot.2018.8355116.

[85] L. M. Feehan, J. Geldman, E. C. Sayre, C. Park, A. M. Ezzat, J. Y. Yoo, C. B. Hamilton, and L. C. Li. “Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data”. In: *JMIR mHealth and uHealth* 6.8 (Aug. 2018). DOI: 10.2196/10527.

[86] G. H. Forman and J. Zahorjan. “The challenges of mobile computing”. In: *Computer* 27.4 (1994), pp. 38–47. DOI: 10.1109/2.274999.

[87] M. Frankenhaeuser and B. Gardell. “Underload and overload in working life: Outline of a multidisciplinary approach”. In: *Journal of human stress* 2.3 (1976), pp. 35–46. DOI: 10.1080/0097840X.1976.9936068.

[88] P. S. Freedson, E. Melanson, and J. Sirard. “Calibration of the Computer Science and Applications, Inc. accelerometer”. In: *Medicine & Science in Sports & Exercise* 30.5 (May 1998), pp. 777–781. DOI: 10.1097/00005768-199805000-00021.

[89] P. S. Freedson and K. Miller. “Objective monitoring of physical activity using motion sensors and heart rate”. In: *Research quarterly for exercise and sport* 71 Suppl 2 (2000), pp. 21–29. DOI: 10.1080/02701367.2000.11082782.

[90] G. M. Friesen, T. C. Jannett, M. A. Jadallah, S. L. Yates, S. R. Quint, and H. T. Nagle. “A comparison of the noise sensitivity of nine QRS detection algorithms”. In: *Biomedical Engineering, IEEE Transactions on* 37.1 (Jan. 1990), pp. 85–98. DOI: 10.1109/10.43620.

[91] E. Galy, J. Paxion, and C. Berthelon. “Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: an example with driving”. In: *Ergonomics* 61.4 (2018), pp. 517–527. DOI: 10.1080/00140139.2017.1369583.

[92] F. X. Gamelin, S. Berthoin, and L. Bosquet. “Validity of the polar S810 heart rate monitor to measure R-R intervals at rest”. In: *Medicine and science in sports and exercise* 38.5 (2006), pp. 887–893. DOI: 10.1249/01.mss.0000218135.79476.9c.

[93] K.-R. Geiß and M. Hamm. *Handbuch Sportler-Ernährung*. ger. Behr, 1990. ISBN: 3-925673-71-7.

[94] A. K. Ghosh. “Anaerobic threshold: its concept and role in endurance sport”. In: *The Malaysian journal of medical sciences: MJMS* 11.1 (2004). PMC3438148, pp. 24–36. ISSN: 1394-195X.

[95] H. Gjoreski, B. Kaluža, M. Gams, R. Milić, and M. Luštrek. “Context-based ensemble method for human energy expenditure estimation”. In: *Applied Soft Computing* 37 (2015), pp. 960–970. DOI: 10.1016/j.asoc.2015.05.001.

[96] A. L. Goldberger, L. A. N. 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”. In: *Circulation* 101.23 (June 2000). DOI: 10.1161/01.cir.101.23.e215.

[97] R. Gravina, P. Alinia, H. Ghasemzadeh, and G. Fortino. “Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges”. In: *Information Fusion* 35 (2017), pp. 68–80. DOI: 10.1016/j.inffus.2016.09.005.

[98] N. K. Gupta. *Inside Bluetooth Low Energy, Second Edition*. Artech House, 2016. ISBN: 978-1-63081-370-3.

[99] D. Hamar. “Advances in the Computerized Breath-by-Breath Method”. In: *Advances in Ergometry*. Ed. by N. Bachl. Springer Berlin Heidelberg, 1991, pp. 289–302. DOI: 10.1007/978-3-642-76442-4_40.

[100] U. Hansmann, L. Merk, M. S. Nicklous, and T. Stober. *Pervasive Computing Handbook*. Springer, 2001. DOI: 10.1007/978-3-662-04318-9.

[101] J. Hanson. “Rotations in three, four, and five dimensions”. In: *ArXiv e-prints* (2011). URL: http://arxiv.org/abs/1103.5263.

[102] J. A. Harris and F. G. Benedict. “A Biometric Study of Human Basal Metabolism”. In: *Proceedings of the National Academy of Sciences* 4.12 (1918), pp. 370–373. DOI: 10.1073/pnas.4.12.370.

[103] S. G. Hart and L. E. Staveland. “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research”. In: *Human Mental Workload*. Vol. 52. North-Holland, 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9.

[104] T. Hastie, R. Tibshirani, and J. H. Friedman. *The elements of statistical learning: Data mining, inference, and prediction*. Second edition, corrected at 12th printing 2017. Springer series in statistics. Springer, 2017. ISBN: 978-0-387-84857-0.

[105] H. A. Haugen, L.-N. Chan, and F. Li. “Indirect calorimetry: a practical guide for clinicians”. In: *Nutrition in clinical practice: official publication of the American Society for Parenteral and Enteral Nutrition* 22.4 (2007), pp. 377–388. DOI: 10.1177/0115426507022004377.

[106] G. Havenith and D. Fiala. “Thermal Indices and Thermophysiological Modeling for Heat Stress”. In: *Comprehensive Physiology* 6.1 (2015), pp. 255–302. DOI: 10.1002/cphy.c140051.

[107] J. He, W. Zhou, H. Yu, X. He, and P. Peng. “Structural Designing of a MEMS Capacitive Accelerometer for Low Temperature Coefficient and High Linearity”. In: *Sensors* 18.2 (2018). DOI: 10.3390/s18020643.

[108] J. He, E. Park, J. Li, and E. Kim. “Physiological and psychological responses while wearing firefighters’ protective clothing under various ambient conditions”. In: *Textile Research Journal* 87.8 (2016), pp. 929–944. DOI: 10.1177/0040517516641364.

[109] J. A. Healey and R. W. Picard. “Detecting Stress During Real-World Driving Tasks Using Physiological Sensors”. In: *IEEE Transactions on Intelligent Transportation Systems* 6.2 (2005), pp. 156–166. DOI: 10.1109/TITS.2005.848368.

[110] D. P. Heil. “Estimating energy expenditure in wildland fire fighters using a physical activity monitor”. In: *Applied ergonomics* 33.5 (2002), pp. 405–413. DOI: 10.1016/s0003-6870(02)00042-x.

[111] I. P. Herman. “Metabolism: Energy, Heat, Work, and Power of the Body”. In: *Physics of the Human Body*. Ed. by I. P. Herman. Vol. 91. Biological and Medical Physics, Biomedical Engineering. Springer International Publishing, 2016, pp. 393–489. DOI: 10.1007/978-3-319-23932-3_6.

[112] J. Hernandez, P. Paredes, A. Roseway, and M. Czerwinski. “Under pressure: sensing stress of computer users”. In: *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI ’14*. Ed. by M. Jones, P. Palanque, A. Schmidt, and T. Grossman. ACM Press, 2014, pp. 51–60. DOI: 10.1145/2556288.2557165.

[113] A. P. Hills, N. Mokhtar, and N. M. Byrne. “Assessment of physical activity and energy expenditure: an overview of objective measures”. In: *Frontiers in nutrition* 1 (2014), p. 5. DOI: 10.3389/fnut.2014.00005.

[114] B. Hoefflinger. “From Microelectronics to Nanoelectronics”. In: *Chips 2020*. Ed. by B. Hoefflinger. The Frontiers Collection. Springer Berlin Heidelberg, 2012, pp. 13–36. DOI: 10.1007/978-3-642-23096-7_2.

[115] *In-Amp Input RFI Protection: MT-070: Tutorial*. Analog Devices, Inc., 2008. URL: https://www.analog.com/media/en/training-seminars/tutorials/MT-070.pdf.

[116] H. Isshiki and Y. Yamamoto. “Instrument for monitoring arousal level using electrodermal activity”. In: *Proceedings of IEEE International Conference on Instrumentation and Measurement Technology*. IEEE, 1994, pp. 975–978. DOI: 10.1109/IMTC.1994.351943.

[117] M. Jeon, ed. *Emotions and affect in human factors and human-computer interaction*. Academic Press is an imprint of Elsevier, 2017. DOI: 10.1016/B978-0-12-801851-4.00025-2.

[118] P. G. A. M. Jorna. “Spectral analysis of heart rate and psychological state: A review of its validity as a workload index”. In: *Biological Psychology* 34.2-3 (Nov. 1992), pp. 237–257. DOI: 10.1016/0301-0511(92)90017-o.

[119] Y. Jung, S. Kim, and B. Choi. “Consumer valuation of the wearables: The case of smartwatches”. In: *Computers in Human Behavior* 63 (2016), pp. 899–905. DOI: 10.1016/j.chb.2016.06.040.

[120] M. Kachuee, S. Fazeli, and M. Sarrafzadeh. “ECG Heartbeat Classification: A Deep Transferable Representation”. In: *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 2018, pp. 443–444. DOI: 10.1109/ichi.2018.00092.

[121] K. Kaewkannate and S. Kim. “A comparison of wearable fitness devices”. In: *BMC Public Health* 16.1 (May 2016), p. 433. DOI: 10.1186/s12889-016-3059-0.

[122] R. Kamaleswaran, R. Mahajan, and O. Akbilgic. “A robust deep convolutional neural network for the classification of abnormal cardiac rhythm using single lead electrocardiograms of variable length”. In: *Physiological measurement* 39.3 (2018), p. 035006. DOI: 10.1088/1361-6579/aaaa9d.

[123] P. Karthikeyan, M. Murugappan, and S. Yaacob. “Detection of Human stress using Short-Term ECG and HRV signals”. In: *Journal of Mechanics in Medicine and Biology* 13.02 (Apr. 2013), p. 1350038. DOI: 10.1142/s0219519413500383.

[124] R. J. Kate, A. M. Swartz, W. A. Welch, and S. J. Strath. “Comparative evaluation of features and techniques for identifying activity type and estimating energy cost from accelerometer data”. In: *Physiological measurement* 37.3 (2016), pp. 360–379. DOI: 10.1088/0967-3334/37/3/360.

[125] G. P. Kenny, S. R. Notley, and D. Gagnon. “Direct calorimetry: a brief historical review of its use in the study of human metabolism and thermoregulation”. In: *European journal of applied physiology* 117.9 (2017), pp. 1765–1785. DOI: 10.1007/s00421-017-3670-5.

[126] L. R. Keytel, J. H. Goedecke, T. D. Noakes, H. Hiiloskorpi, R. Laukkanen, L. van der Merwe, and E. V. Lambert. “Prediction of energy expenditure from heart rate monitoring during submaximal exercise”. In: *Journal of sports sciences* 23.3 (2005), pp. 289–297. DOI: 10.1080/02640410470001730089.

[127] M. A. Khawaja, F. Chen, and N. Marcus. “Measuring Cognitive Load Using Linguistic Features: Implications for Usability Evaluation and Adaptive Interaction Design”. In: *International Journal of Human-Computer Interaction* 30.5 (2014), pp. 343–368. DOI: 10.1080/10447318.2013.860579.

[128] R. Khudorozhkov and D. Podvyaznikov. *Benchmarks of ResNet Architecture for Atrial Fibrillation Classification*. Sept. 2018. URL: http://arxiv.org/pdf/1810.00396v1.

[129] S. Kim, R. Vyas, J. Bito, K. Niotaki, A. Collado, A. Georgiadis, and M. M. Tentzeris. "Ambient RF Energy-Harvesting Technologies for Self-Sustainable Standalone Wireless Sensor Platforms". In: *Proceedings of the IEEE* 102.11 (2014), pp. 1649–1666. DOI: 10.1109/JPROC.2014.2357031.

[130] P. Kligfield. "The centennial of the Einthoven electrocardiogram". In: *Journal of electrocardiology* 35 Suppl (2002), pp. 123–129. DOI: 10.1054/jelc.2002.37169.

[131] N. C. Krishnan and S. Panchanathan. "Analysis of low resolution accelerometer data for continuous human activity recognition". In: *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 3337–3340. DOI: 10.1109/ICASSP.2008.4518365.

[132] S. Kupschick, M. Pendzich, D. Gardas, T. Jürgensohn, S. Wischniewski, and L. Adolph. *Predicting firefighters' exertion based on machine learning techniques*. 2016. DOI: 10.21934/baua:focus20161107.

[133] E. Lanzola, A. Tagliabue, and H. Cena. "Skin temperature and energy expenditure". In: *Annals of nutrition & metabolism* 34.5 (1990), pp. 311–316. DOI: 10.1159/000177603.

[134] J. Lasek and M. Gagolewski. "The Winning Solution to the AAIA'15 Data Mining Competition: Tagging Firefighter Activities at a Fire Scene". In: *2015 Federated Conference on Computer Science and Information Systems*. Annals of Computer Science and Information Systems. IEEE, 2015, pp. 375–380. DOI: 10.15439/2015F418.

[135] E. S. Lee, H. C. Shin, J. H. Lee, Y. J. Yang, J. J. Cho, G. Ahn, Y. S. Yoon, and E. Sung. "Development of the Perceived Stress Inventory: A New Questionnaire for Korean Population Surveys". In: *Korean journal of family medicine* 36.6 (2015), pp. 286–293. DOI: 10.4082/kjfm.2015.36.6.286.

[136] S. Lee, M. Liu, S. Jun, S. Xu, J. Kim, and Arvind. "Application-Managed Flash". In: *14th USENIX Conference on File and Storage Technologies (FAST 16)*. USENIX Association. 2016, pp. 339–353. ISBN: 978-1-931971-28-7.

[137] J. A. Levine. "Measurement of energy expenditure". In: *Public Health Nutrition* 8.7a (2005), p. 79. DOI: 10.1079/PHN2005800.

[138] Q. Li, R. G. Mark, and G. D. Clifford. "Robust heart rate estimation from multiple asynchronous noisy sources using signal quality indices and a Kalman filter". In: *Physiological Measurement* 29.1 (Jan. 2007), pp. 15–32. DOI: 10.1088/0967-3334/29/1/002.

[139] R. Li, P. Deurenberg, and J. G. Hautvast. "A critical evaluation of heart rate monitoring to assess energy expenditure in individuals". In: *The American journal of clinical nutrition* 58.5 (1993), pp. 602–607. DOI: 10.1093/ajcn/58.5.602.

[140] R. T. Li, S. R. Kling, M. J. Salata, S. A. Cupp, J. Sheehan, and J. E. Voos. "Wearable Performance Devices in Sports Medicine". In: *Sports health* 8.1 (2016), pp. 74–78. DOI: 10.1177/1941738115616917.

[141] X. Li, Z. Chen, Q. Liang, and Y. Yang. "Analysis of Mental Stress Recognition and Rating Based on Hidden Markov Model". In: *Journal of Computational Information Systems* 10.18 (2014), pp. 7911–7919. DOI: 10.12733/jcis11559.

[142] C.-W. Lin, Y.-T. C. Yang, J.-S. Wang, and Y.-C. Yang. “A wearable sensor module with a neural-network-based activity classification algorithm for daily energy expenditure estimation”. In: *IEEE transactions on information technology in biomedicine: a publication of the IEEE Engineering in Medicine and Biology Society* 16.5 (2012), pp. 991–998. DOI: 10.1109/titb.2012.2206602.

[143] *LIS2DE12; Datasheet: MEMS digital output motion sensor: ultra-low-power high-performance 3-axis "femto" accelerometer*. STMicroelectronics, 2017. URL: https://www.st.com/resource/en/datasheet/lis2de12.pdf.

[144] S. Liu, R. X. Gao, and P. S. Freedson. “Computational methods for estimating energy expenditure in human physical activities”. In: *Medicine and science in sports and exercise* 44.11 (2012), pp. 2138–2146. DOI: 10.1249/MSS.0b013e31825e825a.

[145] J. Lu, H. Jia, N. Verma, and N. K. Jha. “Genetic Programming for Energy-Efficient and Energy-Scalable Approximate Feature Computation in Embedded Inference Systems”. In: *IEEE Transactions on Computers* 67.2 (2018), pp. 222–236. DOI: 10.1109/TC.2017.2738642.

[146] K. Lu, L. Yang, F. Seoane, F. Abtahi, M. Forsman, and K. Lindecrantz. “Fusion of Heart Rate, Respiration and Motion Measurements from a Wearable Sensor System to Enhance Energy Expenditure Estimation”. In: *Sensors (Basel, Switzerland)* 18.9 (2018). DOI: 10.3390/s18093092.

[147] C. Luo, D. L. Ware, J. B. Zwischenberger, and J. W. Clark. “A mechanical model of the human heart relating septal function to myocardial work and energy”. In: *Cardiovascular engineering (Dordrecht, Netherlands)* 8.3 (2008), pp. 174–184. DOI: 10.1007/s10558-008-9054-z.

[148] K. Lyden, S. L. Kozey, J. W. Staudenmeyer, and P. S. Freedson. “A comprehensive evaluation of commonly used accelerometer energy expenditure and MET prediction equations”. In: *European journal of applied physiology* 111.2 (2011), pp. 187–201. DOI: 10.1007/s00421-010-1639-8.

[149] K. Lyden, T. Swibas, V. Catenacci, R. Guo, N. Szuminsky, and E. L. Melanson. “Estimating energy expenditure using heat flux measured at a single body site”. In: *Medicine and science in sports and exercise* 46.11 (2014), pp. 2159–2167. DOI: 10.1249/MSS.0000000000000346.

[150] R. MacManus. *Health trackers: How technology is helping us monitor and improve our health*. First paperback edition. Rowman & Littlefield, 2015. ISBN: 081089582X.

[151] G. Magenes, D. Curone, L. Caldani, and E. L. Secco. *Fire fighters and rescuers monitoring through wearable sensors: The ProeTEX project: Buenos Aires, Argentina, Aug. 31, 2010 - Sept. 4, 2010*. Vol. 2010. IEEE, 2010. DOI: 10.1109/IEMBS.2010.5627452.

[152] H. Magne. “La depense d’energie dans la marche de 1’homme en terrain horizontal ou incliné”. In: *Journal de physiologie et de pathologie générale* 15 (1919), pp. 1154–1173. URL: https://gallica.bnf.fr/ark:/12148/bpt6k6361318m.

[153] L. Mainardi, L. Sörnmo, and S. Cerutti. “Understanding Atrial Fibrillation: The Signal Processing Contribution, Part II”. In: *Synthesis Lectures on Biomedical Engineering* 3.1 (2008), pp. 1–139. DOI: 10.2200/S00153ED1V01Y200809BME025.

[154] M. Malik, J. T. Bigger, A. J. Camm, R. E. Kleiger, A. Malliani, A. J. Moss, and P. J. Schwartz. "Heart rate variability: Standards of measurement, physiological interpretation, and clinical use". In: *European Heart Journal* 17.3 (1996), pp. 354–381. DOI: 10.1093/oxfordjournals.eurheartj.a014868.

[155] U. E. Manawadu, T. Kawano, S. Murata, M. Kamezaki, J. Muramatsu, and S. Sugano. "Multiclass Classification of Driver Perceived Workload Using Long Short-Term Memory based Recurrent Neural Network". In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 26.06.2018 - 30.06.2018, pp. 1–6. DOI: 10.1109/IVS.2018.8500410.

[156] P. Marwedel. *Embedded System Design*. Springer International Publishing, 2018. DOI: 10.1007/978-3-319-56045-8.

[157] C. Maslach, W. B. Schaufeli, and M. P. Leiter. "Job Burnout". In: *Annual Review of Psychology* 52.1 (2001), pp. 397–422. DOI: 10.1146/annurev.psych.52.1.397.

[158] A. Mason, N. Yazdi, K. Najafi, and K. D. Wise. "A Low-power Wireless Microinstrumentation System For Environmental Monitoring". In: *Proceedings of the International Solid-State Sensors and Actuators Conference - TRANSDUCERS '95*. IEEE, 1995. DOI: 10.1109/sensor.1995.717104.

[159] J. Masterson and F. A. Richardson. *Humidex: a method of quantifying human discomfort due to excessive heat and humidity*. Environment Canada, 1979.

[160] W. D. McArdle, F. I. Katch, and V. L. Katch. *Exercise physiology: Nutrition, energy and human performance*. 7th ed. Wolters Kluwer Health/Lippincott Williams & Wilkins, 2010. ISBN: 9780781797818.

[161] M. McCaul, A. Porter, R. Barrett, P. White, F. Stroiescu, G. Wallace, and D. Diamond. "Wearable Platform for Real-time Monitoring of Sodium in Sweat". In: *Chemphyschem: a European journal of chemical physics and physical chemistry* 19.12 (2018), pp. 1531–1536. DOI: 10.1002/cphc.201701312.

[162] *MEMS Pressure Sensor Comparison 2018*. System Plus Consulting Sarl, 2018. URL: https://www.systemplus.fr/reverse-costing-reports/mems-pressure-sensor-comparison-2018/.

[163] C. Menßen. "Multifaktorielle online Stresserkennung mit einem mobilen Endgerät". Master thesis. Bielefeld University, 2014.

[164] S. Michael, K. S. Graham, and G. M. Davis. "Cardiac Autonomic Responses during Exercise and Post-exercise Recovery Using Heart Rate Variability and Systolic Time Intervals-A Review". In: *Frontiers in physiology* 8 (2017), p. 301. DOI: 10.3389/fphys.2017.00301.

[165] M. D. Mifflin, S. T. St Jeor, L. A. Hill, B. J. Scott, S. A. Daugherty, and Y. O. Koh. "A new predictive equation for resting energy expenditure in healthy individuals". In: *The American journal of clinical nutrition* 51.2 (1990), pp. 241–247. DOI: 10.1093/ajcn/51.2.241.

[166] B. Mittelstadt. "Ethics of the health-related internet of things: a narrative review". In: *Ethics and Information Technology* 19.3 (2017), pp. 157–175. DOI: 10.1007/s10676-017-9426-4.

[167] H. J. Montoye, R. Washburn, S. Servais, A. Ertl, J. G. Webster, and F. J. Nagle. “Estimation of energy expenditure by a portable accelerometer”. In: *Medicine & Science in Sports & Exercise* 15.5 (1983), pp. 403–407. DOI: 10.1249/00005768-198315050-00010.

[168] *MP34DT04; Datasheet: MEMS audio sensor omnidirectional digital microphone*. STMicroelectronics, 2015. URL: https://www.st.com/resource/en/datasheet/MP34DT04.pdf.

[169] M. Nakanishi, S. Izumi, S. Nagayoshi, H. Kawaguchi, M. Yoshimoto, T. Shiga, T. Ando, S. Nakae, C. Usui, T. Aoyama, and S. Tanaka. “Estimating metabolic equivalents for activities in daily life using acceleration and heart rate in wearable devices”. In: *Biomedical engineering online* 17.1 (2018), p. 100. DOI: 10.1186/s12938-018-0532-2.

[170] N. J. Nedergaard, M. A. Robinson, E. Eusterwiemann, B. Drust, P. J. Lisboa, and J. Vanrenterghem. “The Relationship Between Whole-Body External Loading and Body-Worn Accelerometry During Team-Sport Movements”. In: *International journal of sports physiology and performance* 12.1 (2017), pp. 18–26. DOI: 10.1123/ijspp.2015-0712.

[171] R. Negra, I. Jemili, and A. Belghith. “Wireless Body Area Networks: Applications and Technologies”. In: *Procedia Computer Science* 83 (2016), pp. 1274–1281. DOI: 10.1016/j.procs.2016.04.266.

[172] M. B. Nelson, L. A. Kaminsky, D. C. Dickin, and A. H. K. Montoye. “Validity of Consumer-Based Physical Activity Monitors for Specific Activity Types”. In: *Medicine and science in sports and exercise* 48.8 (2016), pp. 1619–1628. DOI: 10.1249/MSS.0000000000000933.

[173] J. Nikolic-Popovic and R. Goubran. “Impact of motion artifacts on Heart Rate Variability measurements and classification performance”. In: *Medical Measurements and Applications (MeMeA), Proceedings of the IEEE International Symposium on*. 2013, pp. 156–159. DOI: 10.1109/MeMeA.2013.6549726.

[174] N. J. Nilsson. *Introduction to machine learning. An early draft of a proposed textbook*. Citeseer, 1996. URL: http://ai.stanford.edu/~nilsson/MLBOOK.pdf.

[175] L. Nokes, J. A. Fairclough, W. J. Mintowt-Czyz, I. Mackie, and J. Williams. “Vibration analysis of human tibia: The effect of soft tissue on the output from skin-mounted accelerometers”. In: *Journal of Biomedical Engineering* 6.3 (1984), pp. 223–226. DOI: 10.1016/0141-5425(84)90107-9.

[176] *nRF51822; Product Specification: Multiprotocol Bluetooth low energy/2.4 GHz RF System on Chip*. Nordic Semiconductor, 2016. URL: https://infocenter.nordicsemi.com/pdf/nRF51822_PS_v3.3.pdf.

[177] H. F. Nweke, Y. W. Teh, M. A. Al-garadi, and U. R. Alo. “Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges”. In: *Expert Systems with Applications* 105 (2018), pp. 233–261. DOI: 10.1016/j.eswa.2018.03.056.

[178] K. Ober. *Calculate Rotation Matrix to align Vector A to Vector B in 3d?* Mathematics Stack Exchange. (version: 2015-08-21). 2014. eprint: http://math.stackexchange.com/q/897677. URL: http://math.stackexchange.com/q/897677.

[179] L. Ong and J. A. Blumenthal. “Assessment of Physical Activity in Research and Clinical Practice”. In: *Handbook of Behavioral Medicine*. Ed. by A. Steptoe. Springer New York, 2010, pp. 31–48. DOI: 10.1007/978-0-387-09488-5_3.

[180] J. Pan and W. J. Tompkins. “A real-time QRS detection algorithm”. In: *IEEE Transactions on Biomedical Engineering* 32.3 (1985), pp. 230–236. DOI: 10.1109/TBME.1985.325532.

[181] S. S. Paul, A. Tiedemann, L. M. Hassett, E. Ramsay, C. Kirkham, S. Chagpar, and C. Sherrington. “Validity of the Fitbit activity tracker for measuring steps in community-dwelling older adults”. In: *BMJ Open Sport & Exercise Medicine* 1.1 (July 2015). DOI: 10.1136/bmjsem-2015-000013.

[182] M. Pedley. *Tilt Sensing Using a Three-Axis Accelerometer: Application Note*. Freescale Semiconductor, 2013. URL: https://cache.freescale.com/files/sensors/doc/app_note/AN3461.pdf.

[183] N. Pinheiro, R. Couceiro, J. Henriques, J. Muehlsteff, I. Quintal, L. Goncalves, and P. Carvalho. “Can PPG be used for HRV analysis?” In: *IEEE Engineering in Medicine and Biology Society* 2016 (2016), pp. 2945–2949. DOI: 10.1109/EMBC.2016.7591347.

[184] L. Piwek, D. A. Ellis, S. Andrews, and A. Joinson. “The Rise of Consumer Health Wearables: Promises and Barriers”. In: *PLOS Medicine* 13.2 (Feb. 2016). DOI: 10.1371/journal.pmed.1001953.

[185] T. Ploetz and J. Healey. “ISWC 2017: Riding the Waves of Wearables”. In: *IEEE Pervasive Computing* 17.2 (2018), pp. 78–83. DOI: 10.1109/MPRV.2018.022511248.

[186] H. Pontzer, R. Durazo-Arvizu, L. R. Dugas, J. Plange-Rhule, P. Bovet, T. E. Forrester, E. V. Lambert, R. S. Cooper, D. A. Schoeller, and A. Luke. “Constrained Total Energy Expenditure and Metabolic Adaptation to Physical Activity in Adult Humans”. In: *Current biology: CB* 26.3 (2016), pp. 410–417. DOI: 10.1016/j.cub.2015.12.046.

[187] N. Quadrianto, K. Kersting, and Z. Xu. “Gaussian Process”. In: *Encyclopedia of Machine Learning and Data Mining*. Ed. by C. Sammut and G. I. Webb. Vol. 6. Springer US, 2016, pp. 1–13. DOI: 10.1007/978-1-4899-7502-7_108-1.

[188] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri. “Heart rate variability: a review”. In: *Medical & biological engineering & computing* 44.12 (2006), pp. 1031–1051. DOI: 10.1007/s11517-006-0119-0.

[189] T. Rault, A. Bouabdallah, Y. Challal, and F. Marin. “A survey of energy-efficient context recognition systems using wearable sensors for healthcare applications”. In: *Pervasive and Mobile Computing* 37 (2017), pp. 23–44. DOI: 10.1016/j.pmcj.2016.08.003.

[190] P. Rawat, K. Singh, H. Chaouchi, and J. Bonnin. “Wireless sensor networks: a survey on recent developments and potential synergies”. English. In: *The Journal of Supercomputing* 68.1 (2014), pp. 1–48. DOI: 10.1007/s11227-013-1021-9.

[191] Raymond Chan. “Wireless heart rate monitoring system”. Pat. US20060247549A1. 2006.

[192] M. Redon. “Strategies for Choosing the Appropriate Microcontroller when Developing Ultra Low Power Systems”. In: *Analog Dialogue* 51.12 (2017), pp. 1–4. ISSN: 0161-3626.

[193] J.-L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita. "Transition-Aware Human Activity Recognition Using Smartphones". In: *Neurocomputing* 171 (2016), pp. 754–767. DOI: 10.1016/j.neucom.2015.07.085.

[194] F. Riemenschneider. "ULPMark goes peripheral". In: *Design & Elektronik* 8 (2017). URL: https://www.elektroniknet.de/design-elektronik/halbleiter/ulpmark-jetzt-mit-peripheriefunktionen-145796.html.

[195] B. A. S. del Rio, T. Lopetegi, and I. Romero. "Assessment of different methods to estimate electrocardiogram signal quality". In: *Computing in Cardiology*. 2011, pp. 609–612.

[196] R. A. Robergs, D. Dwyer, and T. Astorino. "Recommendations for improved data processing from expired gas analysis indirect calorimetry". In: *Sports medicine (Auckland, N.Z.)* 40.2 (2010), pp. 95–111. DOI: 10.2165/11319670-000000000-00000.

[197] M. Robnik-Šikonja and I. Kononenko. "Theoretical and Empirical Analysis of ReliefF and RReliefF". In: *Machine Learning* 53.1/2 (2003), pp. 23–69. DOI: 10.1023/A:1025667309714.

[198] K. Romer and F. Mattern. "The design space of wireless sensor networks". In: *Wireless Communications, IEEE* 11.6 (Dec. 2004), pp. 54–61. DOI: 10.1109/MWC.2004.1368897.

[199] S. Rosati, G. Balestra, and M. Knaflitz. "Comparison of Different Sets of Features for Human Activity Recognition by Wearable Sensors". In: *Sensors (Basel, Switzerland)* 18.12 (2018). DOI: 10.3390/s18124189.

[200] J. Rowley. "The wisdom hierarchy: representations of the DIKW hierarchy". In: *Journal of Information Science* 33.2 (2007), pp. 163–180. DOI: 10.1177/0165551506070706.

[201] A. S. Aghaei, B. Donmez, C. C. Liu, D. He, G. Liu, K. N. Plataniotis, H.-Y. W. Chen, and Z. Sojoudi. "Smart Driver Monitoring: When Signal Processing Meets Human Factors: In the driver's seat". In: *IEEE Signal Processing Magazine* 33.6 (2016), pp. 35–48. DOI: 10.1109/MSP.2016.2602379.

[202] M. Sadrawi, C.-H. Lin, Y.-T. Lin, Y. Hsieh, C.-C. Kuo, J. C. Chien, K. Haraikawa, M. F. Abbod, and J.-S. Shieh. "Arrhythmia Evaluation in Wearable ECG Devices". In: *Sensors (Basel, Switzerland)* 17.11 (2017). DOI: 10.3390/s17112445.

[203] L. Salahuddin, J. Cho, M. G. Jeong, and D. Kim. "Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings". In: *Proceedings of IEEE Engineering in Medicine and Biology Society* 2007 (2007), pp. 4656–4659. DOI: 10.1109/IEMBS.2007.4353378.

[204] S. Sarkar, D. Ritscher, and R. Mehra. "A detector for a chronic implantable atrial tachyarrhythmia monitor". In: *IEEE Transactions on Biomedical Engineering* 55.3 (2008), pp. 1219–1224. DOI: 10.1109/TBME.2007.903707.

[205] E. Sazonov and M. R. Neuman, eds. *Wearable sensors: Fundamentals, implementation and applications*. Elsevier, 2014. DOI: 10.1016/c2013-0-06896-x.

[206] S. Scheurer, S. Tedesco, K. N. Brown, and B. O'Flynn. "Human activity recognition for emergency first responders via body-worn inertial sensors". In: *IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)* (2017), pp. 5–8. DOI: 10.1109/BSN.2017.7935994.

[207] 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 2018 on International Conference on Multimodal Interaction - ICMI '18*. Ed. by S. K. D'Mello, P. P. Georgiou, S. Scherer, E. M. Provost, M. Soleymani, and M. Worsley. ACM Press, 2018, pp. 400–408. DOI: 10.1145/3242969.3242985.

[208] Y. Schutz, R. L. Weinsier, and G. R. Hunter. "Assessment of free-living physical activity in humans: an overview of currently available and proposed new measures". In: *Obesity research* 9.6 (2001), pp. 368–379. DOI: 10.1038/oby.2001.48.

[209] T. M. Seeberg, A.-S. B. Vardøy, M. M. V. Taklo, and H. O. Austad. "Decision support for subjects exposed to heat stress". In: *IEEE journal of biomedical and health informatics* 17.2 (2013), pp. 402–410. DOI: 10.1109/JBHI.2013.2245141.

[210] H. Selye. "A Syndrome produced by Diverse Nocuous Agents". In: *Nature* 138.3479 (1936), p. 32. DOI: 10.1038/138032a0.

[211] H. Selye. "Stress without Distress". In: *Psychopathology of Human Adaptation*. Springer US, 1976, pp. 137–146. DOI: 10.1007/978-1-4684-2238-2_9.

[212] S. Seneviratne, Y. Hu, T. Nguyen, G. Lan, S. Khalifa, K. Thilakarathna, M. Hassan, and A. Seneviratne. "A Survey of Wearable Devices and Challenges". In: *IEEE Communications Surveys & Tutorials* 19.4 (2017), pp. 2573–2620. DOI: 10.1109/COMST.2017.2731979.

[213] D. K. Shaeffer. "MEMS inertial sensors: A tutorial overview". In: *IEEE Communications Magazine* 51.4 (2013), pp. 100–109. DOI: 10.1109/MCOM.2013.6495768.

[214] F. Shaffer and J. P. Ginsberg. "An Overview of Heart Rate Variability Metrics and Norms". In: *Frontiers in public health* 5 (2017), p. 258. DOI: 10.3389/fpubh.2017.00258.

[215] N. Sharma and T. Gedeon. "Objective measures, sensors and computational techniques for stress recognition and classification: a survey". In: *Computer methods and programs in biomedicine* 108.3 (2012), pp. 1287–1301. DOI: 10.1016/j.cmpb.2012.07.003.

[216] S. J. Sherman, W. K. Tsang, T. A. Core, and D. E. Quinn. "A low cost monolithic accelerometer". In: *1992 Symposium on VLSI Circuits Digest of Technical Papers*. IEEE, 1992, pp. 34–35. DOI: 10.1109/VLSIC.1992.229249.

[217] D. Singh, K. Vinod, and S. C. Saxena. "Sampling frequency of the RR interval time series for spectral analysis of heart rate variability". In: *Journal of Medical Engineering & Technology* 28.6 (Jan. 2004), pp. 263–272. DOI: 10.1080/03091900410001662350.

[218] E. Smets, W. de Raedt, and C. van Hoof. "Into the Wild: The Challenges of Physiological Stress Detection in Laboratory and Ambulatory Settings". In: *IEEE journal of biomedical and health informatics* 23.2 (2019), pp. 463–473. DOI: 10.1109/JBHI.2018.2883751.

[219] D. L. Smith, J. M. Haller, B. A. Dolezal, C. B. Cooper, and P. C. Fehling. "Evaluation of a wearable physiological status monitor during simulated fire fighting activities". In: *Journal of occupational and environmental hygiene* 11.7 (2014), pp. 427–433. DOI: 10.1080/15459624.2013.875184.

[220] G. B. Spurr, A. M. Prentice, P. R. Murgatroyd, G. R. Goldberg, J. C. Reina, and N. T. Christman. “Energy expenditure from minute-by-minute heart-rate recording: comparison with indirect calorimetry”. In: *The American journal of clinical nutrition* 48.3 (1988), pp. 552–559. DOI: 10.1093/ajcn/48.3.552.

[221] R. G. Steadman. “The Assessment of Sultriness. Part I: A Temperature-Humidity Index Based on Human Physiology and Clothing Science”. In: *Journal of Applied Meteorology* 18.7 (1979), pp. 861–873. DOI: 10.1175/1520-0450(1979)018<0861:TAOSPI>2.0.CO;2.

[222] R. A. H. Stewart, J. Benatar, and R. Maddison. “Living longer by sitting less and moving more”. In: *Current opinion in cardiology* 30.5 (2015), pp. 551–557. DOI: 10.1097/HCO.0000000000000207.

[223] *STM32L476xx; Datasheet: DS10198*. STMicroelectronics, 2018. URL: https://www.st.com/resource/en/datasheet/stm32l476rg.pdf.

[224] R. M. Stock. “Is Boreout a Threat to Frontline Employees’ Innovative Work Behavior?” In: *Journal of Product Innovation Management* 32.4 (2015), pp. 574–592. DOI: 10.1111/jpim.12239.

[225] S. J. Strath, D. R. Bassett, A. M. Swartz, and D. L. Thompson. “Simultaneous heart rate-motion sensor technique to estimate energy expenditure”. In: *Medicine & Science in Sports & Exercise* 33.12 (Dec. 2001), pp. 2118–2123. DOI: 10.1097/00005768-200112000-00022.

[226] N. Strodthoff and C. Strodthoff. “Detecting and interpreting myocardial infarction using fully convolutional neural networks”. In: *Physiological measurement* 40.1 (2019), p. 015001. DOI: 10.1088/1361-6579/aaf34d.

[227] J. R. Stroop. “Studies of interference in serial verbal reactions”. In: *Journal of Experimental Psychology* 18.6 (1935), pp. 643–662. DOI: 10.1037/h0054651.

[228] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss. “Activity-Aware Mental Stress Detection Using Physiological Sensors”. In: *Mobile Computing, Applications, and Services*. Ed. by M. Gris and G. Yang. Vol. 76. Springer Berlin Heidelberg, 2012, pp. 211–230. DOI: 10.1007/978-3-642-29336-8_12.

[229] F.-T. Sun, C. Kuo, and M. Griss. “PEAR: Power Efficiency through Activity Recognition (for ECG-based Sensing)”. In: *Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare*. Ed. by J. O’Donoghue, G. O’Hare, and P. McCullagh. IEEE, 2011. DOI: 10.4108/icst.pervasivehealth.2011.246037.

[230] J. Sweller. “Cognitive Load During Problem Solving: Effects on Learning”. In: *Cognitive Science* 12.2 (1988), pp. 257–285. DOI: 10.1207/s15516709cog1202_4.

[231] J. Sweller, J. J. G. van Merrienboer, and F. G. W. C. Paas. “Cognitive Architecture and Instructional Design”. In: *Educational Psychology Review* 10.3 (1998), pp. 251–296. DOI: 10.1023/a:1022193728205.

[232] A. Tal. *Two Flash Technologies Compared: NOR vs NAND*. 91-SR-012-04-8L. Ed. by M-Systems Flash Disk Pioneers. M-Systems, 2002. URL: https://www.csd.uoc.gr/~hy428/reading/M-Systems_NANDvsNOR.pdf.

[233] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen. “An advanced detrending method with application to HRV analysis”. In: *IEEE Transactions on Biomedical Engineering* 49.2 (2002), pp. 172–175. DOI: 10.1109/10.979357.

[234] M. Teresa, L. Sileo, and M. De. “Magnetic Field Sensors Based on Microelectromechanical Systems (MEMS) Technology”. In: *Magnetic Sensors - Principles and Applications*. Ed. by K. Kuang. InTech, 2012. DOI: 10.5772/36468.

[235] N. V. Thakor, J. G. Webster, and W. J. Tompkins. “Estimation of QRS complex power spectra for design of a QRS filter”. In: *IEEE Transactions on Biomedical Engineering* 31.11 (1984), pp. 702–706. DOI: 10.1109/TBME.1984.325393.

[236] C. Thurber, L. R. Dugas, C. Ocobock, B. Carlson, J. R. Speakman, and H. Pontzer. “Extreme events reveal an alimentary limit on sustained maximal human energy expenditure”. In: *Science Advances* 5.6 (June 2019). DOI: 10.1126/sciadv.aaw0341.

[237] M. S. Tremblay, R. C. Colley, T. J. Saunders, G. N. Healy, and N. Owen. “Physiological and health implications of a sedentary lifestyle”. In: *Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme* 35.6 (2010), pp. 725–740. DOI: 10.1139/H10-079.

[238] C. Tudor-Locke and D. R. Bassett. “How Many Steps/Day Are Enough?” In: *Sports Medicine* 34.1 (2004), pp. 1–8. DOI: 10.2165/00007256-200434010-00001.

[239] M. D. Tundo, E. Lemaire, and N. Baddour. “Correcting Smartphone orientation for accelerometer-based analysis”. In: *2013 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, 2013, pp. 58–62. DOI: 10.1109/memea.2013.6549706.

[240] U.S. Department of Health and Human Services, ed. *Physical Activity Guidelines for Americans: 2nd edition*. 2018. URL: https://health.gov/paguidelines/second-edition/pdf/Physical_Activity_Guidelines_2nd_edition.pdf.

[241] J. B. de V. Weir. “New methods for calculating metabolic rate with special reference to protein metabolism”. In: *The Journal of Physiology* 109.1-2 (Aug. 1949). 15394301, pp. 1–9. DOI: 10.1113/jphysiol.1949.sp004363.

[242] K. L. Venkatachalam, J. E. Herbrandson, and S. J. Asirvatham. “Signals and signal processing for the electrophysiologist: part I: electrogram acquisition”. In: *Circulation. Arrhythmia and electrophysiology* 4.6 (2011), pp. 965–973. DOI: 10.1161/circep.111.964304.

[243] M. Vollmer. “A robust, simple and reliable measure of heart rate variability using relative RR intervals”. In: *Computing in Cardiology Conference (CinC)*. IEEE, 2015, pp. 609–612. DOI: 10.1109/CIC.2015.7410984.

[244] *W25N01GWxxIG/IT; Datasheet: 1.8V 1G-BIT SERIAL SLC NAND FLASH MEMORY WITH DUAL/QUAD SPI*. Winbond, 2018. URL: https://www.winbond.com/resource-files/w25n01gw%20revf%20082418.pdf.

[245] J. M. Wakeling, B. M. Nigg, and A. I. Rozitis. “Muscle activity damps the soft tissue resonance that occurs in response to pulsed and continuous vibrations”. In: *Journal of applied physiology (Bethesda, Md.: 1985)* 93.3 (2002), pp. 1093–1103. DOI: 10.1152/japplphysiol.00142.2002.

[246] L. Wang, S. W. Su, B. G. Celler, and E. Ambikairajah. “Analysis of Orientation Error of Triaxial Accelerometers on the Assessment of Energy Expenditure”. In: *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, 2005, pp. 3514–3517. DOI: 10.1109/IEMBS.2005.1617237.

[247] D. E. R. Warburton, C. W. Nicol, and S. S. D. Bredin. “Health benefits of physical activity: the evidence”. In: *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne* 174.6 (2006), pp. 801–809. DOI: 10.1503/cmaj.051351.

[248] B. Warneke, M. Last, B. Liebowitz, and K. S. J. Pister. “Smart Dust: communicating with a cubic-millimeter computer”. In: *Computer* 34.1 (Jan. 2001), pp. 44–51. DOI: 10.1109/2.895117.

[249] W. Weber, J. M. Rabaey, and E. Aarts. *Ambient Intelligence*. Springer Berlin Heidelberg, 2005. DOI: 10.1007/b138670.

[250] M. Weippert, M. Kumar, S. Kreuzfeld, D. Arndt, A. Rieger, and R. Stoll. “Comparison of three mobile devices for measuring R-R intervals and heart rate variability: Polar S810i, Suunto t6 and an ambulatory ECG system”. In: *European journal of applied physiology* 109.4 (2010), pp. 779–786. DOI: 10.1007/s00421-010-1415-9.

[251] M. Weiser. “The computer for the 21 st century”. In: *Scientific American* (1991). URL: https://www.ics.uci.edu/~corps/phaseii/Weiser-Computer21stCentury-SciAm.pdf.

[252] E. P. Widmaier, H. Raff, K. T. Strang, and A. J. Vander. *Vander's human physiology: The mechanisms of body function*. Thirteenth edition. McGraw-Hill, 2014. ISBN: 978-0-07-337830-5.

[253] J. Wijsman, B. Grundlehner, H. Liu, H. Hermens, and J. Penders. “Towards mental stress detection using wearable physiological sensors”. In: *Proceedings of IEEE Engineering in Medicine and Biology Society*. Vol. 2011. 2011, pp. 1798–1801. DOI: 10.1109/IEMBS.2011.6090512.

[254] P. Wilhelm. “IT-gestützte Leistungsmessung im Sport: System und Anwendung”. Dissertation. Bielefeld University, 2011.

[255] J. W. Williamson. “Autonomic responses to exercise: where is central command?” In: *Autonomic neuroscience: basic & clinical* 188 (2015), pp. 3–4. DOI: 10.1016/j.autneu.2014.10.011.

[256] S. Windisch, W. Seiberl, D. Hahn, and A. Schwirtz. “Physiological Responses to Firefighting in Extreme Temperatures Do Not Compare to Firefighting in Temperate Conditions”. In: *Frontiers in physiology* 8 (2017), p. 619. DOI: 10.3389/fphys.2017.00619.

[257] K. D. Wise. “Integrated microinstrumentation systems: smart peripherals for distributed sensing and control”. In: *1993 IEEE International Solid-State Circuits Conference Digest of Technical Papers*. IEEE, 1993, pp. 126–127. DOI: 10.1109/ISSCC.1993.280056.

[258] D. H. Wolpert. “The Lack of A Priori Distinctions Between Learning Algorithms”. In: *Neural Computation* 8.7 (1996), pp. 1341–1390. DOI: 10.1162/neco.1996.8.7.1341.

[259] S. P. Won and F. Golnaraghi. “A Triaxial Accelerometer Calibration Method Using a Mathematical Model”. In: *IEEE Transactions on Instrumentation and Measurement* 59.8 (Aug. 2010), pp. 2144–2153. DOI: 10.1109/tim.2009.2031849.

[260] Yamax. *YAMAX - About Us @ONLINE*. 2018. URL: http://www.yamax-yamasa.com/aboutus/ (visited on 08/24/2018).

[261] S.-H. Yang. “Hardware Design for WSNs”. In: *Wireless Sensor Networks*. Ed. by S.-H. Yang. Signals and Communication Technology. Springer London, 2014, pp. 49–72. DOI: 10.1007/978-1-4471-5505-8_3.

[262] K. Y. Yazdandoost and K. Sayrafian-Pour. *Channel Model for Body Area Network (BAN): IEEE P802.15-08-0780-12-0006*. IEEE P802.15 Working Group for Wireless Personal Area Networks, 2010. URL: https://mentor.ieee.org/802.15/dcn/08/15-08-0780-12-0006-tg6-channel-model.pdf.

[263] R. M. Yerkes and J. D. Dodson. “The relation of strength of stimulus to rapidity of habit-formation”. In: *Journal of Comparative Neurology and Psychology* 18.5 (1908), pp. 459–482. DOI: 10.1002/cne.920180503.

[264] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock. “State of science: mental workload in ergonomics”. In: *Ergonomics* 58.1 (2015), pp. 1–17. DOI: 10.1080/00140139.2014.956151.

[265] A. Yurtman, B. Barshan, and B. Fidan. “Activity Recognition Invariant to Wearable Sensor Unit Orientation Using Differential Rotational Transformations Represented by Quaternions”. In: *Sensors (Basel, Switzerland)* 18.8 (2018). DOI: 10.3390/s18082725.

[266] R. Zangróniz, A. Martínez-Rodrigo, J. M. Pastor, M. T. López, and A. Fernández-Caballero. “Electrodermal Activity Sensor for Classification of Calm/Distress Condition”. In: *Sensors (Basel, Switzerland)* 17.10 (2017). DOI: 10.3390/s17102324.

[267] Y. Zhang, N. Suda, L. Lai, and V. Chandra. *Hello Edge: Keyword Spotting on Microcontrollers*. Nov. 20, 2017. URL: http://arxiv.org/pdf/1711.07128v3.

Author’s publications

[268] M. Adams, M. Hesse, T. Hörmann, and U. Rückert. “Visuelles Sensorsystem für die Trainings- und Spielunterstützung im Leistungshandball”. In: *Technologien im Leistungssport 3. Tagungsband zur 19. Frühjahrsschule am 14./15. Mai 2018 in Leipzig*. Ed. by I. Fichtner. Vol. 13. Meyer & Meyer Verlag, 2018, pp. 106–115. ISBN: 978-3-8403-7628-3.

[269] M. Hesse, P. Christ, T. Hörmann, and U. Rückert. “A Respiration Sensor for a Chest-Strap Based Wireless Body Sensor”. In: *SENSORS, 2014 IEEE*. 2014, pp. 490–493. DOI: 10.1109/ICSENS.2014.6985042.

[270] M. Hesse, P. Christ, T. Hörmann, M. Adams, and U. Rückert. “Die Entwicklung zukünftiger körpernaher Sensorsysteme für die autarke und mobile Trainingsunterstützung”. In: *Technologien im Leistungssport*. Ed. by I. Fichtner. Vol. 4. Schriftenreihe für angewandte Trainingswissenschaft. Meyer & Meyer, 2016, pp. 152–161. ISBN: 978-3-89899-949-6.

[271] M. Hesse, M. Adams, T. Hörmann, and U. Rückert. “Towards a Comprehensive Power Consumption Model for Wireless Sensor Nodes”. In: *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 2016, pp. 390–395. DOI: 10.1109/BSN.2016.7516293.

[272] T. Hörmann, M. Hesse, M. Adams, and U. Rückert. “A Software Assistant for User-Centric Calibration of a Wireless Body Sensor”. In: *2016 IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. 2016, pp. 183–188. DOI: 10.1109/BSN.2016.7516256.

[273] T. Hörmann, M. Hesse, P. Christ, M. Adams, C. Menßen, and U. Rückert. “Detailed Estimation of Cognitive Workload with Reference to a Modern Working Environment”. In: *Biomedical Engineering Systems and Technologies*. Ed. by A. Fred and H. Gamboa. Springer Nature, 2017, pp. 205–223. DOI: 10.1007/978-3-319-54717-6_12.

[274] T. Hörmann, M. Hesse, P. Christ, M. Adams, C. Menßen, and U. Rückert. “Fine-Grained Prediction of Cognitive Workload in a Modern Working Environment by Utilizing Short-Term Physiological Parameters”. In: *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies*. Vol. 4. SCITEPRESS, 2016, pp. 42–51. DOI: 10.5220/0005665000420051.

[275] T. Hörmann, P. Christ, M. Hesse, and U. Rückert. “Robust estimation of physical activity by adaptively fusing multiple parameters”. In: *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2015, pp. 1–6. DOI: 10.1109/bsn.2015.7299390.

[276] T. Hörmann and U. Rückert. “Vernetzte Arbeitsumgebungen: Körpernahe und tragbare Sensorik in der Arbeitswelt”. In: *Handbuch Gestaltung digitaler und vernetzter Arbeitswelten*. Ed. by G. W. Maier, G. Engels, and E. Steffen. Springer, 2017, pp. 1–20. DOI: 10.1007/978-3-662-52903-4_9-1.

[277] C. Lian Sang, M. Hesse, S. Zehe, M. Adams, T. Hörmann, and U. Rückert. “An Adaptive Acknowledgement On-demand Protocol for Wireless Sensor Networks”. In: *Proceedings of the 6th International Conference on Sensor Networks*. Vol. 1. 2017, pp. 174–181. DOI: 10.5220/0006208501740181.

[278] C. Lian Sang, M. Adams, T. Hörmann, M. Hesse, M. Porrmann, and U. Rückert. “An Analytical Study of Time of Flight Error Estimation in Two-Way Ranging Methods”. In: *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2018. DOI: 10.1109/ipin.2018.8533697.

[279] C. Lian Sang, M. Adams, T. Hörmann, M. Hesse, M. Porrmann, and U. Rückert. “Numerical and Experimental Evaluation of Error Estimation for Two-Way Ranging Methods”. In: *Sensors* 19.3 (2019). DOI: 10.3390/s19030616.

Supervised work

[280] F. Bohr. “Entwicklung eines Atemsensors für den mobilen Einsatz”. Master project. Bielefeld University, 2017.

[281] M. Bungenstock. “Entwicklung eines minimalistischen Sensorknotens”. Bachelor thesis. Bielefeld University, 2016.

[282] R. Ewers. “Implementierung und Evaluierung eines Hybrid-Tranceivers für ein körpernahes Sensorsystem”. Bachelor thesis. Bielefeld University, 2016.

[283] R. Ewers. “UWB-Lokalisierung für den AMiRo”. Master project. Bielefeld University, 2018.

[284] M. Ghaffar and M. R. Noori. “Comparison of bio signal Compression”. Master project. Bielefeld University, 2015.

[285] F. Herrmann. “Effiziente Signalverarbeitung für den BI-Vital”. Bachelor thesis. Bielefeld University, 2018.

[286] N. Kaden. “Entwicklung und Evaluation eines Systems zur sensorgestützten Zahnreinigung”. Master thesis. Bielefeld University, 2019.

[287] T. Krahn. “Weiterentwicklung eines minimalen Sensorknotens”. Master project. Bielefeld University, 2017.

[288] S. M. Nemati. “Accuracy of DW1000 in game-sports tracking”. Master project. Bielefeld University, 2019.

[289] S. M. Nemati. “Ultra-Low Power ECG Break-Out-Board”. Master project. Bielefeld University, 2016.

[290] K. Penner. “Sensorfusion von inertialen und visuellen Sensordaten für das Tracking im Hallensport”. Master project. Bielefeld University, 2019.

[291] P. C. Schneider. “Unüberwachte Feature Extraktion mit Variational Autoencodern für die multisensorische Erkennung menschlicher Aktivitäten”. Master project. Bielefeld University, 2019.

[292] D. Spannagel. “Konzeption und Entwicklung eines graphischen Sensorfusionstools für ein Sport-Videoanalyse-Systems”. Master project. Bielefeld University, 2017.

[293] B. Steinhagen. “Multi-Protokoll Tranceiver für den BI-Vital”. Master project. Bielefeld University, 2019.

[294] B. Steinhagen. “Realisierung eines multi-modales Trackingsystems zur Spielerverfolgung im Hallensport”. Master thesis. Bielefeld University, 2019.

[295] S. Welzel. “Portierung des AMiRo-OS auf den BI-Vital”. Bachelor thesis. Bielefeld University, 2019.

A BI-Vital: schematics and layout

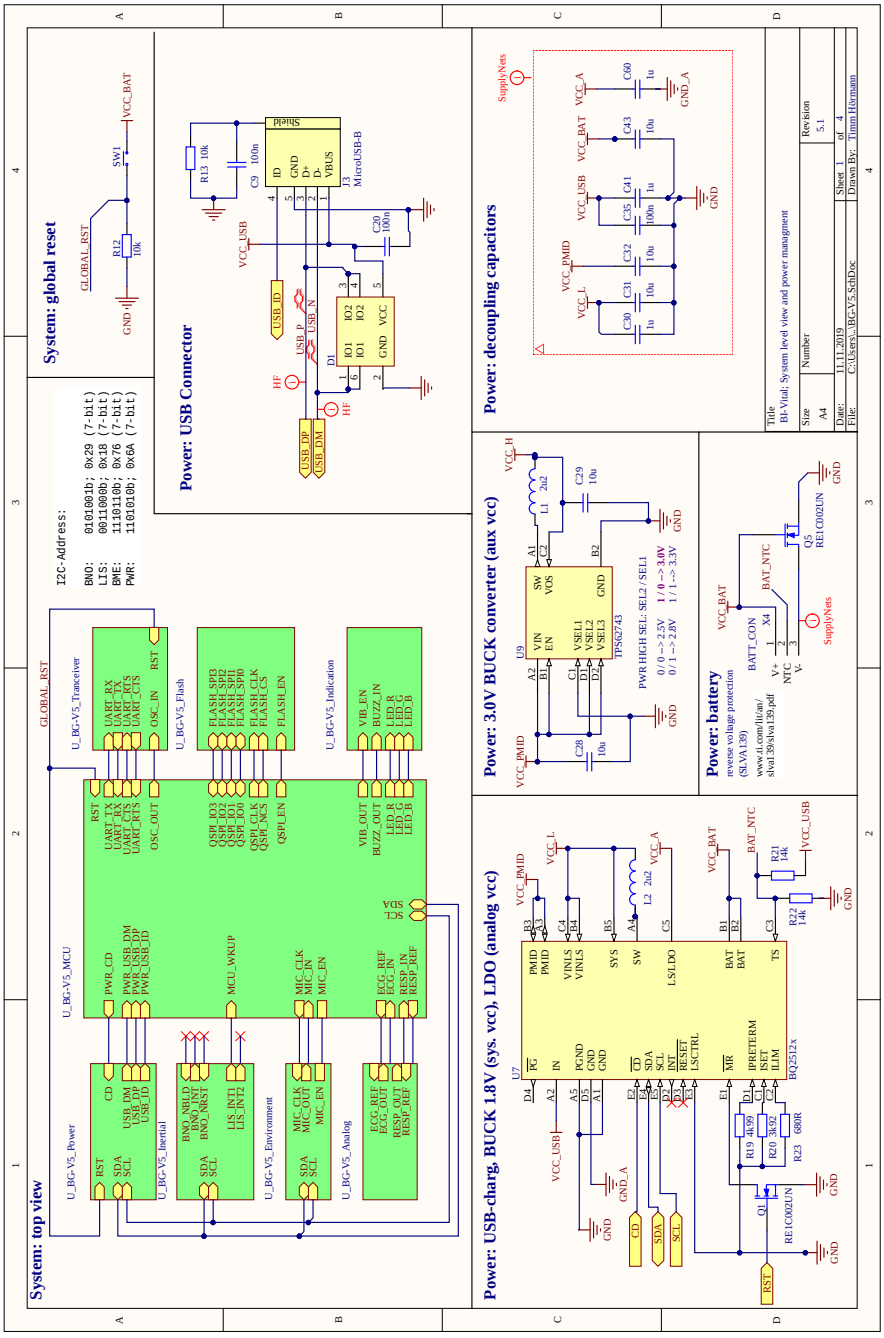


Figure A.1: Schematic, system level

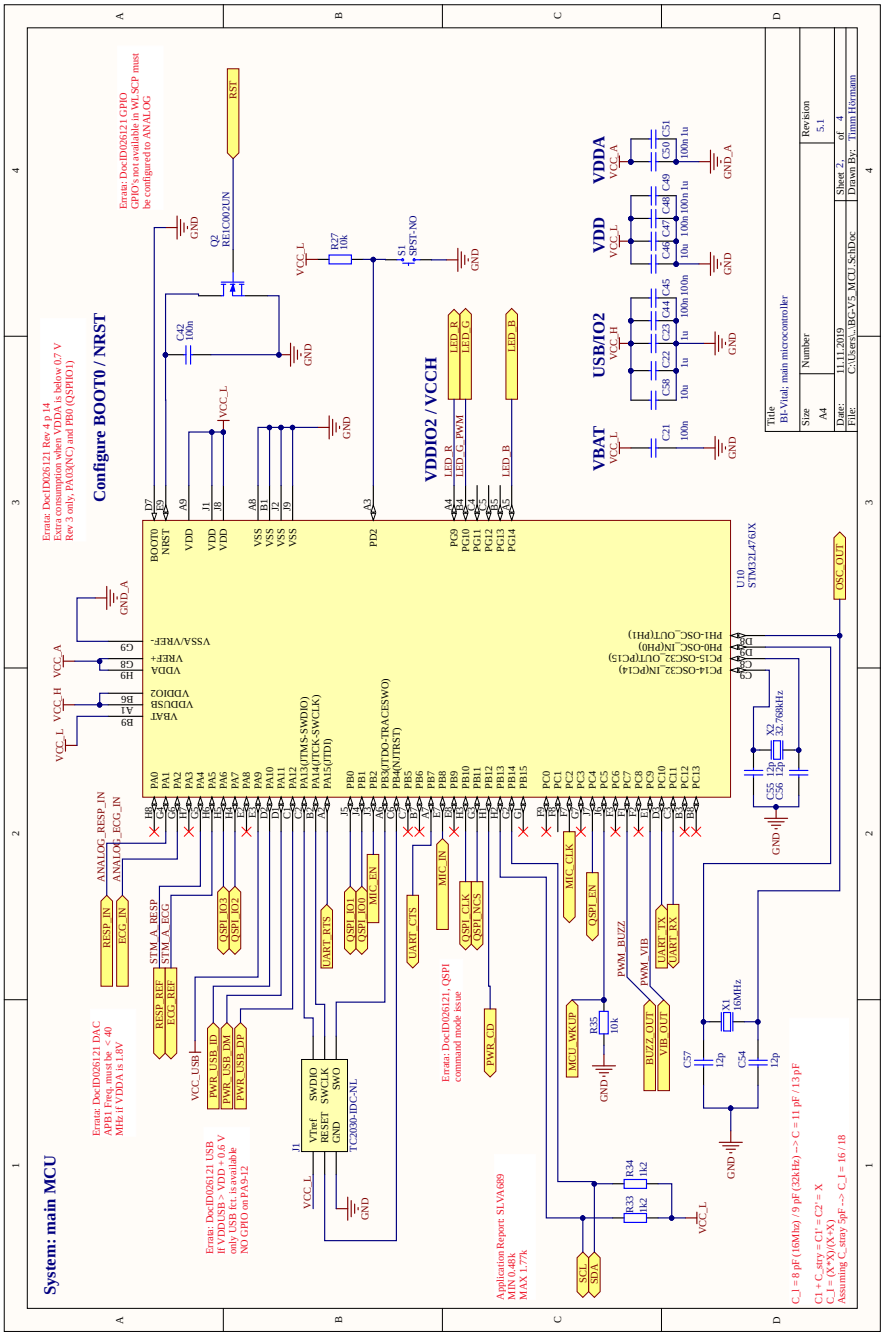


Figure A.2: Schematic, main micro controller

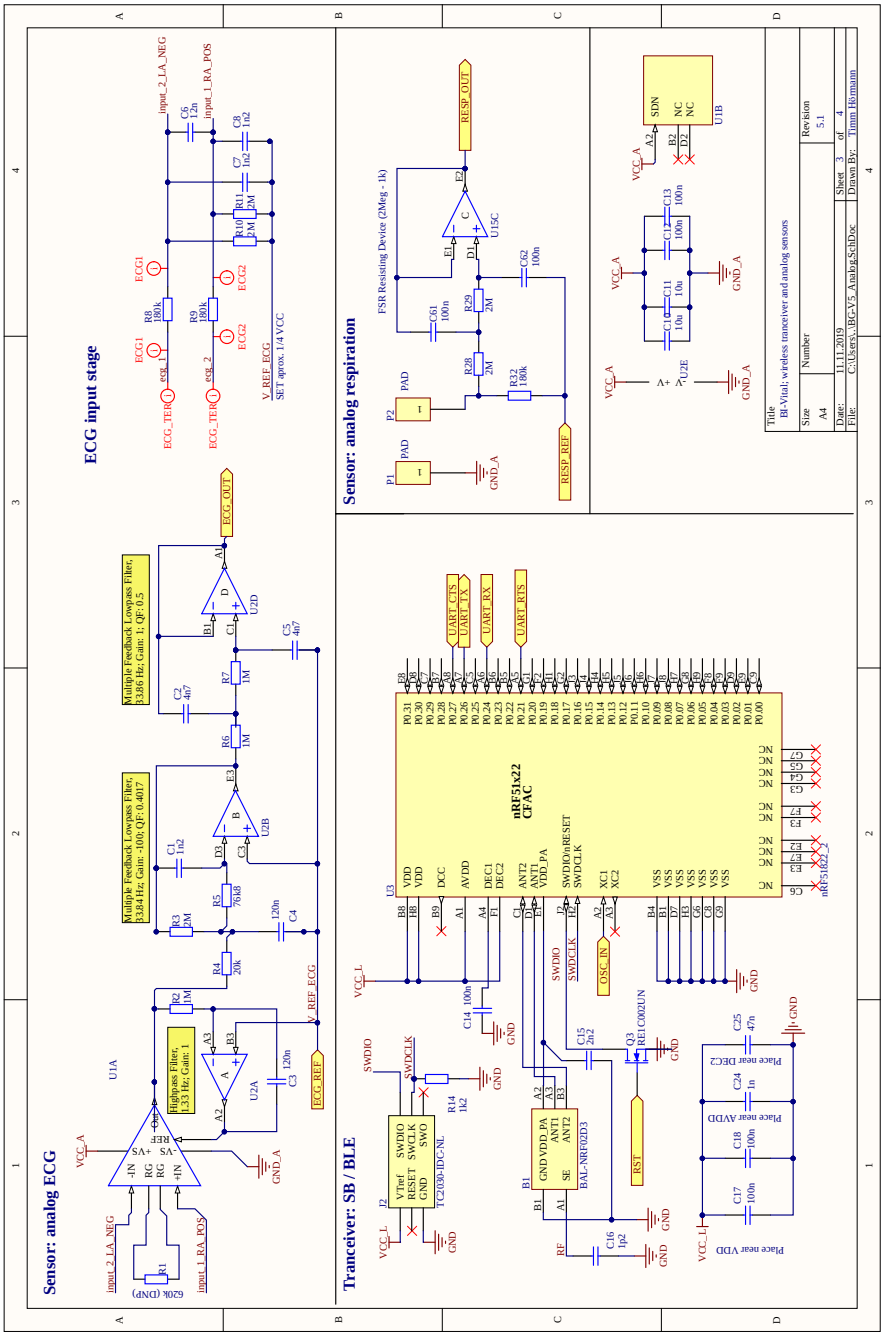


Figure A.3: Schematic, wireless transceiver and analog sensors

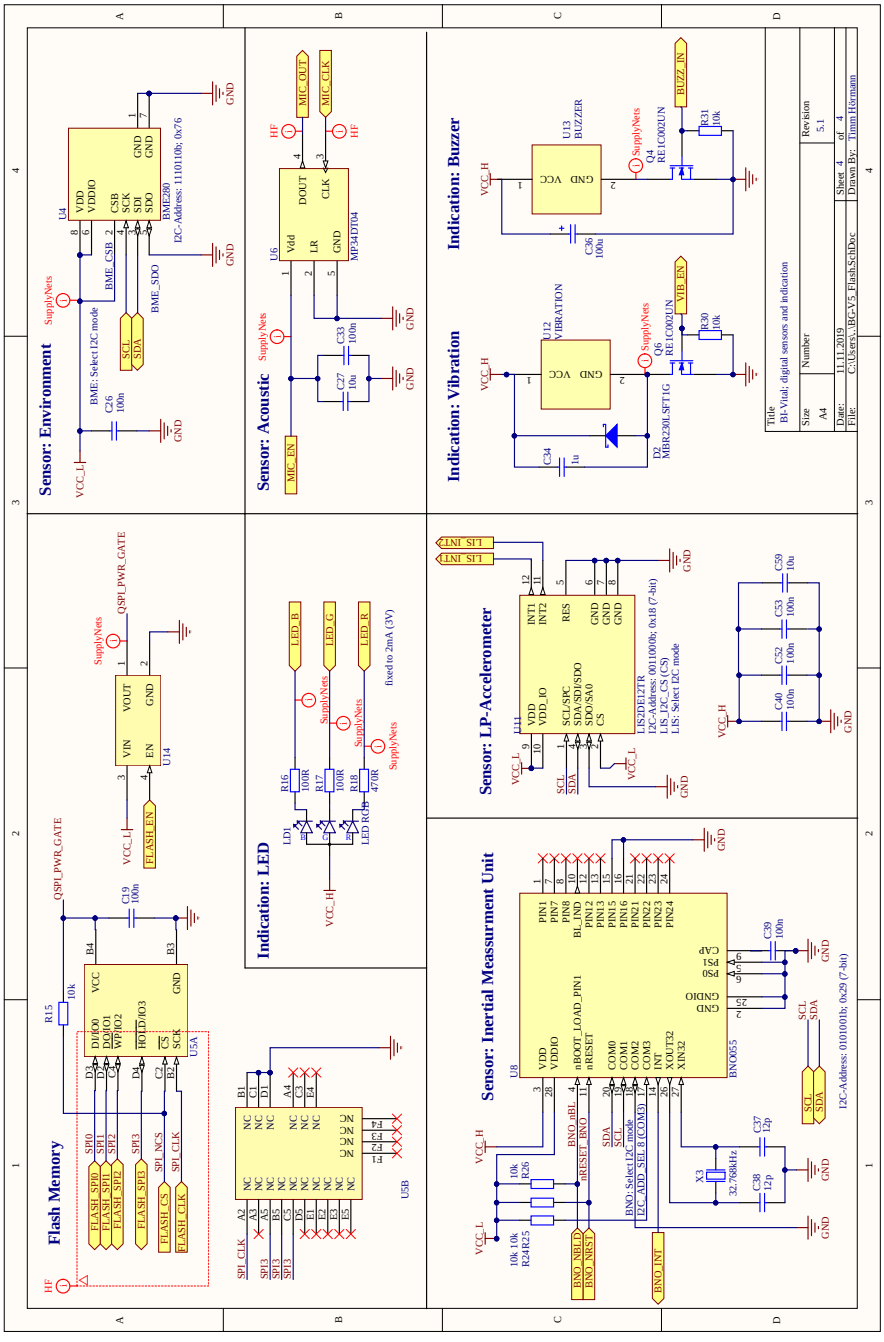


Figure A.4: Schematic, digital sensors and indication

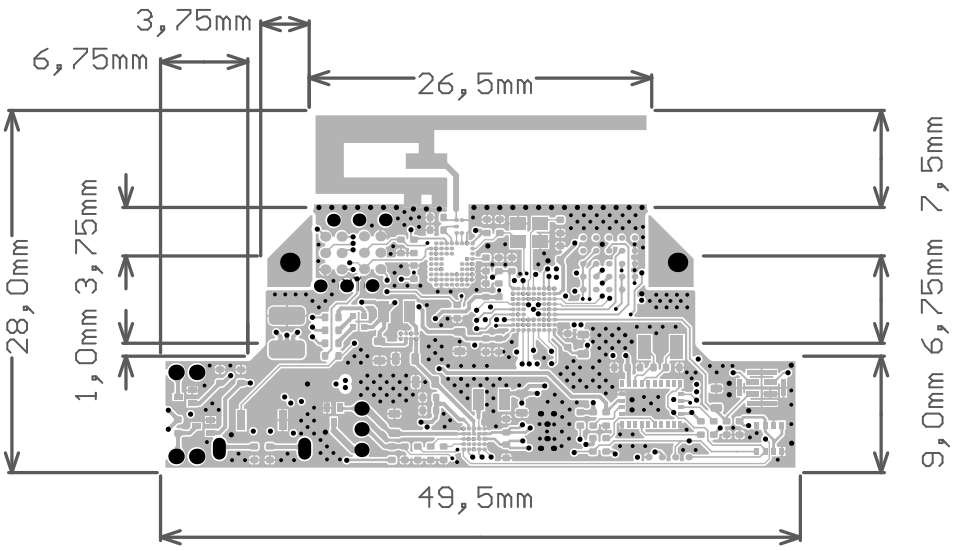


Figure A.5: Top Layer, digital components

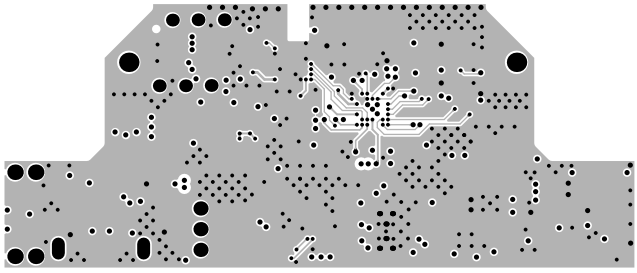


Figure A.6: Internal Layer 1, digital ground

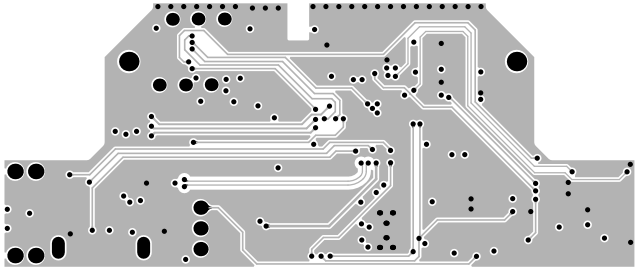


Figure A.7: Internal layer 2, digital signal

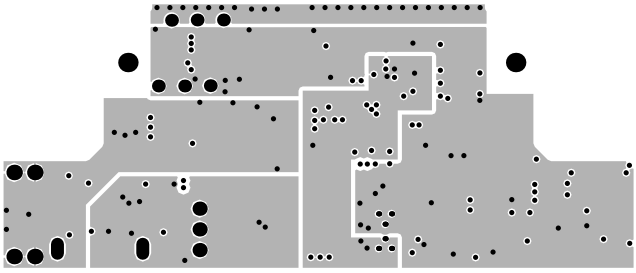


Figure A.8: Internal layer 3, supply split plane

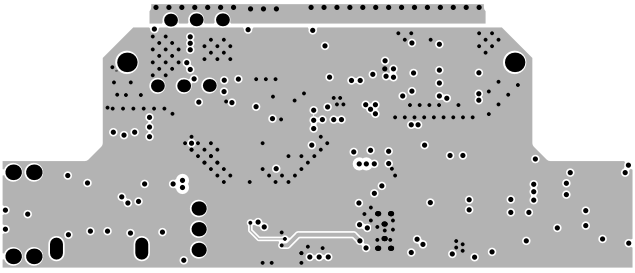


Figure A.9: Internal layer 4, analog ground

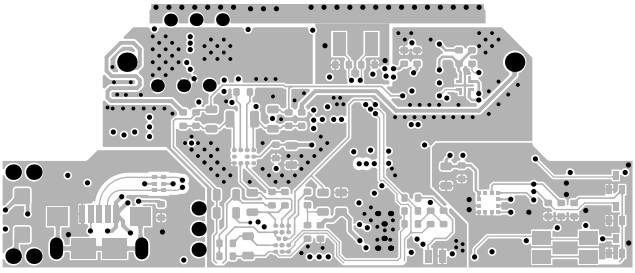


Figure A.10: Bottom layer, analog components

B Cognitive workload: detailed results

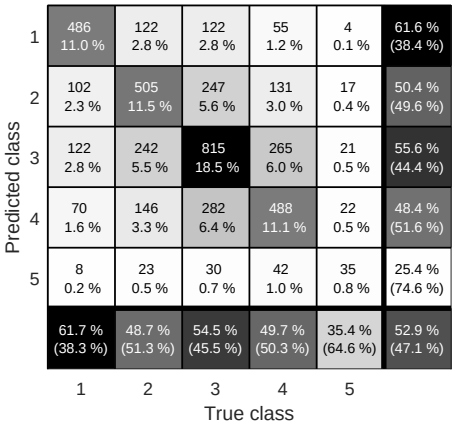


Figure B.11: CW classification, 10-fold, DT

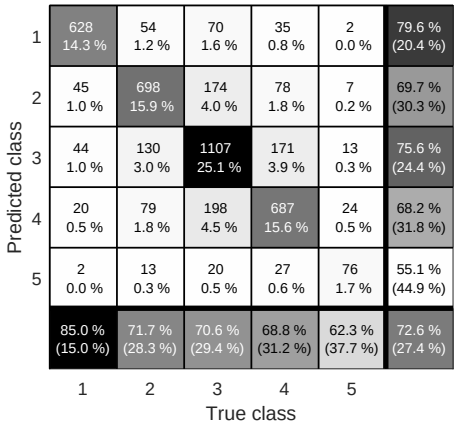


Figure B.12: CW classification, 10-fold, KNN

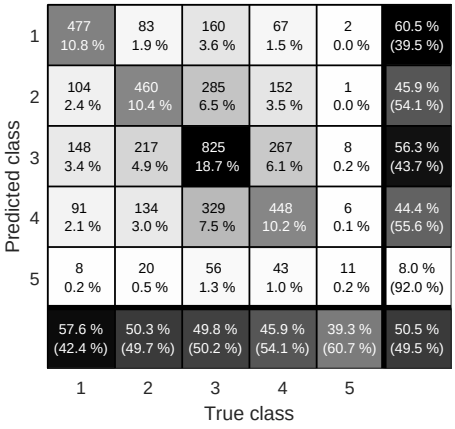


Figure B.13: CW classification, 10-fold, NB

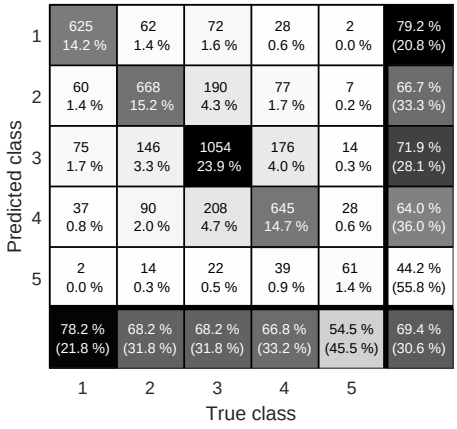


Figure B.14: CW classification, 10-fold, SVM

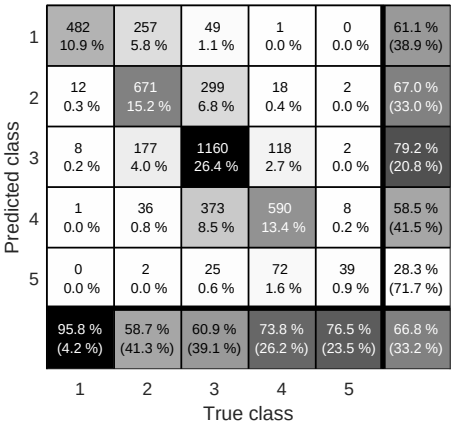


Figure B.15: CW classification, 10-fold, GP

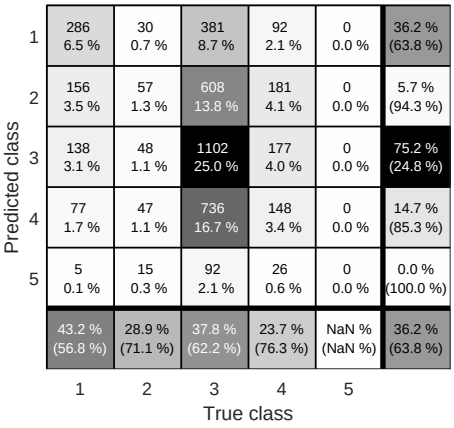


Figure B.16: CW classification, LOGO, DT

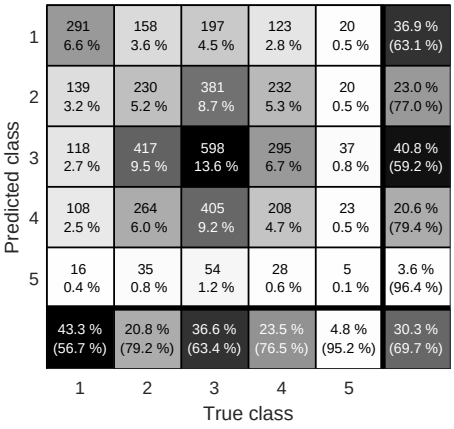


Figure B.17: CW classification, LOGO, KNN

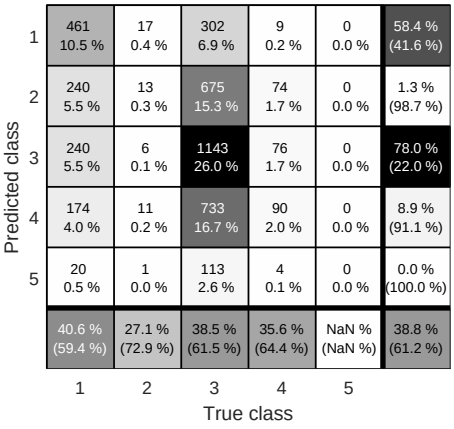


Figure B.18: CW classification, LOGO, NB

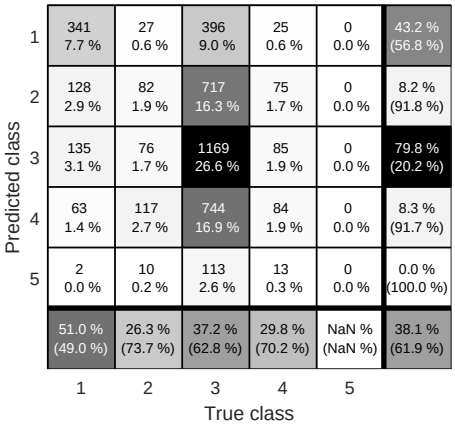


Figure B.19: CW classification, LOGO, SVM

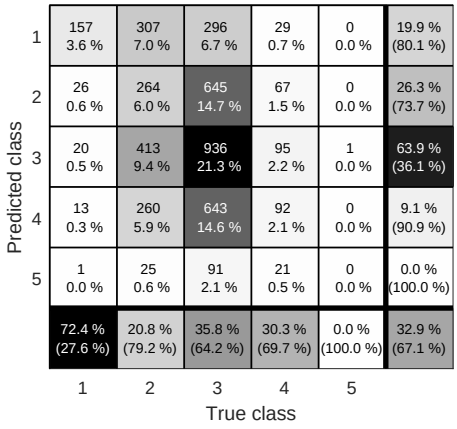


Figure B.20: CW classification, LOGO, GP

C Physical activity: detailed results

Neural network model to estimate energy expenditure as metabolic equivalents (METs):

$$\begin{bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{bmatrix}^T = \begin{bmatrix} HR_{min} \\ HR_{range} \\ HR_{slope} \\ RR_{min} \\ RR_{range} \\ RR_{slope} \\ Pitch_{mean} \\ Pitch_{range} \\ Roll_{mean} \\ Roll_{range} \\ VO_2max \end{bmatrix}^T \cdot \begin{bmatrix} 10.3215 & 8.6777 & 17.6274 & 5.7927 \\ 3.3402 & 6.5841 & 1.9801 & 2.3567 \\ 1.0859 & 1.0851 & -5.6478 & 2.1461 \\ -0.7447 & 2.0253 & 2.4175 & 17.5141 \\ -2.3898 & -1.1770 & -1.7389 & -0.5156 \\ 0.347 & 0.1439 & -0.0405 & -0.0414 \\ -4.2886 & 1.3016 & -2.6122 & -16.0202 \\ -1.6158 & 0.3662 & -2.4179 & 4.9293 \\ 0.0807 & 0.0692 & -2.9385 & 0.7067 \\ -14.4024 & 1.5491 & 2.1833 & -1.4884 \\ 13.2875 & 3.686 & 2.1061 & 1.8055 \end{bmatrix} + \begin{bmatrix} -13.5157 \\ -13.6158 \\ -3.4493 \\ -1.1061 \end{bmatrix}^T \tag{C.1}$$

$$\text{Output / MET} = \begin{bmatrix} \text{sig}(N_1) \\ \text{sig}(N_2) \\ \text{sig}(N_3) \\ \text{sig}(N_4) \end{bmatrix}^T \cdot \begin{bmatrix} 1.4069 \\ 8.7691 \\ 2.5294 \\ 3.8606 \end{bmatrix} + 1.2400 \tag{C.2}$$

$$\text{sig}(x) = \frac{1}{1+e^{-x}} \tag{C.3}$$

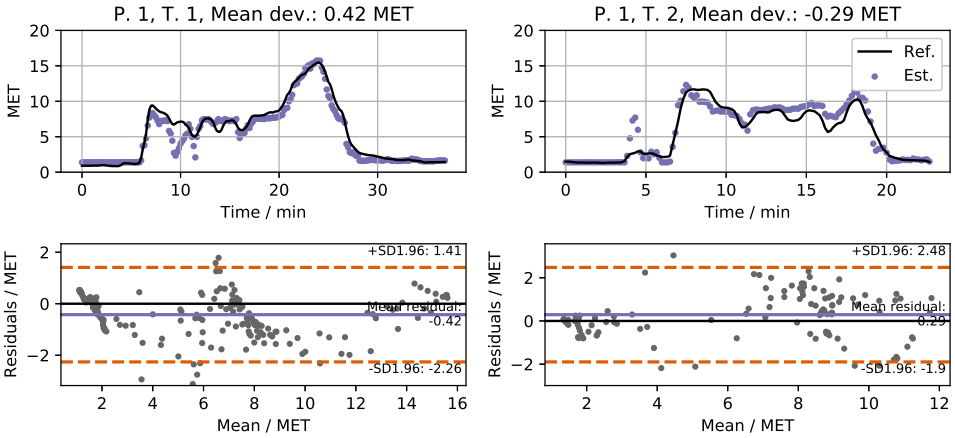


Figure C.21: PA regression, participant 1

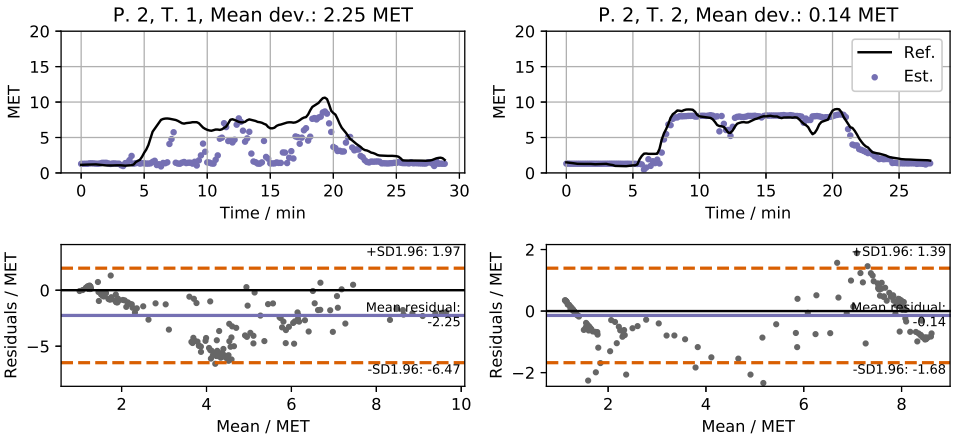


Figure C.22: PA regression, participant 2

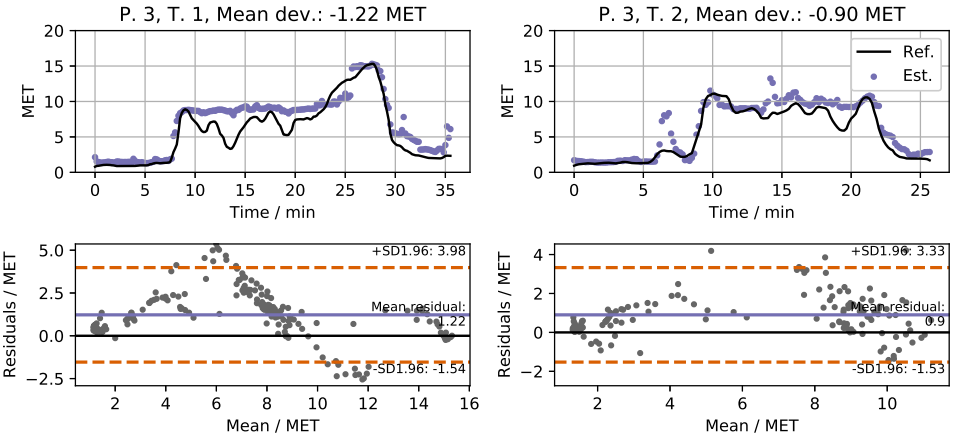


Figure C.23: PA regression, participant 3

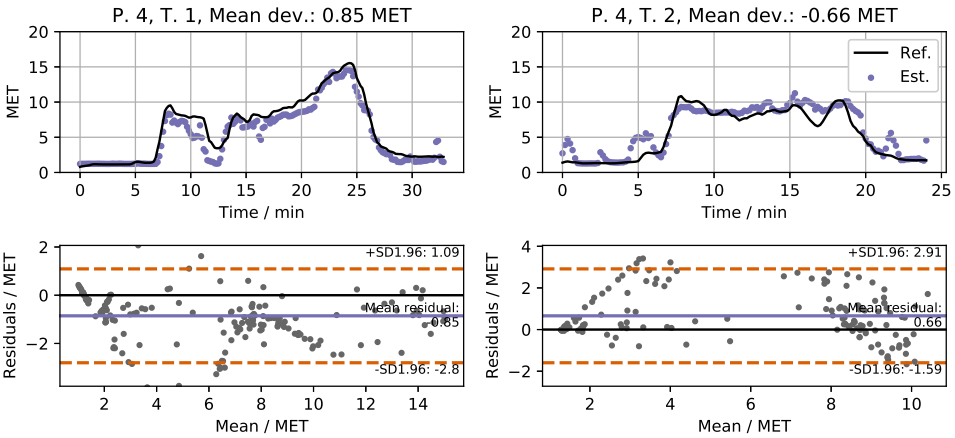


Figure C.24: PA regression, participant 4

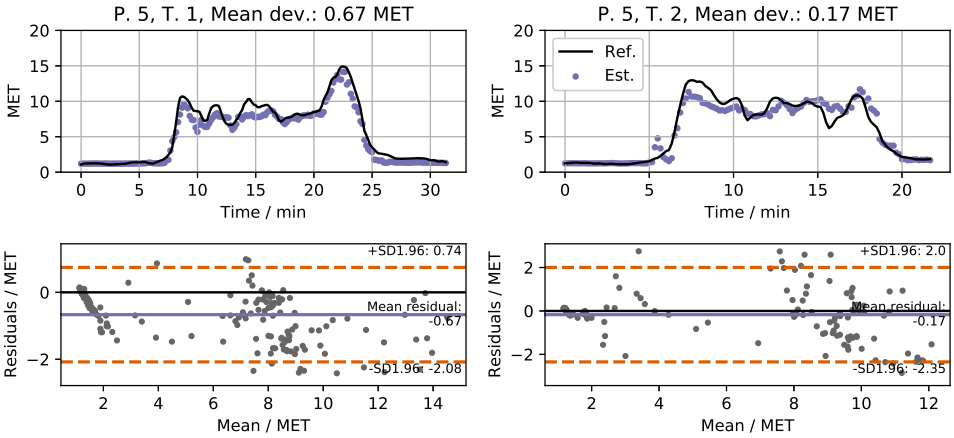


Figure C.25: PA regression, participant 5

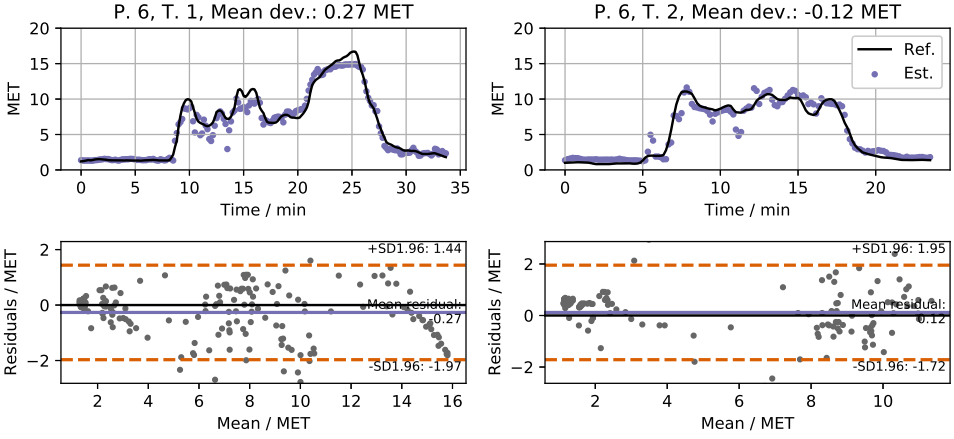


Figure C.26: PA regression, participant 6

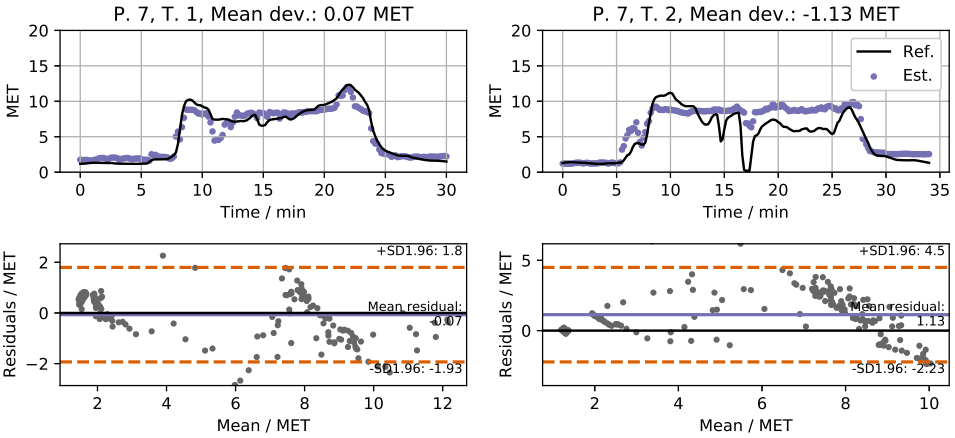


Figure C.27: PA regression, participant 7

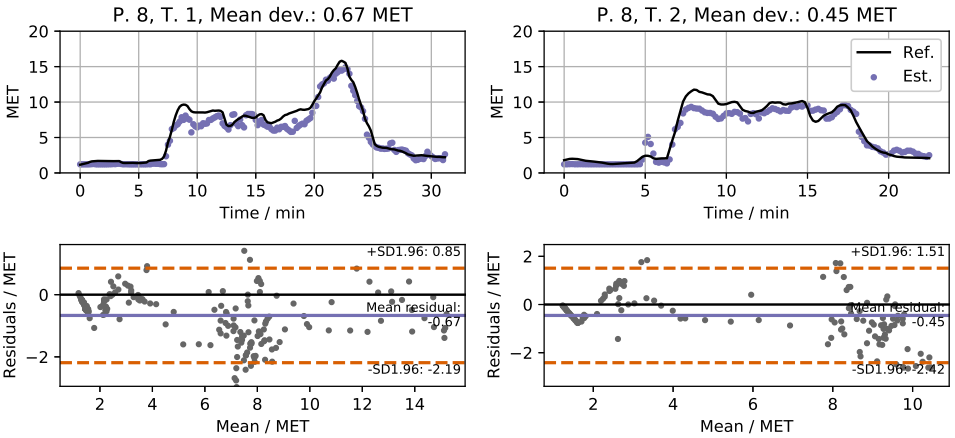


Figure C.28: PA regression, participant 8

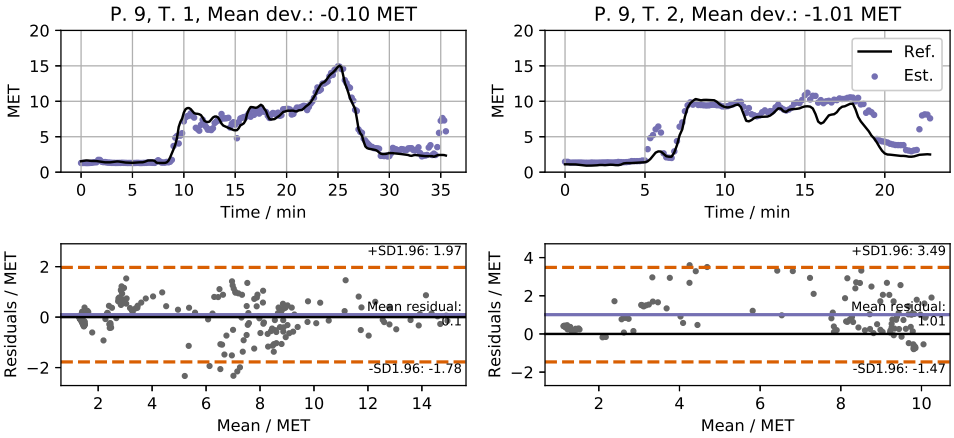


Figure C.29: PA regression, participant 9

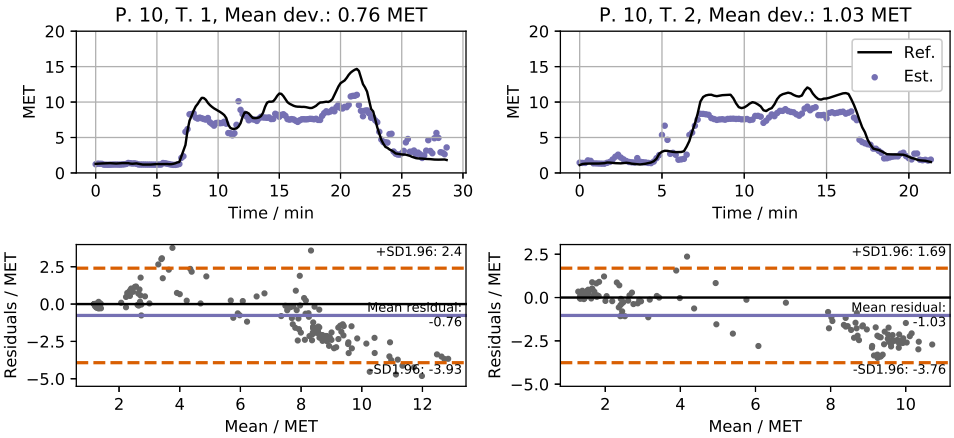


Figure C.30: PA regression, participant 10

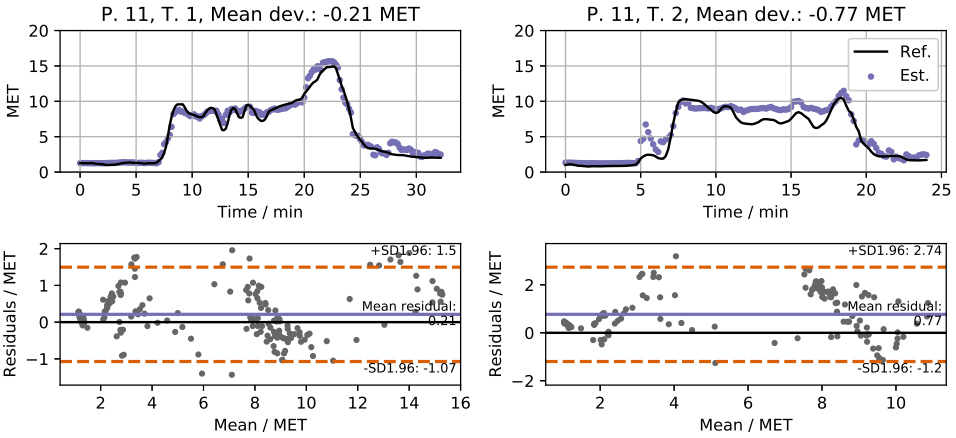


Figure C.31: PA regression, participant 11

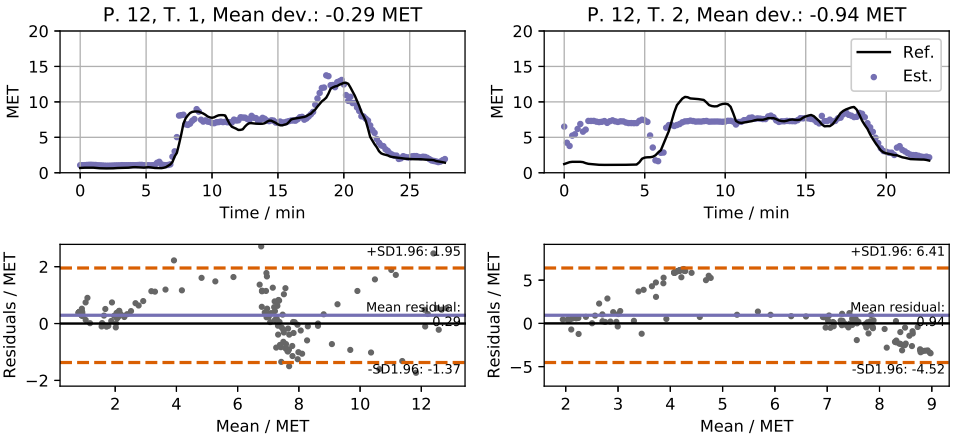


Figure C.32: PA regression, participant 12

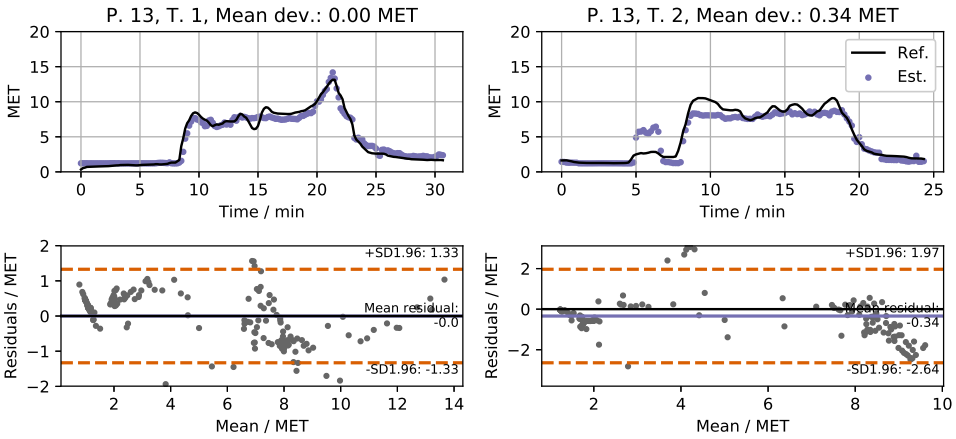


Figure C.33: PA regression, participant 13