Politecnico di Torino

Politecnico di Torino

Mechatronic Engineering
Academic Year 2023/2024
Graduation Session of October 2024

Correlation Analysis between Vigilance Tests and Body Battery

Supervisor:
Massimo Violante

Candidate:
Dario Mendolicchio

Abstract

Fatigue and sleep loss are crucial factors in everyday life, as they can deeply affect human productivity and health. Fatigue assessment is really important in certain applications: having detailed information about workers' fatigue levels may be vital in preventing possible catastrophic accidents. Monotonous activities such as highway driving may be severely affected by high levels of fatigue. For this reason, increasing the number of tools able to collect such information is really important for improving road safety and various other everyday situations. For this purpose, analyzing potential correlation between vigilance tests and body battery could provide valuable insights. In particular, vigilance tests like Psychomotor Vigilance Task (PVT) and Three Choice Vigilance Task (3CVT) are widely used in fatigue and sleep loss research fields due to their easy of use. On the other hand, Body Battery (BB) is an example of how new technologies allow for the easy collection of complex parameters that can assist in fatigue assessment. Garmin smartwatches give the opportunity to collect many vital signs useful for computing this complex parameter, providing valuable details about the body's energy levels throughout the day. However, these smartwatches cannot provide an immediate, reliable Body Battery value, as its computation requires continuous data collection during both day and night. For this reason, two vigilance tests, Psychomotor Vigilance Task (PVT) and Three Choice Vigilance Task (3CVT), were considered to explore possible correlation with it. Having a simple tool like these tests to estimate Body Battery, even if the device was not worn in the previous days, could be crucial for assessing fatigue levels before any critical activity. To this purpose, outcome metrics of both tests were evaluated. Initially, three different versions of both PVT and 3CVT (3, 5, and 10 minutes) were tested to determine which duration was more suitable for this application. Then, once the 5 minute test demonstrated to be the best trade-off between time consumed and amount of data collected, the data collection campaign was extended to 5 subjects to understand which of the two tests could achieve the best correlation with Body Battery. At this point, Psychomotor Vigilance Task was selected for further data collection, reaching the total number of 14 subjects: 7 men and 7 female. The protocol was fine tuned step by step, and in the final phase, each subject took the test twice a day (morning and evening) for 5 days. In the end, PVT showed potential as a possible estimator of the Body Battery. However, physiological research studies, such as this one on Body Battery, require a broad and heterogeneous audience to ensure full population representativeness. In this sense, extending the data collection campaign to a larger sample could increase the robustness and validity of the results obtained in this study.

Table of Contents

List of Tables

List of Figures

Acronyms

3CVT Three Choice Vigilance Task

BB Body Battery

EEG Electroencephalogram

GUI Graphical User Interface

ISI Interstimulus Interval

KSS Karolinska Sleepiness Scale

MCI Mind Cognitive Impairement

PVT Psychomotor Vigilance Task

RT Reaction Time

Chapter 1

Introduction

Fatigue is one of the main influencing factors in humans' lives. Together with sleep loss and vigilance impairment, it can severely impact health and reduce productivity. However, it does not only affect working efficiency, but it is very impacting also in everyday tasks and subjective well-being, having a very important influence in every aspect.

1.1 Background and Context

Modern technologies have introduced a wide range of tools and innovative solutions that allow to retrieve crucial information about fatigue-related fields. It is essential to continuously enlarge the number of available instruments that help assessing these specific conditions, as they can be very useful in several application such as real-driving scenarios. In these situations, vigilance impairment due to high level of fatigue or sleep loss can have catastrophic consequences. In these regard, vigilance tests, such as Psychomotor Vigilance Task (PVT) and Three Choice Vigilance Task (PVT), implement specific task to emulate monotonous situation, helping in having an objective measure of vigilance impairment. However, fatigue is further complex than this, as it is a subjective feeling affected by several factors, like sleep loss, and it cannot be simply reduced to a vigilance attention impairment. For this reason, Garmin's introduction of the Body Battery (BB) concept contributed to give a wider overview in residual energy evaluation, and, consequently, in fatigue estimation. It is a complex 24h energy monitoring parameter, computed on the base of multiple information retrieved continuously by different embedded sensors inside Garmin's smartwatches, that gives a quantitative value of the actual level of residual energy[1]. However, the Body Battery (BB) accuracy depends on the average duration for which the smartwatch is worn[1]. In absence of continuous valuable information due to long periods for which the device is not worn, its

value is reset according to a model, leading to less accurate results. This critical situation might be mitigated if this starting point could be computed according to the effective current energy level of the user. In this regard, vigilance tests could represent a quick and effective potential tool to estimate the current Body Battery (BB). Moreover, combining the data obtained from them and the BB, could help to enlarge the amount of information related to fatigue and energy level, speeding up the estimation of these conditions.

1.2 Objective of the Study

The goal of this study is to explore new solution based on the integration between vigilance tests and Garmin's Body Battery (BB). Analysing the relationship between them could represent a preliminary step to start investigating this possible integration. Finding evidence of such a relationship might open to new application for vigilance test that might have clear and powerful effect in fatigue estimation. In this research, two vigilance tests, the Psychomotor Vigilance Task (PVT) and Three Choice Vigilance Task (3CVT), were taken into account, and various Garmin smartwatches have been used to retrieve BB parameter. First, a pilot study was carried out to well define the study protocol and to understand which vigilance test could best fit for our purpose. Then, a deeper investigation was conducted on the definitive dataset collected in the final phase to examine potential monotonic correlation through a multivariate analyses.

Chapter 2

State of the Art

2.1 Fatigue and Sleep Loss

Fatigue is mainly associated to a subjective feeling which deeply affects everyday life. It is difficult to have a unique and well-established way to measure it. There exist various triggers for this debilitating condition: chronic diseases, cancer or sleep loss are just some examples[2]. For this reason several scale have been developed over years, both to detect fatigue in general or in specific application field. They can be both unidimensional, if they focus only on one aspect, or multidimensional, in case they explore different aspects to give a full overview about the subject's feeling[2]. Fatigue is also linked to sleep loss, as it can be one of its main consequences[3]. In particular, as stated by Lim and Dinges (2008), prolonged wakefulness can lead to fatigue resulting in reduced reactivity and attention impairment[3]. Moreover, sleep loss can also affect circadian rhythms and homeostatic sleep drives, increasing further fatigue levels[3]. For these reasons, dealing with both fatigue and sleep loss is always more relevant in real world scenario, leading to the continuous development of new tools able to asses them objectively.

2.1.1 The Cardiorespiratory System and Fatigue

Fatigue is a complex feeling and can be analyzed under different perspectives. The distinction between physical, mental and emotional fatigue are just some examples of this complexity. Physical fatigue is the reduction in physical endurance, it causes general tiredness and muscle weakness. On the other hand, mental fatigue is related with cognitive function, influencing decision-making ability and concentration. In the end, emotional fatigue is caused by emotional exhaustion, such as increasing sense of irritability[2]. Despite its subjective nature, it is possible to estimate it objectively using some physiological parameter. One of the most significant marker is heart rate variability (HRV), which has been shown to be a reliable indicator of

fatigue levels[4]. HRV reflects the balance between the sympathetic ("fight or flight") and the parasympathetic ("rest and digest") autonomous nervous systems, both of which play an important role in sleep quality and recovery processes[5]. In particular, low HRV provides insights in parasympathetic system activity, suggesting poor sleep quality and increased fatigue[5].

2.1.2 Karolinska Sleepiness Scale and Sleep Loss

Similar to fatigue, there are also some subjective scales to evaluate the perceived level of sleepiness. They are useful tools to quickly assess participants' feelings. In particular, the Karolinska Sleepiness Scale (KSS) is a subjective scale designed to assess the instantaneous level of perceived sleepiness. It reflects the participants' real-time feelings over the previous 10 minutes[6]. Due to its correlation with Electroencephalogram (EEG) and reaction times, KSS is widely used in research on fatigue and sleep loss[6]. Its higher sensitivity to short-term application with respect to long-term, makes it suitable for integration in shift work studies or driving safety research[6]. In general, KSS is scored from 1 to 10, with 1 reflecting "extremely alert" and 10 indicating "extremely sleepy, can't keep awake" (for more details, refer to Appendix A)[7]. However, there are different versions of KSS, according to their complexity. Åkerstedt et al. (2016) demonstrated how different configurations of these scales can yield the same results[6]. Specifically, they evaluated two versions of KSS: one simplified, which labels only odd-numbered points, to reduce the overload risk related to a higher number of option, and the standard one, which provides 10 options to choose from[6].

2.2 Psychomotor Vigilance Task

The Psychomotor Vigilance Task (PVT) consists in a sustained attention and stimulus-response task which is widely used in fatigue and sleep deprivation researches[8]. The first version of this test can be attributed to Dinges and Powell (1985) and it was a simple audio cassette recorder which had been modified to record Reaction Time (RT), which is the time elapsed between the stimulus appearance and the user reaction[9]. Newest form of PVT developed during the following decades became always more powerful, providing not only RT but also several other outcome metrics which can be really useful in fatigue studies. The PVT-192 is one of the latest version and it is considered the gold standard: it is an handheld device measuring 21cm x 11cm x 6cm with a built-in four digit LED screen which provides the visual stimulus as a stopwatch advancing, two response button to deal with the device and react to the stimulus, and an alphanumeric display[10]. Stimuli are completely random with an Interstimulus Interval (ISI) of 2-10 seconds and the standard duration of the test is considered 10 minutes. All data are stored

inside the device and they must be uploaded later on a computer where they can be post-processed using several available software, such as REACT[8].

2.2.1 Outcome Metrics

PVT gives several information that can help researchers to understand the subject condition: main outcomes in standard versions of this test are *reaction times*, expressed in milliseconds, *lapses*, defined as RT slower than 500 milliseconds and *false starts*, which consists in pressing the button before the stimulus appearance[11]. Usually, all RTs are post-processed to obtain their reciprocal transform, in this way the effect of possible extreme values is mitigated; then, information such as mean value, variability, fastest 10% and slowest 10% are extracted from raw data[11].

2.2.2 Application field

PVT is an useful tool in vigilant attention researches and its outcome metrics can also give information about sleep disorders, helping also to distinguish patients from healthy people[12]. In particular, as demonstrated by Thomann et al. (2014) test results for patients can be significantly worst for all the metrics and also age effect can be lost due to this performance impairment[12]. Moreover, not every sleep disorder has the same effect on PVT metrics, giving additional information about specific disturbances[12].

Sleep deprivation can be a crucial point related to safety in real life situations, as vigilant attention reduction can have severe consequences in real-driving circumstances. As Anderson&Horne (2006)[13] carried out in their work, monotonous task such as driving a car in an highway can be deeply affected by excessive daytime sleepiness. In this study, people were asked to make the PVT under different conditions: after a full night sleep (8 hour) or a night restricted sleep (5 hour) and with or without a source of disturbance like a television program which was running in their peripheral view. Results shown that sleep deprived subjects had worst PVT results and increased number of lapses compared to full night sleep individuals; moreover, head turn were higher both with and without the source of distraction for sleep deprived subject, demonstrating how dangerous can be this condition in real life driving[13].

2.2.3 Sensitivity to Fatigue and Sleep Loss

Psychomotor Vigilance Task is considered one of the best option to asses both fatigue and sleep loss since it allows to collect a significant quantity of information in a relative short period. Anyway each specific outcome metric of this test can be more or less relevant according to what we are interested in. Basner and Dinges

(2011) have evidenced how mean 1/RT and number of lapses can be really helpful in sleep deprivation studies[11]. Anyway, lapses are not only relevant in this specific analyses but they can also give precious information when dealing with fatigue[14]. In their research, Lee et al. (2010) have shown the important correlation between this vigilance test and fatigue symptoms giving evidences of its importance in this field of research[14].

Deeper study on the Psychomotor Vigilance Task and its relationship with fatigue have also stated the importance of the randomness range of the Interstimulus Interval. Longer ISI (5-10 seconds) lead to a richer correlation, suggesting that PVT sensitivity to fatigue can be improved removing shorter ISI (2-5 seconds) from the analysis or including this variable inside the statistical method[10].

2.2.4 Effects of Test Duration on PVT Performance

Standard duration of the Psychomotor Vigilance Task is 10 minutes and although the framework has always remained unchanged, several variants have been tested. In particular, the optimal duration of the test to optimize the data collection has been deeply discussed over years. Longest version can be really restrictive for on-field application and data collection can become really harsh, leading also to the possibility of corrupted data. For this reason the validation of shorter test has been investigated in several researches. The 3-minute PVT have been demonstrated to be a viable alternative for self-monitoring in several field, showing reliability and sensitivity to sleep deprivation[15]; however, this shorter version have shown less sensitivity to the percentage of lapses, suggesting that longer version of PVT should still be preferred in context where there are not time limitation[16]. Nevertheless, 5-minutes PVT represents the best trade-off for on-field application: ensuring good sensibility to fatigue and sleep loss compared with the standard version, while halving the duration[17].

2.2.5 PVT Implementation on Different Platforms

Spread of new hardware and fastest computers have soon led to the possibility of using computers not only for the post-process and analysis part but also for the development of the test, allowing to have everything on a single device capable to administer the test, to store data and to process and analyze the outcome metrics. PC-versions of PVT ensure accessibility and flexibility, giving also the possibility to set-up some test parameter such as the duration. The biggest obstacle has always been to replicate the precision of the the PVT-192, gold standard, in particular its error margin of 1 millisecond[18]. Computer version of the test should be calibrated since even small milliseconds of error in RT can affect the results[18]. Several recent studies have used PC versions of PVT, in particular Khitrov et al. (2013) has

demonstrated that their alternative can be considered reliable and valid as the gold standard PVT-192 [19]. In their specific application they validated their system with a sophisticate setup: a light sensor to detect the stimulus appearance on the monitor, a button sensor built inside the gaming mouse which is used to react, and the RTBox, a sub-millisecond measuring device (Figure 2.1)[19].

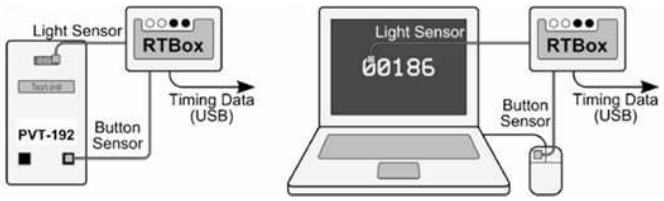

Figure 2.1: Calibration setup for PC-PVT. Source: Khitrov et al.. (2013)[19].

Test performed on various PC configurations showed that older hardware have negligible effects on the accuracy while using a gaming mouse instead of a standard one can reduce input lag and improve RT measurements accuracy[19]. In their latest version, they also introduced some additional functionality related to the integration of sleep and caffeine habits of the subjects[8].

Another alternative implementation of PVT is touch-screen devices implementation, that improves portability and usability . Nevertheless, devices like smartphones introduce several additional drawback that should be taken into account. Touchscreen latency for example is higher with respect to PVT-192 or computers, for this reason it should be measured using high speed camera and then subtracted to raw RT to clean the dataset[20]. Also device orientation have been demonstrated to be an influencing factor: horizontal orientation and thumb usage setup revealed faster results with respect to the vertical one and the index finger usage[20]. Touchscreen implementation opens also to different scenario for the user reaction technique: Kay et al. (2013) evidenced that the *touch down* technique, where the user can touch any part of the screen as soon as the stimulus appears, can be a viable option for touchscreen PVT[21]. There exist also validated version, like the NASA-PVT, which shown acceptable sensitivity to sleep loss but slightly slower Reaction Time with respect to the PVT-192[22]. Nevertheless, it could still represent a viable option for fatigue research if the touchscreen latency of the device is properly compensated [22].

2.3 Three Choice Vigilance Task

The Three Choice Vigilance Task (3CVT) is an alternative implementation of vigilance tests. Its main difference compared to PVT is its cognitive evaluation part. It consists of a 20-minute test characterized by 3 geometric shapes which

randomly appear one by one on the screen for 200 milliseconds[23]. In contrast to PVT, the user should not only react to the visual stimulus, but also distinguish between target and distracting shape by pressing the corresponding button[23]. The primary shape is the target, it appears 70% of the time, and the corresponding button is left arrow; secondary and tertiary shape are the distracting ones, each appearing 15% of the time, and the corresponding button is the right arrow (Figure 2.2)[23]. Interstimulus Interval in this case is slightly different: it goes from 1.5 seconds up to a maximum of 10 seconds.

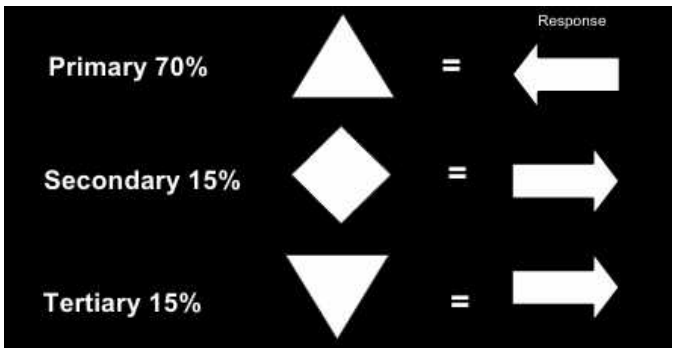

Figure 2.2: 3CVT visual stimuli and corresponding button. Source: Berka et al. (2005)[23].

2.3.1 Outcome Metrics

Main outcomes of the 3CVT are related to Reaction Time (RT): in addition to this variable, which is crucial to analyze speed and vigilance evolution, also accuracy is computed, defined as the percentage of correct answers given by the subject[24]. Also omissions are counted whenever no reaction is registered before 1.5 seconds after the stimulus appearance[23]. In certain cases, using more complex outcomes can be useful to have also additional information. For example, the F-Measure combines both reaction speed and accuracy inside a single score, allowing to go deeper in the evaluation of the cognitive performance[24].

2.3.2 Application Field

3CVT is an useful non-invasive tool in various clinical research settings: in particular, in combination with EEG this cognitive-vigilance test can give additional information to understand the evolution of various physiological phenomena[24]. 3CVT can be used in Mind Cognitive Impairement (MCI) diagnostic: this kind of cognitive degradation can be really invalidating and its early recognition can be really helpful for disease treatment[24]. Outcome of the test, in fact, can be sensitive biomarkers for early stages of dementia[24].

Furthermore, 3CVT have been demonstrated to be sensitive to sleep loss and consequent fatigue. In particular, as stated by Berka et al. (2005) this vigilance test can give information about sensitivity to sleep loss, enabling the possibility to identify people most susceptible to this condition [23]. Highly vigilance demanding application, such as aviation or driving situation, deals deeply with the necessity to asses both actual and future driver level of fatigue. To this purpose, outcomes like reaction speed or percentage of accuracy can give precious information about eventual performance drop, which can be really dangerous for these sensitive application[25].

2.4 Body Battery

Garmin's Body Battery (BB) is a 24h energy monitoring parameter. It is based on the combination of stress, physical activity, rest and sleep quality[1]. Particularly, Garmin's smartwatches analyze continuously Heart Rate (HR), Heart Rate Variability (HRV) and physical activity data[1]. BB parameter moves within the range between 5 and 100. Higher BB values are expected in the morning, when full night of restorative sleeping recharged the energy for the next day. On the other hand, lower values are expected in the evening after a full energy demanding day[1]. A full night sleep and rest have positive effects on energy level, but also taking a nap could give a little contribution to the BB[1]. For compatible devices, sleep pressure and relative homeostatic sleep drive can be integrated to retrieve additional information for BB estimation[1]. A negative influencing factor, instead, is the stress. Its level is evaluated using a combination of HR and HRV to obtain a value in the range between 0 and 100[26]. Then, 0-25 stress-level is classified as rest, contributing in recharging BB, since it reflects parasympathetic autonomous nervous system dominance[1]. Greater values of stress, instead, reflect progressively the degree of sympathetic autonomous nervous system supremacy, resulting in BB progressive draining[1]. However, although physical activity reduces BB in short term, having an higher fitness level allows to reduce stress influence, having positive effects in long-term[1].

Chapter 3

Methodology

3.1 Introduction

As stated in section 2.2.5, modern devices allow to develop finely tuned version of vigilance tests which can be tailored on specific requirements. The simple nature of both vigilance tasks gives the opportunities to develop specific applications for each of them. Despite all limitation related to the validation of these self-developed tools, the decision to proceed towards this direction was driven by the great flexibility and the great customization that these kinds of solution offer, which align with the goals of this study. Furthermore, this choice is justified given that already validated PC-versions requires the computation of the specific device latency, otherwise their greater precision would be lost. Since this investigation is based on delivering the tests on participants' devices, compute all latencies would have been very time consuming and costly. Moreover, to create a standard and concise procedure, the 20 minute duration of the 3CVT was discarded, in order to evaluate only shorter version of it, equal to the PVT ones. Although in this case there were not study supporting their validity, longer duration would have been to complex for the experimental implementation, resulting in poor datasets. These choices are also justified given that this study represents a first step in evaluating the potential correlations between these tests and body battery, for this reason having tailored versions could offer significant advantages in term of flexibility.

3.2 Development of the Tests

In this work, both tests were developed using the Python programming language to ensure future flexibility and compatibility with any potential device. Following this philosophy, built-in libraries, such as *Tkinter* for the creation of the Graphical User Interface (GUI), have been preferred to external ones where possible. To ensure

reusability of the code, the same structure was used for both the tests. Then, each of them was enhanced with specific functions to cover all their features.

3.2.1 Test Structure

An introductory page appears as soon as the file is run: it gives all the essential information necessary to understand how to deal with the test. The structure is the same for both, with the specific description for each one, as shown in Figure 3.1 and Figure 3.2. Moreover, the PVT includes the Karolinska Sleepiness Scale, since it was added for completeness in the final phase protocol.

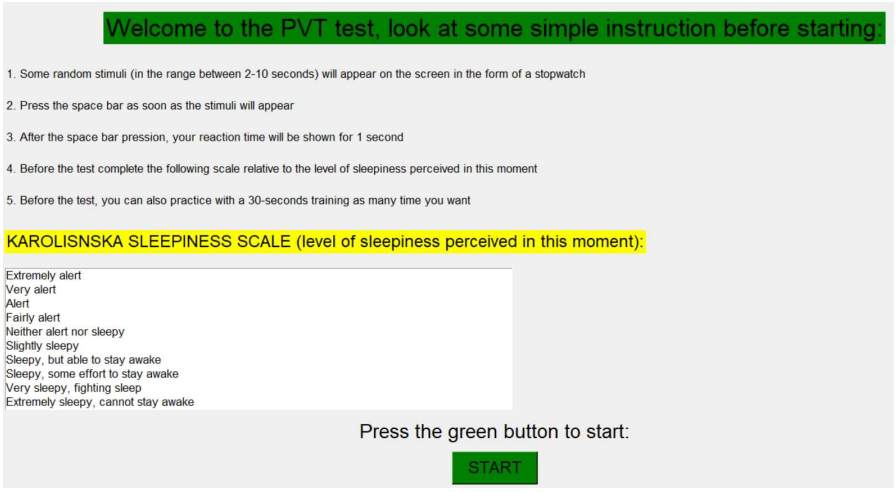

Figure 3.1: PVT introductory page

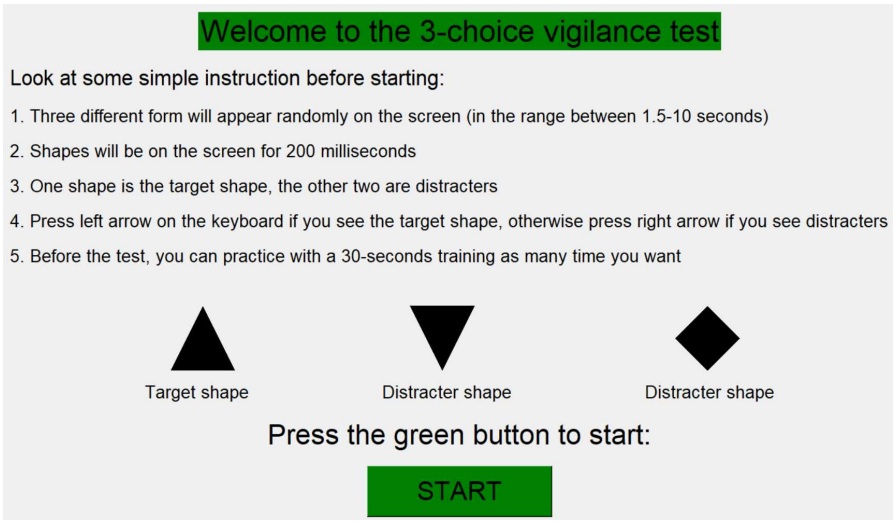

Figure 3.2: 3CVT introductory page

Ones the user submits their answer to the KSS, they are allowed to access the *set-up page*, which has the same structure for both tests (Figure 3.3). This page is designed to store all important information: first section is about subject's name, while the second part is for test duration, which can be set between 3, 5 or 10 minutes according to the protocol adopted. The default duration is 5 minutes. Both tests include an initial trial lasting 30 seconds, providing the opportunity to new users to become familiar with them. This training part can be repeated endlessly just pushing the corresponding button, if necessary.

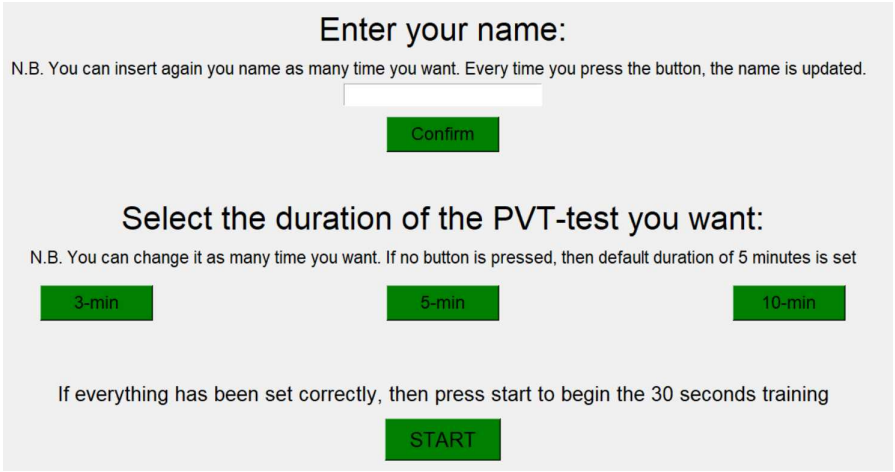

Figure 3.3: PVT set-up page

Ones the test is completed, a *recap page* (Figure 3.5 and 3.7) appears on the

screen, summarizing all the outcomes from the test and providing the user with an immediate response about their performance.

3.2.2 Psychomotor Vigilance Task (PVT)

As described in section 2.2, Psychomotor Vigilance Task (PVT) is based on the appearance of visual stimulus in the form of a stopwatch. In particular, the running stopwatch appears randomly in the range between 2 and 10 seconds on a white background (Figure 3.4).

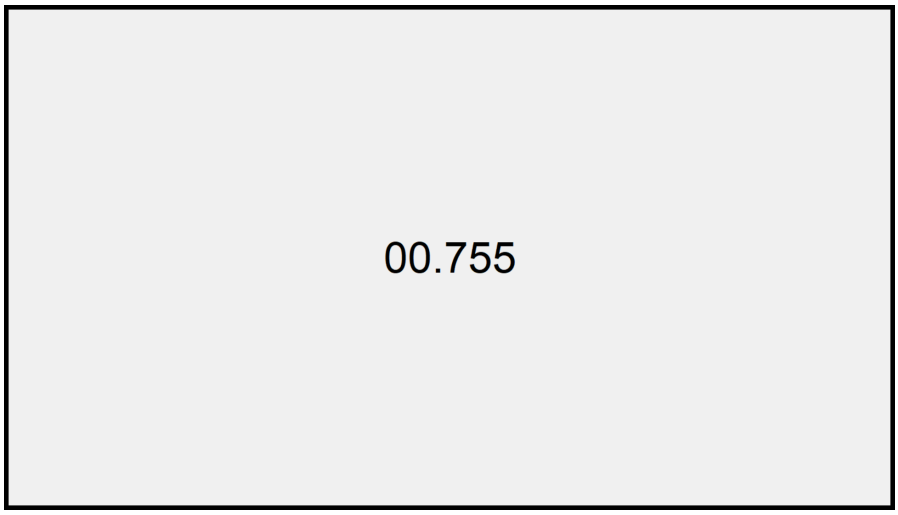

Figure 3.4: PVT stimulus example

Outcomes metrics are directly computed once the test is over: a specific page is shown to the participant in order to have a quick look at it. The PVT outcome metrics are outlined below:

- **Lapses percentage**: Reaction times slower than 500 milliseconds
- **Invalid starts**: User's reaction before the stimulus appearance
- **Mean value of the reciprocal transform of the reaction times excluding lapses**
- **Standard deviation of the reciprocal transform of the reaction times excluding lapses**
- **Mean value of the reciprocal transform of the reaction times including lapses**

- **Standard deviation of the reciprocal transform of the reaction times including lapses**
- **Mean value of the reciprocal transform of the 10% fastest reaction times**
- **Standard deviation of the reciprocal transform of the 10% fastest reaction times**
- **Mean value of the reciprocal transform of the 10% slowest reaction times**
- **Standard deviation of the reciprocal transform of the 10% slowest reaction times**

Notice that the computation of some outcome metrics was added to the PVT application just for the final phase, once the optimization phase was over. At the end of the test, all these information are provided for a quick overview to the participant. Figure 3.5 shows the specific recap page that is shown to the participant after the PVT is finished.

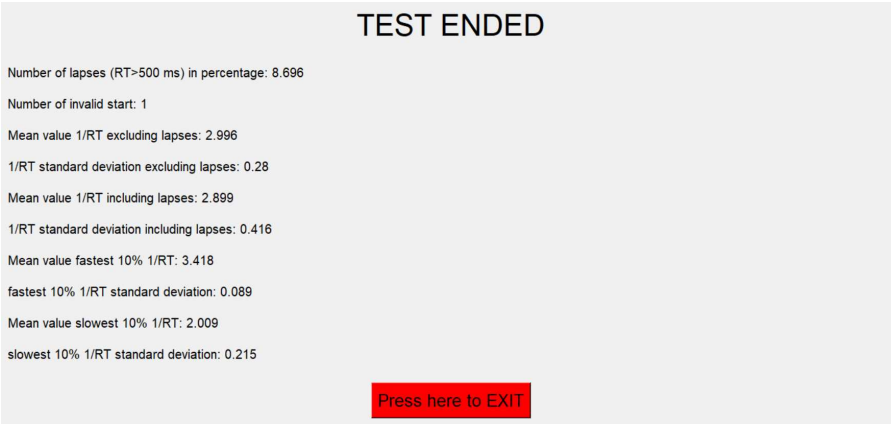

Figure 3.5: PVT recap page example

To accomplish all PVT functionalities, several key functions have been specifically designed. The most important ones are listed below:

- **random_timer()**: To deal with the random ISI (2-10 seconds) of the stimuli
- **start_stimulus()**: To start the stimulus and save the stimulus appearance time
- **button_pressed()**: To deal with the participant's reaction

- **reaction_times_management()**: To distinguish between lapses, false starts, and valid Reaction Time (RT)
- **compute_outcomes()**: To compute all the outcomes of the test
- **convert_outcomes_reciprocal()**: To compute all the inverse transform outcomes of the test
- **save_results()**: To save all outcome metrics inside an excel file

3.2.3 Three choice Vigilance Task (3CVT)

Three Choice Vigilance Task (3CVT) works quite similar to the PVT but introduce some important additional functionalities. As stated in section 2.3, the significant difference between these two tests is related to the cognitive part: visual stimuli are no more of a single type, but the participant has also to distinguish between target and distracting shape, pressing the corresponding button. Visual stimuli are in the form of three different shape: if the target appears (triangle shape) on the white background, the user should press the *left arrow*, while if one of the two distractors (diamond shape or inverted triangle shape) is shown, the corresponding button is the right arrow. An example of visual stimuli (diamond shape) is provided in Figure 3.6.

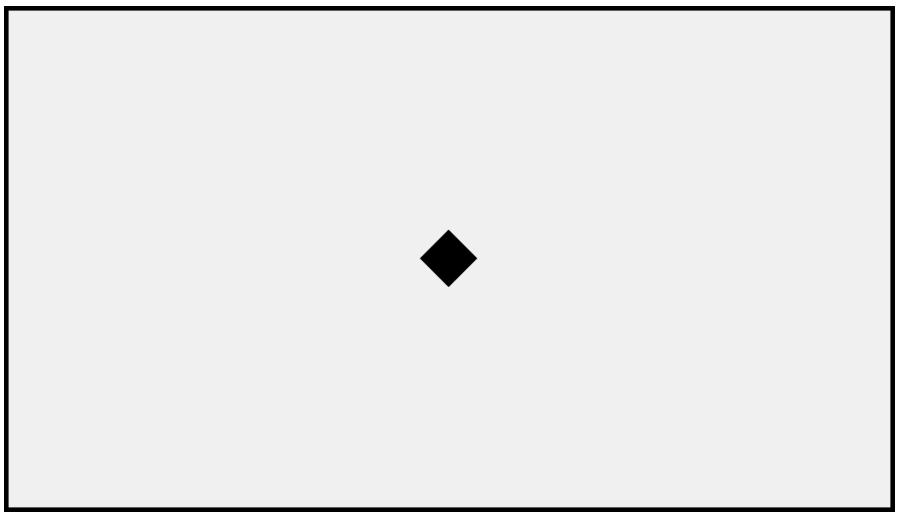

Figure 3.6: 3CVT stimulus example

The presence of the cognitive part, implies the necessity to distinguish also between correct and wrong reactions. Outcome metrics, for this reason, are similar but not exactly the same. 3CVT outcome metrics are reported below:

- **Omissions**: No reaction before 1.5 seconds after stimulus appearance, expressed in percentage
- **Accuracy**: Percentage of correct reactions out of the total
- **Invalid starts**: User's reaction before the stimulus appearance
- **Mean value of the reciprocal transform of the correct reactions**
- **Standard deviation of the reciprocal transform of the correct reactions**
- **Mean value of the reciprocal transform of the incorrect reactions**
- **Standard deviation of the reciprocal transform of the incorrect reactions**
- **Mean value of the reciprocal transform of all reactions**
- **Standard deviation of the reciprocal transform of all reactions**

Also for 3CVT the recap page is shown once the test is over, an example is provided in Figure 3.7, providing the user with immediate feedback about their performance.

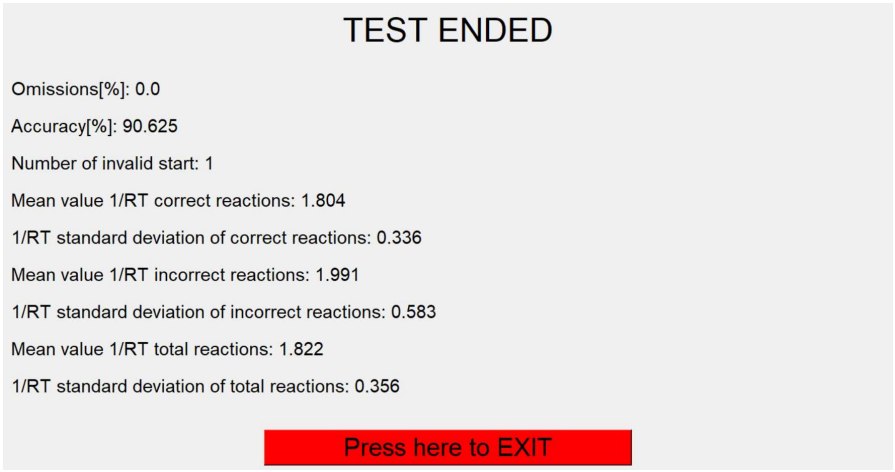

Figure 3.7: 3CVT recap page example

While both tests have a similar structure, 3CVT requires some additional functions to implement the cognitive part. For example, the distinction between correct and wrong reactions, led to the splitting of the PVT function **button_pressed()** into two separate entities: one to deal with target reaction and the other for the

distractor responses. The most important functions developed for this test are outlined below:

- **random_timer()**: To implement the random behaviour of the stimuli appearance
- **start_stimulus()**: To make one of the shape appear randomly (70% probability for target while 15% for each distractor) on the screen, while storing the corresponding time
- **omission_check()**: To check any omission, defined as user's reactions longer than 1,5 seconds
- **button_left_pressed(event)**: To manage the user response to the target stimuli
- **button_right_pressed(event)**: To manage the user response to distractor stimuli
- **reaction_times_management()**: To distinguish between omissions (RT bigger than 1.5 seconds), correct and incorrect responses
- **compute_outcomes()**: To compute all the outcomes of the test
- **convert_outcomes_reciprocal()**: To compute all the inverse transform outcomes of the test
- **save_results()**: To save all outcome metrics inside an excel file

3.2.4 Data Saving and Storage

As mentioned in the section describing the functions implemented for every test, a specific function was dedicated to the storage of the outcomes. Each application was designed to create, if it was not already existing, a specific excel file, named respectively *PVT_output* and *3CVT_output* where all the output of the test were stored. Once the file was created, at each iteration the specific sheet of the file was updated according to the duration selected. Each time, the participant's name, the time of the day and the results obtained during that specific session were saved, ensuring a clear and organized history for both tests. In the final phase of the data collection, also Karolinska Sleepiness Scale score was integrated between the data saved automatically inside the excel file. Moreover, to ensure further robustness to the PVT test in the last phase, a second excel file named *PVT_output_raw_data* was introduced in the final saving step of the application. In this way, a raw data history is added alongside the processed data contained in the *PVT_outcome*, giving more flexibility for the post-process data analysis.

3.3 Experimental Protocol

The aim of this study is to find possible correlations between vigilance tasks and Garmin's Body Battery in order to assess whether these tests can serve as useful tools to estimate BB. As a result, the level of fatigue could be estimated objectively giving valuable insight about subject's actual and future performance. The underlying idea of this work is to correlate the BB value at a specific moment with the outcome metrics from the considered vigilance tests.

3.3.1 Protocol Description and Evolution

Participants were evaluated over five days, collecting data during different times to evaluate the evolution both from a daily perspective and over the medium term. Furthermore, the test was not required to be completed during consecutive days to give more flexibility to participants and to enhance the robustness of any potential correlations. Each subject was equipped with a Garmin smartwatch at least 3 days before the beginning of the procedure to have a reliable value for the Body Battery. They were also asked to keep wearing the smartwatch until the end of the 5-day data collection period, while continuing to live their normal lives. Vigilance task applications were shared with them in order to allow its administration on subject's personal computer. When both tests were performed, PVT was always done first. At the end of the fifth day, they were asked to share the output excel files created by the tests.

Preliminary Phase

Initial step was dedicated to understand which duration best fit for the final purpose, considering both the amount of data obtained during each session and the time required by this kind of tasks. Only one subject was involved during this phase. The test protocol required the administration of both PVT and 3CVT with duration of 3, 5 and 10 minutes. The subject had to complete both tests at three different moments of the day: in the morning (about 10:00 A.M.), at lunchtime (about 01:30 P.M.) and in the evening (about 04:30 P.M.). During this preliminary phase BB was collected both at the beginning and at the end of each test.

Optimization Phase

Then, focus was moved towards the selection of a single vigilance test, in order to lighten the load on participants and to obtain more realistic results. Five participants took part to this phase. They were asked to perform both PVT and 3CVT with a duration of 5 minutes three times a day. Considering the feedback received after the first phase, at least 30 minutes were introduced between one

PVT session and the following 3CVT to reduce the influence between successive administrations.

Final Phase

At this point, once the final set-up was clear and well-defined, the number of participants was increased to have as much data as possible. Fourteen people took part to this phase: seven men and seven women. Each subject was asked to perform the 5-minute PVT only two times per day: in the morning (about 10:00 A.M.) and in the evening (about 04:30 P.M.), for five days long.

3.3.2 Tools and Procedure

As stated in section 3.3.1, each participant was equipped with a Garmin smartwatch. However, due to limits of feasibility, not every subject received the same model. To overcome the limited number of devices in the last phase, test administration was performed during different moments in order to allow reusability of the same smartwatches with different subjects. At the end of every usage they were initialized in order to avoid any possible data overlap from different users. Moreover, no supervision was scheduled during data collection, as it was not compatible with the increasing number of participants. The BB value relative to each session was retrieved at a later time from the database on the *Garmin Connect* website and added in the output excel files manually for both the tests.

3.3.3 Statistical Analysis

Data manipulation was performed both in the tests themselves and in the post-processing phase. Both applications, have a specific function (section 3.2.2 and section 3.2.3) responsible for the calculation of the reciprocal transform of the Reaction Time (RT) in order to reduce sensitivity to potential extreme values [11]. This first step was crucial since it was the standard in vigilance task application. To have a wider and more robust overview of the possible correlation between tests outcome metrics and Body Battery (bb), different types of correlations were computed. Initially, Pearson (r), Spearman (rho) and Kendall (tau) coefficients were all considered. However, since Pearson's correlation is based on the assumption that the dataset is normally distributed, it was soon discarded due to the cyclic nature of the BB. Both these correlation coefficients investigate the monotonic nature of the correlation. Spearman's rho indicates direction and strength of the correlation by analyzing the rank of the two variable investigated, that is the position of the observations when sorted in ascending or descending order. On the other hand, Kendall's tau investigates the concordance between pairs of observations to

evaluate the agreement of the two variables. Moreover, Kendall's correlation is less sensitive to small errors or rank variations, leading to results that are more robust to outliers. The idea was to look for any possible monotonous correlation, both positive or negative and linear or non-linear, between every outcome metric of each test and the BB. This choice was led by the intention to investigate two slightly different perspectives, obtaining a sort of cross validation. Values close to 1 (or -1 for negative correlation) for both rho and tau should be considered good. This means that if the correlation coefficient is close to 1, indicating positive monotonic relationship, then when one variable increases (or decreases), the other follow a similar trend in a predictable manner. On the other hand, values close to -1, indicates a negative monotonic relationship, which means that when one variable increases, the other decreases (or vice versa) following a predictable but opposite trend. In the final phase, also the p-value calculation was introduced: low values of this parameter (p-value < 0.05) provide further valuable insights about the statistical significance of the correlation coefficient, adding more robustness to the analyses.

Chapter 4

Results

4.1 Preliminary Results

The aim of the preliminary phase was to assess which test duration best fits with our research goal. To this purpose, the daily trends were evaluated for each test under different duration condition (3,5 and 10 minutes). Both PVT and 3CVT outcome metrics were analyzed to determine whether there was a performance impairment that could reflect the BB decline throughout the day. Specifically, BB follows a decreasing trend during the day, starting from higher values in the morning and reaching the smallest in the evening. Therefore, the outcome metrics of the tests were expected to follow a similar deteriorating pattern, reflecting the progressive decline in energy. In particular, for 3CVT, omissions percentage, invalid starts and all the standard deviations should increase throughout the day, while accuracy and all the mean values (as we are considering their reciprocal transform) should decrease towards evening. On the other side, for PVT, lapses percentage, invalid starts and all the standard deviations should have the lowest values in the morning, while all mean values should follow a negative trend throughout the day. To highlight the performance impairment trends, a blue color scale was used, where lighter shades represent better performance while darker shades represent poorer results. In case where the trend was not consistent throughout the entire day, but only significant in the morning and evening, the lunchtime variable was left uncoloured to avoid confusion.

4.1.1 Three Choice Vigilance Task

3-minute 3CVT

In the 3-minute 3CVT, both standard deviation of the reciprocal of the correct reaction times and the mean value of the reciprocal of the incorrect reaction times

varied as expected between the morning and evening, but did not maintain the same trend during lunch. In particular:

- The omissions percentage showed a general consistent deteriorating trend, reaching 1.3% in the evening.
- The accuracy decreased constantly throughout the day by around 3 point percentage.
- The invalid starts passed from 0 in the morning to around 1 in the evening.
- The mean value of the reciprocal of the correct reaction times decreased by about 60 1/ms.
- The standard deviation of the reciprocal of the incorrect reaction times showed a fluctuating trend, but consistent between morning and evening deteriorating by around 80 1/ms.

	MORNING	LUNCH TIME	EVENING
Omiss. (%)	0,000	0,476	1,333
Acc. (%)	95,380	92,695	89,801
Inv. Starts	0,000	0,867	0,933
Mean Corr. (1/s)	2,118	2,153	2,060
Std Corr. (1/s)	0,350	0,324	0,290
Mean Incorr. (1/s)	1,494	2,383	1,762
Std Incorr. (1/s)	0,178	0,373	0,258
BB Start (score)	12,667	59,133	48,133
BB End (score)	12,000	49,000	31,500

Table 4.1: 3-min 3CVT preliminary results

5-minute 3CVT

In the 5-minute 3CVT, omissions and mean value of the reciprocal of both correct and incorrect reaction times varied as expected between the morning and evening but do not maintain the same trend during lunch. Specifically:

- Omissions increased slightly by about 0.5 percentage points.
- The mean value of reciprocal of the correct reaction times increased slightly by about 50 1/ms.

- The mean value of the reciprocal of the incorrect reaction times decreased significantly by about 285 1/ms.

	MORNING	LUNCH TIME	EVENING
Omiss. (%)	0,884	0,455	1,430
Acc. (%)	92,950	95,357	94,910
Inv. Starts	1,000	0,600	0,333
Mean Corr. (1/s)	2,091	2,017	2,049
Std Corr. (1/s)	0,394	0,306	0,348
Mean Incorr. (1/s)	2,191	1,786	1,905
Std Incorr. (1/s)	0,292	0,156	0,203
BB Start (score)	54,800	36,000	15,500
BB End (score)	54,400	35,800	15,500

Table 4.2: 5-min 3CVT preliminary results

10-minute 3CVT

In the 10-minute 3CVT, accuracy, invalid starts and the standard deviation of the reciprocal of the incorrect reactions reflected the worsening trend of BB throughout the entire day. Moreover, also the standard deviation of the reciprocal of the correct reactions had a similar trend, remaining quite constant between lunch and evening. Particularly:

- Accuracy followed a consistent deteriorating trend throughout all the sessions, decreasing by about 0.5 percentage point in each.

- Invalid starts increased from about zero in the morning to around 2 by the evening.

- The standard deviation of the reciprocal of the correct reactions remained relatively constant throughout the day, with only a slight decrease between morning and lunch.

- The standard deviation of the reciprocal of the incorrect reactions increased throughout all the day, rising by about 500 1/ms between morning and evening.

	MORNING	LUNCH TIME	EVENING
Omiss. (%)	1,427	0,421	1,080
Acc. (%)	94,518	93,929	93,273
Inv. Starts	0,200	1,000	1,600
Mean Corr. (1/s)	2,029	2,047	2,055
Std Corr. (1/s)	0,316	0,324	0,323
Mean Incorr. (1/s)	2,409	2,224	2,512
Std Incorr. (1/s)	0,189	0,272	0,670
BB Start (score)	63,200	41,200	17,000
BB End (score)	62,200	40,200	16,000

Table 4.3: 10-min 3CVT preliminary results

4.1.2 Psychomotor Vigilance Task

3-min PVT

In the 3-minute PVT, lapses percentage varied as expected between morning and evening, but did not maintain the same trend during lunch. Invalid starts, on the other hand, remained relatively constant between morning and evening, with a slight increase during lunch session. In particular:

- Lapses increased by about 2 percentage points, starting at 3.919% in the morning and finishing at 5.767% in the evening.

- Invalid starts slightly increased by 0.2 between morning session and evening.

	MORNING	LUNCH TIME	EVENING
Lapses(%)	3,919	0,800	5,767
Inv. Starts	1,000	1,800	1,200
Mean No Lapses (1/s)	3,264	3,539	3,371
Std No Lapses (1/s)	0,598	1,641	0,481
Mean Lapses (1/s)	3,205	3,541	3,274
Std Lapses (1/s)	0,659	1,668	0,609
BB Start (score)	66,800	41,000	19,800
BB End (score)	66,400	40,400	19,400

Table 4.4: 3-min PVT preliminary results

5-min PVT

In the 5-minute PVT, invalid starts and standard deviations of the reciprocal of the reactions, both including and excluding lapses, varied as expected between morning and evening, but do not follow the same trend during lunch. Specifically:

- Invalid starts increased slightly, reaching a maximum of 1.8 in the evening.
- The standard deviation of the reciprocal of the reaction times without lapses increased slightly by about 60 1/ms.
- The standard deviation of the reciprocal of the reaction times with lapses increased slightly by about 30 1/ms.

	MORNING	LUNCH TIME	EVENING
Lapses(%)	3,746	2,493	3,335
Inv. Starts	1,600	0,600	1,800
Mean No Lapses (1/s)	3,290	3,223	3,306
Std No Lapses (1/s)	0,427	0,516	0,489
Mean Lapses (1/s)	3,234	3,189	3,257
Std Lapses (1/s)	0,514	0,549	0,546
BB Start (score)	67,400	41,800	21,600
BB End (score)	66,400	41,600	21,200

Table 4.5: 5-min PVT preliminary results

10-min PVT

In the 10-minute PVT, only invalid starts followed a trend similar to BB throughout the day. Particularly:

- Invalid starts were 3.2 in the morning, increased to 3.8 at lunch and rose to 4 in the evening.

	MORNING	LUNCH TIME	EVENING
Lapses(%)	3,989	5,140	3,924
Inv. Starts	3,200	3,800	4,000
Mean No Lapses (1/s)	3,270	3,495	3,337
Std No Lapses (1/s)	0,548	1,939	0,461
Mean Lapses (1/s)	3,192	3,409	3,276
Std Lapses (1/s)	0,630	1,952	0,541
BB Start (score)	63,200	41,200	15,800
BB End (score)	61,600	40,200	15,400

Table 4.6: 10-min PVT preliminary results

4.1.3 Preliminary results discussion

The first clear finding from the preliminary examination was that it was not necessary to evaluate the Body Battery twice (before and after the test), as it did not give additional information. Although this was expected for the 3-minute version of both tests, due to Garmin's BB having a sampling rate of 3 minutes, the results suggested that for future experimental steps, it was sufficient to collect the BB value once before the test begins, regardless of the selected test duration. On the other hand, trend analyses indicated that the optimal configuration for both tests was the 5-minute version, as it offered the best trade-off between the amount of data collected, the time spent on task, and the trend evolution throughout the day. Specifically, the 5-minute versions of both tests yielded 3 outcomes that exhibited similar trends with respect to BB. On the other hand, the 10-minute version of the 3CVT produced 3 relevant outcomes out of 7, while the 10-minute version of the PVT showed only 1 relevant outcome out of 6. In the end, for the 3-minute versions, the main drawback is the very small number of data samples that can be collected in this short time. However, the heavy load involved in administering all test configurations to a single subject simplified this initial step but may have impacted the results. Thus, considering these preliminary results and adopting a cautious approach, the 5-minute configurations seem to represent the best trade-off for the next experimental steps.

4.2 Optimization Phase Results

Once the optimal trade-off for test duration was determined, the next step was to identify which test could deliver better performance, allowing the final phase to focus on a single test. Both correlation performance and trends throughout the day were evaluated during this step. This process enabled the fine-tuning and

simplification of the test protocol for the final phase, avoiding potential protocol overload that could impact the correlation results. Two men and two women participated in the data collection campaign, in addition to the male subject from the preliminary phase, bringing the total number of participants to five: three men and two women. The administration protocol for both the 5-minute 3CVT and the 5-minute PVT remained unchanged: the tests were administered three times a day (morning, lunchtime, and evening) for five days (not necessarily consecutive).

In the following table, each subject is identified by a capital letter and a number, which represent the subject's gender (M for male and F for female) and the progressive participant number. To illustrate the daily performance trends, the same colour approach of the preliminary phase is used: lighter shades of blue indicating better performance and darker shades representing performance deterioration. Given the exploratory nature of this study and the complexity of the physiological parameters involved, a cautious approach was adopted in the interpretation of correlation results. Specifically, a conservative threshold of ± 0.3 for the strength of both Spearman and Kendall correlation coefficients was selected to highlight potentially promising relationships, which are shown in light blue. Although this threshold might appear low in comparison to standard evaluations, it is appropriate in this context, considering the limited sample size and the complexity of the physiological phenomena being involved. At this stage, the aim was to avoid the underestimation of any potentially valuable information, as these insights could become more significant with a larger and more representative sample. In detail: correlation coefficients below ± 0.3 indicate weak correlations, while values between ± 0.3 and ± 0.5 suggest moderate correlations. Correlations between ± 0.5 and ± 0.7 were considered strong, and values greater than ± 0.7 represented very strong or near-perfect correlations. It is important to note that single-subject analyses relied on a limited amount of data. For this reason, the "ALL" category, where data from all subjects were combined, was the most significant. This larger dataset provided more statistically robust results, even though it did not account for potential subgroup distinctions, such as gender. Therefore, the results analysis primarily focused on these aggregated outcomes. However, individual subject values were also reported for completeness and commented when relevant. Moreover, a special focus was given to the potential impact of removing the lunch session from both tests, in order to determine if this protocol simplification could be useful in the final phase.

4.2.1 Three Choice Vigilance Task

3CVT Daily Trend Analyses

Omissions percentages increased between morning and evening for 3 out of 5 participants, with the "ALL" category showing a particularly significant rise of

around 1 percentage point (Table 4.7). Accuracy followed the same trend as the BB for most participants, with the exception of two subjects, who were still aligned with the overall trend when the lunch session was excluded. In the "ALL" category, accuracy decreased by approximately 1 percentage point per session, with a total drop of 2.4% from morning to evening (Table 4.7). Invalid starts did not appear to be significant, as the performance decline was minimal for all participants and nearly constant in the combined dataset (Table 4.7). The mean values of the reciprocal of both correct and incorrect reaction times also seemed less relevant, as they showed variability or even opposite trends. However, the standard deviations of the reciprocal reaction times provided more insight: in the combined dataset, both correct and incorrect reciprocal reaction times showed a slight or moderate worsening trend from morning to evening, with a continuous deterioration also during the lunch session for the incorrect reaction time standard deviation (Table 4.7). Finally, the mean value of all reaction times (both correct and incorrect) remained relatively constant throughout the day in the combined dataset, whereas the standard deviation became more significant, increasing by around 60 1/ms during the evening session (Table 4.7).

ALL	MORNING	LUNCH TIME	EVENING
Omiss. (%)	1,201	1,872	1,950
Acc. (%)	93,546	92,540	91,162
Inv. Starts	0,600	0,500	0,560
Mean Corr. (1/s)	1,827	1,858	1,848
Std Corr. (1/s)	0,330	0,354	0,337
Mean Incorr. (1/s)	1,661	1,989	2,020
Std Incorr. (1/s)	0,218	0,299	0,461
Mean Tot. (1/s)	1,835	1,872	1,872
Std Tot. (1/s)	0,335	0,366	0,397
BB (score)	52,913	38,136	28,318

Table 4.7: 5-min 3CVT daily trend for ALL subjects

M1	MORNING	LUNCH TIME	EVENING
Omiss. (%)	0,000	2,084	2,044
Acc. (%)	90,491	89,592	84,766
Inv. Starts	0,400	0,500	0,800
Mean Corr. (1/s)	2,088	2,079	2,060
Std Corr. (1/s)	0,346	0,393	0,387
Mean Incorr. (1/s)	2,247	2,496	2,186
Std Incorr. (1/s)	0,289	0,524	0,354
Mean Tot. (1/s)	2,107	2,112	2,081
Std Tot. (1/s)	0,349	0,413	0,391
BB (score)	41,000	30,000	26,000

Table 4.8: 5-min 3CVT daily trend for subject M1

M2	MORNING	LUNCH TIME	EVENING
Omiss. (%)	0,426	1,304	2,127
Acc. (%)	96,281	94,848	92,793
Inv. Starts	0,200	0,000	0,200
Mean Corr. (1/s)	1,824	1,736	1,714
Std Corr. (1/s)	0,293	0,321	0,266
Mean Incorr. (1/s)	1,447	1,607	1,717
Std Incorr. (1/s)	0,482	0,458	0,521
Mean Tot. (1/s)	1,823	1,747	1,734
Std Tot. (1/s)	0,303	0,332	0,297
BB (score)	49,400	30,600	22,000

Table 4.9: 5-min 3CVT daily trend for subject M2

M3	MORNING	LUNCH TIME	EVENING
Omiss. (%)	1,242	0,000	0,784
Acc. (%)	95,661	96,385	94,704
Inv. Starts	0,200	0,200	0,800
Mean Corr. (1/s)	2,053	1,982	1,998
Std Corr. (1/s)	0,317	0,323	0,366
Mean Incorr. (1/s)	1,873	2,171	2,348
Std Incorr. (1/s)	0,099	0,340	0,327
Mean Tot. (1/s)	2,061	1,994	2,009
Std Tot. (1/s)	0,317	0,338	0,372
BB (score)	55,000	38,600	24,800

Table 4.10: 5-min 3CVT daily trend for subject M3

F1	MORNING	LUNCH TIME	EVENING
Omiss. (%)	1,233	1,491	0,879
Acc. (%)	90,157	90,172	90,490
Inv. Starts	0,800	1,000	0,800
Mean Corr. (1/s)	1,690	1,902	1,840
Std Corr. (1/s)	0,407	0,413	0,384
Mean Incorr. (1/s)	1,888	1,927	2,557
Std Incorr. (1/s)	0,511	0,323	0,533
Mean Tot. (1/s)	1,707	1,910	1,909
Std Tot. (1/s)	0,419	0,421	0,410
BB (score)	62,800	46,250	35,200

Table 4.11: 5-min 3CVT daily trend for subject F1

F2	MORNING	LUNCH TIME	EVENING
Omiss. (%)	3,103	5,089	3,917
Acc. (%)	95,138	90,165	93,059
Inv. Starts	1,400	1,000	0,200
Mean Corr. (1/s)	1,479	1,590	1,630
Std Corr. (1/s)	0,298	0,344	0,296
Mean Incorr. (1/s)	0,848	1,793	1,290
Std Incorr. (1/s)	0,003	0,161	0,043
Mean Tot. (1/s)	1,476	1,600	1,626
Std Tot. (1/s)	0,297	0,350	0,297
BB (score)	58,667	47,000	41,500

Table 4.12: 5-min 3CVT daily trend for subject F2

3CVT Correlation strength

This analysis was conducted following two parallel paths: one including the lunch session (Table 4.13 and 4.14) and the other excluding it (Table 4.15 and 4.16). In this context, results for the "ALL" category showed a slight to moderate increase in the absolute value of most coefficients, indicating that removing the lunch session may lead to improvement. Notably, for the "ALL" category, only the Spearman's rho coefficient of the mean value of the reciprocal of the incorrect reaction times exceeded ± 0.3, suggesting that, when the sample size will be increased, dividing the data based on characteristics such as gender or age could be a more effective strategy. Additionally, the removal of the lunch session positively affected individual subject correlations, as the total number of significant coefficients (greater than ± 0.3) increased for both Spearman's and Kendall's coefficients.

	M1	M2	M3	F1	F2	ALL
Omiss. (%) - BB	-0,347	-0,568	0,181	-0,064	-0,037	-0,057
Acc. (%)- BB	0,464	0,247	0,217	-0,176	0,401	0,062
Inv. Starts - BB	-0,377	0,272	-0,205	0,100	0,599	0,165
Mean Corr. (1/s) - BB	0,073	0,439	0,295	-0,233	-0,316	-0,180
Std Corr. (1/s) - BB	-0,009	0,236	-0,468	-0,103	-0,311	0,076
Mean Incorr. (1/s) - BB	-0,281	-0,043	-0,154	-0,405	-0,537	-0,324
Std Incorr. (1/s) - BB	-0,344	0,109	-0,345	0,077	0,408	-0,141
Mean Tot. (1/s) - BB	0,053	0,343	0,250	-0,389	-0,274	-0,214
Std Tot. (1/s) - BB	-0,135	0,218	-0,614	-0,346	-0,292	-0,018

Table 4.13: 5-min 3CVT morning/lunch/evening Spearman correlations

	M1	M2	M3	F1	F2	ALL
Omiss. (%) - BB	-0,268	-0,434	0,166	-0,063	-0,023	-0,049
Acc. (%)- BB	0,371	0,232	0,137	-0,144	0,315	0,042
Inv. Starts - BB	-0,321	0,230	-0,174	0,053	0,510	0,129
Mean Corr. (1/s) - BB	0,056	0,333	0,191	-0,133	-0,135	-0,112
Std Corr. (1/s) - BB	-0,022	0,144	-0,306	-0,044	-0,250	0,048
Mean Incorr. (1/s) - BB	-0,224	-0,039	-0,067	-0,287	-0,296	-0,209
Std Incorr. (1/s) - BB	-0,233	0,090	-0,268	0,066	-0,303	-0,106
Mean Tot. (1/s) - BB	0,056	0,257	0,172	-0,221	-0,296	-0,134
Std Tot. (1/s) - BB	-0,112	0,153	-0,471	-0,221	-0,303	-0,006

Table 4.14: 5-min 3CVT morning/lunch/evening Kendall correlations

	M1	M2	M3	F1	F2	ALL
Omiss. (%) - BB	-0,380	-0,560	-0,272	-0,410	0,266	-0,091
Acc. (%)- BB	0,576	0,370	-0,030	0,500	0,116	0,165
Inv. Starts - BB	-0,284	0,348	0,104	0,894	-0,244	0,128
Mean Corr. (1/s) - BB	-0,127	0,564	-0,261	-0,600	0,455	-0,177
Std Corr. (1/s) - BB	0,103	0,207	0,055	-0,975	-0,321	0,114
Mean Incorr. (1/s) - BB	-0,248	-0,091	-0,733	-0,564	-0,139	-0,357
Std Incorr. (1/s) - BB	-0,382	-0,075	-0,079	-0,224	-0,381	-0,164
Mean Tot. (1/s) - BB	-0,127	0,576	-0,455	-0,600	0,406	-0,212
Std Tot. (1/s) - BB	0,067	0,085	-0,333	-1,000	-0,413	-0,015

Table 4.15: 5-min 3CVT morning/evening Spearman correlations

	M1	M2	M3	F1	F2	ALL
Omiss. (%) - BB	-0,304	-0,435	-0,205	-0,316	0,307	-0,065
Acc. (%)- BB	0,511	0,333	-0,045	0,400	0,090	0,111
Inv. Starts - BB	-0,243	0,298	0,053	0,837	-0,209	0,097
Mean Corr. (1/s) - BB	-0,111	0,467	-0,156	-0,400	0,333	-0,105
Std Corr. (1/s) - BB	0,067	0,159	0,067	-0,949	-0,200	0,080
Mean Incorr. (1/s) - BB	-0,200	-0,045	-0,556	-0,316	-0,067	-0,219
Std Incorr. (1/s) - BB	-0,244	-0,072	-0,067	-0,120	-0,328	-0,122
Mean Tot. (1/s) - BB	-0,111	0,511	-0,244	-0,400	0,289	-0,128
Std Tot. (1/s) - BB	0,022	0,135	-0,244	-1,000	-0,270	-0,005

Table 4.16: 5-min 3CVT morning/evening Kendall correlations

4.2.2 Psychomotor Vigilance Task

PVT Daily Trend Analyses

Considering the "ALL" category as the most significant, both mean of the reciprocal of the reaction times, including and excluding lapses, showed a negative trend, indicating a clear performance decline from morning to evening, while remaining relatively stable until the lunch session. In both cases, there was a decrease of around

150 1/ms between morning and evening. Similarly, lapses follow this negative trend, although a temporary improvement in performance was observed during the lunch session. When analyzing individual subjects, mean of the reciprocal of the reaction times, both with and without lapses, were the only parameters consistently showing a negative trend between morning and evening (with lunchtime not being consistent for all participants). In contrast, other parameters, such as standard deviations and invalid starts, did not follow this trend consistently and instead exhibited fluctuating or even opposite patterns across different participants.

ALL	MORNING	LUNCH	EVENING
Lapses(%)	6,470	5,812	7,113
Inv. Starts	1,480	0,739	1,360
Mean No Lapses (1/s)	3,024	3,020	2,873
Std No Lapses (1/s)	0,521	0,460	0,439
Mean Lapses (1/s)	2,947	2,946	2,794
Std Lapses (1/s)	0,584	0,537	0,518
BB (score)	55,174	40,273	29,636

Table 4.17: 5-min PVT daily trend for ALL subjects

M1	MORNING	LUNCH	EVENING
Lapses(%)	2,860	0,000	7,313
Inv. Starts	0,800	0,500	1,200
Mean No Lapses (1/s)	3,147	3,034	2,918
Std No Lapses (1/s)	0,449	0,371	0,358
Mean Lapses (1/s)	3,111	3,034	2,829
Std Lapses (1/s)	0,489	0,371	0,475
BB (score)	42,400	32,250	27,200

Table 4.18: 5-min PVT daily trend for subject M1

M2	MORNING	LUNCH	EVENING
Lapses(%)	5,418	0,931	2,506
Inv. Starts	1,4	0,4	1,4
Mean No Lapses (1/s)	3,088	3,105	2,875
Std No Lapses (1/s)	0,384	0,407	0,405
Mean Lapses (1/s)	3,007	3,090	2,850
Std Lapses (1/s)	0,515	0,439	0,430
BB (score)	51,600	31,000	22,600

Table 4.19: 5-min PVT daily trend for subject M2

M3	MORNING	LUNCH	EVENING
Lapses(%)	2,771	2,883	5,993
Inv. Starts	1,2	0,2	0,8
Mean No Lapses (1/s)	3,220	3,303	3,093
Std No Lapses (1/s)	0,461	0,484	0,470
Mean Lapses (1/s)	3,180	3,262	3,014
Std Lapses (1/s)	0,515	0,523	0,557
BB (score)	58,600	41,600	26,600

Table 4.20: 5-min PVT daily trend for subject M3

F1	MORNING	LUNCH	EVENING
Lapses(%)	6,086	5,983	4,406
Inv. Starts	1,800	0,750	1,200
Mean No Lapses (1/s)	3,002	2,879	2,849
Std No Lapses (1/s)	0,431	0,470	0,562
Mean Lapses (1/s)	2,914	2,805	2,803
Std Lapses (1/s)	0,500	0,547	0,589
BB (score)	65,600	50,500	37,200

Table 4.21: 5-min PVT daily trend for subject F1

F2	MORNING	LUNCH	EVENING
Lapses(%)	14,006	18,755	15,746
Inv. Starts	1,800	1,800	2,000
Mean No Lapses (1/s)	2,680	2,743	2,668
Std No Lapses (1/s)	0,761	0,503	0,382
Mean Lapses (1/s)	2,541	2,509	2,515
Std Lapses (1/s)	0,816	0,695	0,518
BB (score)	54,000	46,500	42,000

Table 4.22: 5-min PVT daily trend for subject F2

PVT Correlation strength

The performance correlation analyses were conducted similarly to those for the 3CVT, calculating Spearman's rho and Kendall's tau coefficients, both including and excluding the lunch session. The "ALL" category generally exhibited weak correlation strength across most parameters, except for the standard deviations of the reciprocal of the reaction times, both with and without lapses. Specifically, Spearman's coefficient exceeded the threshold of +0.3 when the standard protocol (morning/lunch/evening) was considered, indicating a moderate positive

correlation between these two parameters and the BB score. However, in the morning/lunch/evening configuration, the variability of the coefficients among individual subjects was higher, leading to less consistent overall results. Moreover, removing the lunch session for the combined dataset suggested that greater overall coherence between the Spearman and Kendall correlations could be achieved, despite a slight worsening in terms of strength, improving the robustness of the correlation analysis.

	M1	M2	M3	F1	F2	ALL
Lapses(%) - BB	-0,226	0,263	-0,336	-0,145	-0,435	0,190
Inv. Starts - BB	-0,097	-0,175	0,106	0,052	0,220	0,157
Mean No Lapses (1/s) - BB	0,584	0,347	0,483	0,299	0,217	0,043
Std No Lapses (1/s) - BB	0,770	-0,234	0,326	0,365	0,100	0,321
Mean Lapses (1/s) - BB	0,538	0,211	0,490	0,339	0,283	-0,003
Std Lapses (1/s) - BB	0,262	0,365	0,181	0,139	-0,200	0,382

Table 4.23: 5-min PVT morning/lunch/evening Spearman correlations

	M1	M2	M3	F1	F2	ALL
Lapses(%) - BB	-0,166	0,183	-0,254	-0,079	-0,366	0,120
Inv. Starts - BB	-0,093	-0,149	0,074	0,065	0,189	0,121
Mean No Lapses (1/s) - BB	0,402	0,279	0,325	0,221	0,222	0,040
Std No Lapses (1/s) - BB	0,603	-0,164	0,232	0,256	0,167	0,217
Mean Lapses (1/s) - BB	0,402	0,164	0,345	0,243	0,278	0,005
Std Lapses (1/s) - BB	0,246	0,221	0,115	0,066	-0,085	0,255

Table 4.24: 5-min PVT morning/lunch/evening Kendall correlations

	M1	M2	M3	F1	F2	ALL
Lapses(%) - BB	-0,173	0,377	-0,582	0,030	-0,200	0,051
Inv. Starts - BB	-0,182	-0,273	0,073	0,151	0,474	0,074
Mean No Lapses (1/s) - BB	0,488	0,340	0,413	0,273	0,000	0,105
Std No Lapses (1/s) - BB	0,817	-0,255	0,274	0,455	0,200	0,216
Mean Lapses (1/s) - BB	0,488	0,152	0,565	0,273	-0,100	0,086
Std Lapses (1/s) - BB	0,280	0,675	0,134	0,297	-0,308	0,329

Table 4.25: 5-min PVT morning/evening Spearman correlations

	M1	M2	M3	F1	F2	ALL
Lapses(%) - BB	-0,141	0,283	-0,477	0,045	-0,200	0,031
Inv. Starts - BB	-0,206	-0,226	0,058	0,158	0,447	0,057
Mean No Lapses (1/s) - BB	0,386	0,270	0,270	0,200	0,200	0,090
Std No Lapses (1/s) - BB	0,705	-0,135	0,159	0,333	0,200	0,148
Mean Lapses (1/s) - BB	0,386	0,135	0,405	0,200	0,000	0,060
Std Lapses (1/s) - BB	0,250	0,494	0,090	0,200	-0,105	0,225

Table 4.26: 5-min PVT morning/evening Kendall correlations

4.2.3 Optimization Phase Results Discussion

Trend analyses and correlation strength for both the PVT and 3CVT evidenced similar results. In both cases, it was still not possible to draw clear and robust conclusions about the correlations between psychovigilance tests and BB. The main limitation was the small data sample available. Even the combined dataset, which aggregated all the collected data, was somewhat limited due to the heterogeneous characteristics of the subjects. The limited number of participants did not allow for the creation of sufficiently large datasets grouped by factors such as gender. These characteristics could be significant in physiological phenomena, such as those involved in Body Battery (BB) evaluations. However, the outcome metrics from both tests showed a deteriorating trend throughout the day, reflecting the typical trend observed in BB. These considerations highlight the necessity of extending data collection to a larger and more diverse population. Simplifying the protocol may also be important to reduce external influencing factors as much as possible. Despite the fact that the current results did not provide a clear or definitive preference between the PVT and 3CVT, the Psychomotor Vigilance Task (PVT) was chosen for the final experimental phase. This choice was based on its greater popularity in fatigue-related research, which offered more flexibility for future comparisons with other studies. Furthermore, the lunch session could be omitted in the final data collection campaign. The results from the optimization phase indicated that this intermediate session did not provide significant additional information and, in some cases, contributed to increasing variability. Its removal often led to more coherent results in both trend and correlation performance analyses.

4.3 Final Experiment Results

The final experimental phase was intended to obtain a larger data sample following all considerations retrieved in previous steps. In particular, fourteen participants took part to this campaign: seven male and seven female. The age range was between twenty and thirty years old, except for two subjects (a man and a woman) whose age was over sixty. The simplified protocol adopted, included PVT autonomous administration on subject's personal computer only two times a day (morning around 10:00 A.M. and evening around 5:00 P.M.) for five days long (not necessarily consecutive). Results were grouped according to some main influencing factor such as gender, age and gender-age combination in order to understand their influence. A slight modification in the PVT application was performed in order to introduce some additional parameter that could give more insight about performance variability throughout the day. Specifically, the 10% fastest and 10% slowest mean reaction time and their relative standard deviations were added. Moreover, alongside the outcomes of the PVT test, also the Karolinska Sleepiness

Scale (KSS) score collected just before each test administration was analysed with respect the BB to have an additional information about perceived sleepiness state. In this final phase, to further validate the obtained results, the corresponding p-value was computed for each correlation coefficient to have a concrete measure of their significance. Specifically, first the p-value was evaluated to highlight in lighter shades of blue values smaller than 0.075. Traditional threshold for p-value was 0.05 to consider the correlation coefficient statistically relevant. However, considering the exploratory and innovative nature of this study, a quite safer limit was chosen to avoid the exclusion of limit values in this preliminary evaluation. The same reasoning was conducted for the correlation coefficient values, whose threshold remains ± 0.3 as explained in the optimization phase results introduction (4.2). Interesting moderate correlations above ± 0.3 were evidenced using a darker shade of blue. Moreover, as the amount of available data samples significantly increased, a parallel analysis was conducted on cleaned datasets. Specifically, the standard datasets previously described were cleaned by removing all the outliers identified in each specific group of data.

4.3.1 Final Results Analyses

All (all data samples combined together)

The combined dataset included all data samples collected during this entire final experimental phase. Three PVT outcome metrics out of ten, along with the KSS score, were below the threshold for statistical relevance, being also consistent between both Spearman and Kendall analyses. However, no correlation coefficient exceeded the ± 0.3 threshold selected for correlation strength, with an overall low average. This suggested that dividing the general dataset based on factors such as gender, age, or their combination might provide more insights.

	rho	p value
Lapses(%) - BB	-0,223	0,009
Inv. Starts - BB	0,080	0,357
Mean No Lapses (1/s) - BB	-0,034	0,693
Std No Lapses (1/s) - BB	-0,031	0,718
Mean Lapses (1/s) - BB	-0,012	0,886
Std Lapses (1/s) - BB	-0,208	0,015
10% fastest - BB	-0,017	0,843
std 10% fastest - BB	0,010	0,904
10% slowest - BB	0,104	0,228
std 10% slowest - BB	-0,159	0,065
KKS - BB	-0,287	0,001

Table 4.27: Raw data Spearman correlations for "all" dataset

	tau	p value
Lapses(%) - BB	-0,158	0,011
Inv. Starts - BB	0,060	0,341
Mean No Lapses (1/s) - BB	-0,018	0,760
Std No Lapses (1/s) - BB	-0,012	0,845
Mean Lapses (1/s) - BB	-0,001	0,982
Std Lapses (1/s) - BB	-0,137	0,019
10% fastest - BB	-0,003	0,958
std 10% fastest - BB	0,007	0,907
10% slowest - BB	0,077	0,189
std 10% slowest - BB	-0,109	0,063
KKS - BB	-0,226	0,000

Table 4.28: Raw data Kendall correlations for "all" dataset

	rho	p value
Lapses(%) - BB	-0,219	0,014
Inv. Starts - BB	0,068	0,446
Mean No Lapses (1/s) - BB	-0,034	0,693
Std No Lapses (1/s) - BB	0,007	0,935
Mean Lapses (1/s) - BB	-0,012	0,886
Std Lapses (1/s) - BB	-0,191	0,032
10% fastest - BB	0,013	0,881
std 10% fastest - BB	-0,008	0,932
10% slowest - BB	0,101	0,243
std 10% slowest - BB	-0,183	0,033
KKS - BB	-0,287	0,001

Table 4.29: Cleaned data Spearman correlations for "all" dataset

	tau	p value
Lapses(%) - BB	-0,161	0,014
Inv. Starts - BB	0,053	0,427
Mean No Lapses (1/s) - BB	-0,018	0,760
Std No Lapses (1/s) - BB	0,015	0,811
Mean Lapses (1/s) - BB	-0,001	0,982
Std Lapses (1/s) - BB	-0,126	0,038
10% fastest - BB	0,015	0,797
std 10% fastest - BB	-0,004	0,953
10% slowest - BB	0,075	0,200
std 10% slowest - BB	-0,124	0,034
KKS - BB	-0,226	0,000

Table 4.30: Cleaned data Kendall correlations for "all" dataset

Female

The p-value evaluation revealed statistical significance for several parameters, with lapses percentage as the most interesting. This outcome of the PVT was the only one to achieve also a moderate correlation strength. However, even if Spearman's rho was over -0.3 threshold, it was not the same for the Kendall's tau, that was only close but below this limit. Furthermore, its statistical significance was the highest in this dataset, reinforcing its importance. Although there were also other statistically significant outcome metrics, they didn't exceed the selected threshold.

In general, correlation coefficients of these parameters tended to be lower for the Kendall's tau with respect to the Spearman's rho.

	rho	p value
Lapses(%) - BB	-0,350	0,003
Inv. Starts - BB	0,196	0,104
Mean No Lapses (1/s) - BB	0,061	0,618
Std No Lapses (1/s) - BB	0,033	0,785
Mean Lapses (1/s) - BB	0,103	0,394
Std Lapses (1/s) - BB	-0,246	0,040
10% fastest - BB	0,075	0,539
std 10% fastest - BB	0,009	0,940
10% slowest - BB	0,282	0,018
std 10% slowest - BB	-0,260	0,030
KKS - BB	-0,293	0,014

Table 4.31: Raw data Spearman correlations for "female" dataset

	tau	p value
Lapses(%) - BB	-0,241	0,005
Inv. Starts - BB	0,145	0,103
Mean No Lapses (1/s) - BB	0,057	0,490
Std No Lapses (1/s) - BB	0,038	0,652
Mean Lapses (1/s) - BB	0,088	0,285
Std Lapses (1/s) - BB	-0,167	0,042
10% fastest - BB	0,071	0,391
std 10% fastest - BB	0,005	0,960
10% slowest - BB	0,185	0,025
std 10% slowest - BB	-0,177	0,032
KKS - BB	-0,234	0,009

Table 4.32: Raw data Kendall correlations for "female" dataset

	rho	p value
Lapses(%) - BB	-0,394	0,001
Inv. Starts - BB	0,112	0,379
Mean No Lapses (1/s) - BB	0,061	0,618
Std No Lapses (1/s) - BB	0,033	0,796
Mean Lapses (1/s) - BB	0,103	0,394
Std Lapses (1/s) - BB	-0,247	0,047
10% fastest - BB	0,064	0,605
std 10% fastest - BB	-0,087	0,508
10% slowest - BB	0,282	0,018
std 10% slowest - BB	-0,260	0,030
KKS - BB	-0,293	0,014

Table 4.33: Cleaned data Spearman correlations for "female" dataset

	tau	p value
Lapses(%) - BB	-0,277	0,002
Inv. Starts - BB	0,083	0,383
Mean No Lapses (1/s) - BB	0,057	0,490
Std No Lapses (1/s) - BB	0,038	0,659
Mean Lapses (1/s) - BB	0,088	0,285
Std Lapses (1/s) - BB	-0,163	0,057
10% fastest - BB	0,062	0,458
std 10% fastest - BB	-0,059	0,515
10% slowest - BB	0,185	0,025
std 10% slowest - BB	-0,177	0,032
KKS - BB	-0,234	0,009

Table 4.34: Cleaned data Kendall correlations for "female" dataset

Male

The male subgroup showed bad results in general. All test outcomes revealed both low statistical significance and low correlation strength. The only parameter providing some additional insight was the Karolinska Sleepiness Scale (KSS) score, which showed good p-values for both Sperman and Kendall, although it slightly exceeded the moderate threshold only for the rho coefficient in the raw dataset. Overall, even after cleaning the dataset, the results did not improved significantly.

	rho	p value
Lapses(%) - BB	-0,051	0,686
Inv. Starts - BB	-0,046	0,711
Mean No Lapses (1/s) - BB	-0,043	0,729
Std No Lapses (1/s) - BB	-0,076	0,546
Mean Lapses (1/s) - BB	-0,037	0,766
Std Lapses (1/s) - BB	-0,187	0,132
10% fastest - BB	-0,042	0,740
std 10% fastest - BB	-0,002	0,987
10% slowest - BB	-0,102	0,417
std 10% slowest - BB	-0,052	0,676
KKS - BB	-0,307	0,012

Table 4.35: Raw data Spearman correlations for "male" dataset

	tau	p value
Lapses(%) - BB	-0,043	0,641
Inv. Starts - BB	-0,018	0,850
Mean No Lapses (1/s) - BB	-0,032	0,711
Std No Lapses (1/s) - BB	-0,036	0,674
Mean Lapses (1/s) - BB	-0,022	0,799
Std Lapses (1/s) - BB	-0,111	0,193
10% fastest - BB	-0,026	0,761
std 10% fastest - BB	-0,002	0,982
10% slowest - BB	-0,064	0,455
std 10% slowest - BB	-0,035	0,686
KKS - BB	-0,244	0,010

Table 4.36: Raw data Kendall correlations for "male" dataset

	rho	p value
Lapses(%) - BB	0,069	0,594
Inv. Starts - BB	0,023	0,859
Mean No Lapses (1/s) - BB	0,118	0,362
Std No Lapses (1/s) - BB	-0,001	0,992
Mean Lapses (1/s) - BB	0,080	0,535
Std Lapses (1/s) - BB	-0,153	0,227
10% fastest - BB	0,112	0,386
std 10% fastest - BB	0,042	0,745
10% slowest - BB	-0,102	0,417
std 10% slowest - BB	-0,098	0,438
KKS - BB	-0,235	0,073

Table 4.37: Cleaned data Spearman correlations for "male" dataset

	tau	p value
Lapses(%) - BB	0,046	0,632
Inv. Starts - BB	0,033	0,733
Mean No Lapses (1/s) - BB	0,071	0,422
Std No Lapses (1/s) - BB	0,017	0,851
Mean Lapses (1/s) - BB	0,053	0,545
Std Lapses (1/s) - BB	-0,088	0,310
10% fastest - BB	0,073	0,408
std 10% fastest - BB	0,028	0,754
10% slowest - BB	-0,064	0,455
std 10% slowest - BB	-0,064	0,455
KKS - BB	-0,189	0,069

Table 4.38: Cleaned data Kendall correlations for "male" dataset

Male Young (20-30 years old)

The subgroup composed of young male participants (20-30 years old) revealed some improvement with respect to the general male group. In this case, only one PVT outcome, alongside the KSS score, met the p-value significance threshold of 0.075. However, the standard deviation of the reciprocal of the reaction times excluding lapses and the mean value of reciprocal of the 10% fastest reaction times also approached significance. They achieved values close to the selected threshold in the raw dataset for both Spearman and Kendall correlation analyses. Overall statistical

significance of all parameters was improved compared to the full general male group, suggesting that this age-gender combination can lead to valuable insights. Regarding the correlation strength, the raw data Spearman's rho coefficient of the standard deviation of the reciprocal of the reaction times including lapses was the only parameter to reach moderate correlation strength. However, this was not completely reflected also in Kendall's tau, that reached only -0.200. Lastly, the data cleaning process in this case did not produced improvements. In particular, both p-values and correlation coefficients showed a declining trend with respect to the initial raw data analyses.

	rho	p value
Lapses(%) - BB	-0,053	0,697
Inv. Starts - BB	0,044	0,746
Mean No Lapses (1/s) - BB	-0,134	0,326
Std No Lapses (1/s) - BB	-0,231	0,087
Mean Lapses (1/s) - BB	-0,120	0,380
Std Lapses (1/s) - BB	-0,308	0,021
10% fastest - BB	-0,233	0,084
std 10% fastest - BB	-0,097	0,477
10% slowest - BB	-0,121	0,375
std 10% slowest - BB	-0,056	0,682
KKS - BB	-0,428	0,001

Table 4.39: Raw data Spearman correlations for " young male" dataset

	tau	p value
Lapses(%) - BB	-0,046	0,645
Inv. Starts - BB	0,057	0,575
Mean No Lapses (1/s) - BB	-0,089	0,340
Std No Lapses (1/s) - BB	-0,145	0,118
Mean Lapses (1/s) - BB	-0,074	0,428
Std Lapses (1/s) - BB	-0,200	0,031
10% fastest - BB	-0,158	0,090
std 10% fastest - BB	-0,073	0,432
10% slowest - BB	-0,074	0,428
std 10% slowest - BB	-0,035	0,713
KKS - BB	-0,335	0,001

Table 4.40: Raw data Kendall correlations for "young male" dataset

	rho	p value
Lapses(%) - BB	0,045	0,749
Inv. Starts - BB	0,122	0,385
Mean No Lapses (1/s) - BB	-0,001	0,994
Std No Lapses (1/s) - BB	-0,185	0,181
Mean Lapses (1/s) - BB	0,015	0,917
Std Lapses (1/s) - BB	-0,265	0,053
10% fastest - BB	-0,200	0,142
std 10% fastest - BB	-0,023	0,868
10% slowest - BB	-0,121	0,375
std 10% slowest - BB	-0,109	0,428
KKS - BB	-0,365	0,010

Table 4.41: Cleaned data Spearman correlations for "young male" dataset

	tau	p value
Lapses(%) - BB	0,029	0,782
Inv. Starts - BB	0,112	0,290
Mean No Lapses (1/s) - BB	-0,004	0,969
Std No Lapses (1/s) - BB	-0,115	0,227
Mean Lapses (1/s) - BB	0,012	0,902
Std Lapses (1/s) - BB	-0,174	0,066
10% fastest - BB	-0,137	0,144
std 10% fastest - BB	-0,024	0,806
10% slowest - BB	-0,074	0,428
std 10% slowest - BB	-0,069	0,463
KKS - BB	-0,287	0,011

Table 4.42: Cleaned data Kendall correlations for "young male" dataset

Female Young (20-30 years old)

The young female subgroup (20-30 years old) showed the best overall results. In particular, there was a general improvement compared to the full general female group, both in p-values and correlation coefficients. Six out of ten PVT outcome metrics, along with the KSS score, reached very high statistically significant values, which remained consistent between both Spearman and Kendall correlation analyses. However, despite the overall Kendall's tau coefficients were close to the ± 0.3 threshold, only the mean value of the 10% fastest reciprocal of the reaction times consistently achieved moderate correlation strength in both Spearman and Kendall analyses. In this case, the data cleaning process led to an improvement in both p-values and correlation coefficients computation. Specifically, this procedure highlighted a second PVT outcome metric, alongside the previously discussed "10% fastest", which was the lapses percentage. Alongside the KSS score, these results revealed moderate and statistically relevant correlations which were consistent between both Spearman and Kendall analyses.

	rho	p value
Lapses(%) - BB	-0,393	0,002
Inv. Starts - BB	0,196	0,133
Mean No Lapses (1/s) - BB	0,396	0,002
Std No Lapses (1/s) - BB	0,306	0,017
Mean Lapses (1/s) - BB	0,410	0,001
Std Lapses (1/s) - BB	-0,035	0,794
10% fastest - BB	0,423	0,001
std 10% fastest - BB	-0,048	0,715
10% slowest - BB	0,377	0,003
std 10% slowest - BB	-0,084	0,525
KKS - BB	-0,586	0,000

Table 4.43: Raw data Spearman correlations for " young female" dataset

	tau	p value
Lapses(%) - BB	-0,277	0,003
Inv. Starts - BB	0,147	0,124
Mean No Lapses (1/s) - BB	0,279	0,002
Std No Lapses (1/s) - BB	0,219	0,015
Mean Lapses (1/s) - BB	0,289	0,001
Std Lapses (1/s) - BB	-0,018	0,848
10% fastest - BB	0,303	0,001
std 10% fastest - BB	-0,035	0,702
10% slowest - BB	0,252	0,005
std 10% slowest - BB	-0,057	0,527
KKS - BB	-0,449	0,000

Table 4.44: Raw data Kendall correlations for "young female" dataset

	rho	p value
Lapses(%) - BB	-0,426	0,001
Inv. Starts - BB	0,122	0,378
Mean No Lapses (1/s) - BB	0,396	0,002
Std No Lapses (1/s) - BB	0,315	0,018
Mean Lapses (1/s) - BB	0,410	0,001
Std Lapses (1/s) - BB	-0,064	0,643
10% fastest - BB	0,449	0,000
std 10% fastest - BB	-0,204	0,151
10% slowest - BB	0,377	0,003
std 10% slowest - BB	-0,084	0,525
KKS - BB	-0,586	0,000

Table 4.45: Cleaned data Spearman correlations for "young female" dataset

	tau	p value
Lapses(%) - BB	-0,303	0,002
Inv. Starts - BB	0,093	0,367
Mean No Lapses (1/s) - BB	0,279	0,002
Std No Lapses (1/s) - BB	0,223	0,016
Mean Lapses (1/s) - BB	0,289	0,001
Std Lapses (1/s) - BB	-0,037	0,695
10% fastest - BB	0,315	0,001
std 10% fastest - BB	-0,136	0,165
10% slowest - BB	0,252	0,005
std 10% slowest - BB	-0,057	0,527
KKS - BB	-0,449	0,000

Table 4.46: Cleaned data Kendall correlations for "young female" dataset

Male-Female Young (20-30 years old)

The young male and female subgroup (20-30 years old) results, reinforced one again the importance of the group division in this analyses. The raw dataset provided no meaningful correlations, while the KSS score remained the only parameter with significant and moderate correlations, consistent throughout all the analyses. However, the overall results were worse compared to the general dataset containing all the data samples, suggesting that this might not be the best split option to improve correlation significance and strength.

	rho	p value
Lapses(%) - BB	-0,191	0,078
Inv. Starts - BB	0,159	0,144
Mean No Lapses (1/s) - BB	0,097	0,372
Std No Lapses (1/s) - BB	-0,029	0,791
Mean Lapses (1/s) - BB	0,095	0,384
Std Lapses (1/s) - BB	-0,148	0,175
10% fastest - BB	0,058	0,596
std 10% fastest - BB	-0,012	0,914
10% slowest - BB	0,090	0,409
std 10% slowest - BB	-0,027	0,803
KKS - BB	-0,452	0,000

Table 4.47: Raw data Spearman correlations for "male and female young" dataset

	tau	p value
Lapses(%) - BB	-0,139	0,083
Inv. Starts - BB	0,131	0,107
Mean No Lapses (1/s) - BB	0,072	0,328
Std No Lapses (1/s) - BB	-0,011	0,884
Mean Lapses (1/s) - BB	0,072	0,334
Std Lapses (1/s) - BB	-0,090	0,224
10% fastest - BB	0,048	0,516
std 10% fastest - BB	-0,011	0,884
10% slowest - BB	0,067	0,367
std 10% slowest - BB	-0,017	0,826
KKS - BB	-0,348	0,000

Table 4.48: Raw data Kendall correlations for "male and female young" dataset

	rho	p value
Lapses(%) - BB	-0,197	0,082
Inv. Starts - BB	0,238	0,031
Mean No Lapses (1/s) - BB	0,136	0,233
Std No Lapses (1/s) - BB	0,003	0,978
Mean Lapses (1/s) - BB	0,145	0,204
Std Lapses (1/s) - BB	-0,122	0,270
10% fastest - BB	0,036	0,752
std 10% fastest - BB	-0,045	0,692
10% slowest - BB	0,079	0,478
std 10% slowest - BB	-0,061	0,579
KKS - BB	-0,452	0,000

Table 4.49: Cleaned data Spearman correlations for "male and female young" dataset

	tau	p value
Lapses(%) - BB	-0,149	0,079
Inv. Starts - BB	0,187	0,025
Mean No Lapses (1/s) - BB	0,093	0,231
Std No Lapses (1/s) - BB	0,011	0,887
Mean Lapses (1/s) - BB	0,102	0,188
Std Lapses (1/s) - BB	-0,074	0,327
10% fastest - BB	0,031	0,690
std 10% fastest - BB	-0,033	0,675
10% slowest - BB	0,059	0,435
std 10% slowest - BB	-0,039	0,606
KKS - BB	-0,348	0,000

Table 4.50: Cleaned data Kendall correlations for "male and female young" dataset

Male-Female adults (over 60 years old)

The adult male and female subgroup (over 60 years old) results were reported despite their low significance. Only two participants met these criteria, leading to a very limited data sample. Results were neither significant nor strong for both Spearman and Kendall analyses. Additionally, the KSS score was useless as both participants reported the same perceived level of sleepiness before all test administration, making impossible the computation of p-value, rho and tau coefficients. However, a larger and more representative dataset could potentially lead to different and more meaningful results.

	rho	p value
Lapses(%) - BB	-0,386	0,093
Inv. Starts - BB	0,072	0,763
Mean No Lapses (1/s) - BB	0,077	0,747
Std No Lapses (1/s) - BB	-0,090	0,705
Mean Lapses (1/s) - BB	0,166	0,485
Std Lapses (1/s) - BB	-0,205	0,385
10% fastest - BB	-0,021	0,930
std 10% fastest - BB	-0,050	0,833
10% slowest - BB	0,217	0,359
std 10% slowest - BB	-0,171	0,471

Table 4.51: Raw data Spearman correlations for "male and female adult" dataset

	tau	p value
Lapses(%) - BB	-0,291	0,100
Inv. Starts - BB	0,041	0,841
Mean No Lapses (1/s) - BB	0,048	0,795
Std No Lapses (1/s) - BB	-0,032	0,871
Mean Lapses (1/s) - BB	0,123	0,474
Std Lapses (1/s) - BB	-0,134	0,434
10% fastest - BB	0,005	1,000
std 10% fastest - BB	0,011	0,974
10% slowest - BB	0,139	0,416
std 10% slowest - BB	-0,144	0,398

Table 4.52: Raw data Kendall correlations for "male and female adult" dataset

	rho	p value
Lapses(%) - BB	-0,386	0,093
Inv. Starts - BB	0,067	0,792
Mean No Lapses (1/s) - BB	0,077	0,747
Std No Lapses (1/s) - BB	0,056	0,819
Mean Lapses (1/s) - BB	0,166	0,485
Std Lapses (1/s) - BB	-0,205	0,385
10% fastest - BB	-0,021	0,930
std 10% fastest - BB	0,107	0,662
10% slowest - BB	0,217	0,359
std 10% slowest - BB	-0,171	0,471

Table 4.53: Cleaned data Spearman correlations for "male and female adult" dataset

	tau	p value
Lapses(%) - BB	-0,291	0,100
Inv. Starts - BB	0,029	0,906
Mean No Lapses (1/s) - BB	0,048	0,795
Std No Lapses (1/s) - BB	0,071	0,699
Mean Lapses (1/s) - BB	0,123	0,474
Std Lapses (1/s) - BB	-0,134	0,434
10% fastest - BB	0,005	1,000
std 10% fastest - BB	0,119	0,505
10% slowest - BB	0,139	0,416
std 10% slowest - BB	-0,144	0,398

Table 4.54: Cleaned data Kendall correlations for "male and female adult" dataset

4.3.2 Final Results Discussion

These final examinations provided a general overview of the potential correlation between PVT outcome metrics and the BB. The main result was clearly the importance of the different influencing factor, such as gender and age. Despite not all the subgroup divisions led to clear improvements, it was important to highlight how they affected both correlation significance and strength. The most promising group division was the one based on gender. Leading to moderate and significant correlation for a great number of PVT outcomes. Clearly the limited number of participants provided some limitations. However, these results contributed to create a solid and tested procedure for future investigations with a larger population.

A larger data sample could give the possibility to further reinforce the moderate correlation strength results obtained for some subgroups in this study. Moreover, it might give the possibility to extend the statistical significance to more parameters, providing additional insights beyond correlation strength.

Chapter 5

Conclusion

In this study, a preliminary evaluation of the possible relationship between vigilance tests and Garmin's Body Battery (BB) was conducted. The goal was to give evidence of potential monotonic correlation between them, providing a solid foundation for future estimation of the BB through a vigilance task. This could open to new applications of the vigilance tests, as they could represent a valid tool to estimate the starting point of the energy levels when the smartwatch is not warn for prolonged periods.

5.1 Interpretation of the Results

The pilot study results settled the basis for the final phase deeper analysis. Specifically, once preliminary phase results showed that 5-minute version of the tests were the best trade-off between amount of data samples and time on task, the pilot study began, to obtain a definitive fine-tuned protocol for our purposes. Due to the limited size of the data sample, these intermediate results did not lead to a clear and strong correlation between vigilance tests and BB. However, the deteriorating trend throughout the day similar to the BB supported the decision to go deeper in this relationship with the final phase. In particular, Psychomotor Vigilance Task (PVT) was selected for the final phase correlation analyses, despite both tests showing comparable results, due to its greater popularity in fatigue and sleep-related research, offering more flexibility for future comparisons. The protocol simplification adopted in the final step, gave the opportunity to reduce the overload influence that could have slightly impacted previous phases. Moreover, a larger data sample allowed to obtain more representative and robust conclusions. Specifically, PVT and BB correlation results from this final phase, evidenced the importance of various influencing factors, such as gender and age. The comparison between the combined dataset containing all data sample from all the participants

of the experiment and the single subgroup division analyses, evidenced a clear improvement in this second configuration. The most promising subgroup division was the gender based. In particular, female group reached moderate and significant correlation for a great number of PVT outcomes, suggesting that the key for further investigation in this relationship could be this influencing factor based subgroup classification.

5.2 Limitations of the Study and Implication for Future Research

As repeatedly stated, the main limitation of this study is represented by the limited number of subjects that took part to the experiment. Having a more representative and sufficiently larger number of participants able to reflect the general population, could lead to a further reinforcement of the already obtained correlations or to the extension of the statistical significance to more parameters. Moreover, a greater data sample would give the possibility to identify a wider number of influencing factors that could not be considered in this exploratory study. Adopting a more restrictive and controlled protocol, based on the administration of the PVT in a controlled environment, could further increase robustness and validity of the correlation analysis. Ultimately, future investigations could rely on the foundations offered by this exploratory research, extending the analysis to a wider population or investigating alternative correlation perspectives that were not considered in this study.

Appendix A

Karolinska Sleepiness Scale (KSS)

Karolinska Sleepiness Scale is a scale to asses subjective level of sleepiness. It reflects subject's feelings in the previous 10 minutes. It associates a score, from 1 to 10, to the perceived level of sleepiness. 1 corresponds to "extremely alert" while 10 corresponds to "Extremely sleepy, can't keep awake".

Value	**Description**
1	Extremely alert
2	Very alert
3	Alert
4	Rather alert
5	Neither alert nor sleepy
6	Some signs of sleepiness
7	Sleepy, but no effort to keep awake
8	Sleepy, but some effort to keep awake
9	Very sleepy, great effort to keep awake, fighting sleep
10	Extremely sleepy, can't keep awake

Table A.1: Karolinska Sleepiness Scale (KSS). Source: Shahid et al., (2012) [7]

Bibliography

[1] Garmin. *Body Battery Energy Monitoring*. `https://www.garmin.com/en-US/garmin-technology/health-science/body-battery/`. n.d. (Cit. on pp. 1, 9).

[2] Lisa Whitehead. «The Measurement of Fatigue in Chronic Illness: A Systematic Review of Unidimensional and Multidimensional Fatigue Measures». In: *Journal of Pain and Symptom Management* 37 (Jan. 2009), pp. 107–128 (cit. on p. 3).

[3] Julian Lim and David F. Dinges. «*Sleep Deprivation and Vigilant Attention*». In: *Annals of the New York Academy of Sciences* 1129.1 (May 2008), pp. 305–322 (cit. on p. 3).

[4] Zhiqiang Ni, Fangmin Sun, and Ye Li. «Heart Rate Variability-Based Subjective Physical Fatigue Assessment». In: *Sensors* 22.9 (Apr. 2022), p. 3199 (cit. on p. 4).

[5] A. R. Burton, K. Rahman, Y. Kadota, A. Lloyd, and U. Vollmer-Conna. «Reduced heart rate variability predicts poor sleep quality in a case–control study of chronic fatigue syndrome». In: *Experimental Brain Research* 204.1 (July 2010), pp. 71–78 (cit. on p. 4).

[6] Anna Åkerstedt Miley, Göran Kecklund, and Torbjörn Åkerstedt. «Comparing two versions of the Karolinska Sleepiness Scale (KSS)». en. In: *Sleep and Biological Rhythms* 14.3 (July 2016), pp. 257–260 (cit. on p. 4).

[7] Azmeh Shahid, Kate Wilkinson, Shai Marcu, and Colin M. Shapiro. «Karolinska Sleepiness Scale (KSS)». In: *STOP, THAT and One Hundred Other Sleep Scales*. Ed. by Azmeh Shahid, Kate Wilkinson, Shai Marcu, and Colin M Shapiro. New York, NY: Springer New York, 2011, pp. 209–210 (cit. on pp. 4, 59).

[8] Jaques Reifman, Kamal Kumar, Maxim Y. Khitrov, Jianbo Liu, and Sridhar Ramakrishnan. «PC-PVT 2.0: An updated platform for psychomotor vigilance task testing, analysis, prediction, and visualization». In: *Journal of Neuroscience Methods* 304 (2018), pp. 39–45 (cit. on pp. 4, 5, 7).

[9] David F. Dinges and Jeffrey W. Powell. «Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations». In: *Behavior Research Methods, Instruments, & Computers* 17.6 (Dec. 1985), pp. 652–655 (cit. on p. 4).

[10] Raymond W. Matthews, Sally A. Ferguson, Charli Sargent, Xuan Zhou, Anastasi Kosmadopoulos, and Gregory D. Roach. «Using interstimulus interval to maximise sensitivity of the Psychomotor Vigilance Test to fatigue.» In: *Accident; analysis and prevention* Pt B (Feb. 2017), pp. 406–410 (cit. on pp. 4, 6).

[11] Mathias Basner and David F. Dinges. «Maximizing sensitivity of the psychomotor vigilance test (PVT) to sleep loss.» In: *Sleep* 34.5 (May 2011), pp. 581–591 (cit. on pp. 5, 6, 20).

[12] Janine Thomann, Christian R. Baumann, Hans-Peter Landolt, and Esther Werth. «Psychomotor vigilance task demonstrates impaired vigilance in disorders with excessive daytime sleepiness.» In: *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine* 10.9 (), pp. 1019–1024 (cit. on p. 5).

[13] Clare Anderson and James Horne. «Sleepiness Enhances Distraction During a Monotonous Task». In: *Sleep* 29 (May 2006), pp. 573–6 (cit. on p. 5).

[14] In-Soo Lee, Wayne A. Bardwell, Sonia Ancoli-Israel, and Joel E. Dimsdale. «Number of lapses during the psychomotor vigilance task as an objective measure of fatigue.» In: *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine* 6 (Apr. 2010), pp. 163–168 (cit. on p. 6).

[15] Al-Baraa Abdulrahman Al-Mekhlafi, Ahmad Shahrul Nizam Isha, Maged S. Al-Quraishi, and Noreen Kanwal. «Implementation of a psychomotor vigilance test to investigate the effects of driving fatigue on oil and gas truck drivers' performance». In: *Frontiers in Public Health* 11 (Oct. 2023), p. 1160317 (cit. on p. 6).

[16] Sibylle Benderoth, Hans-Jürgen Hörmann, Caroline Schießl, and Eva-Maria Elmenhorst. «Reliability and validity of a 3-min psychomotor vigilance task in assessing sensitivity to sleep loss and alcohol: fitness for duty in aviation and transportation.» In: *Sleep* 44.11 (Nov. 2021), zsab151 (cit. on p. 6).

[17] Sylvia Loh, Nicole Lamond, Jill Dorrian, Gregory Roach, and Drew Dawson. «The validity of psychomotor vigilance tasks of less than 10-minute duration.» In: *Behavior research methods, instruments, & computers : a journal of the Psychonomic Society, Inc* 36.2 (May 2004), pp. 339–346 (cit. on p. 6).

[18] Mathias Basner, Tyler M. Moore, Jad Nasrini, Ruben C. Gur, and David F. Dinges. «Response speed measurements on the psychomotor vigilance test: how precise is precise enough?» In: *Sleep* 44.1 (Jan. 2021) (cit. on p. 6).

[19] Maxim Y. Khitrov, Srinivas Laxminarayan, David Thorsley, Sridhar Ramakrishnan, Srinivasan Rajaraman, Nancy J. Wesensten, and Jaques Reifman. «PC-PVT: a platform for psychomotor vigilance task testing, analysis, and prediction.» In: *Behavior research methods* 46.1 (Mar. 2014), pp. 140–147 (cit. on p. 7).

[20] Lucia Arsintescu, Jeffrey B. Mulligan, and Erin E. Flynn-Evans. «Evaluation of a Psychomotor Vigilance Task for Touch Screen Devices.» In: *Human factors* 59.4 (June 2017), pp. 661–670 (cit. on p. 7).

[21] Matthew Kay, Kyle Rector, Sunny Consolvo, Ben Greenstein, Jacob O. Wobbrock, Nathaniel F. Watson, and Julie A. Kientz. «PVT-touch: Adapting a reaction time test for touchscreen devices». In: *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. 2013, pp. 248–251 (cit. on p. 7).

[22] Lucia Arsintescu, Kenji H. Kato, Patrick F. Cravalho, Nathan H. Feick, Leland S. Stone, and Erin E. Flynn-Evans. «Validation of a touchscreen psychomotor vigilance task.» In: *Accident; analysis and prevention* 126 (May 2019), pp. 173–176 (cit. on p. 7).

[23] Chris Berka, Daniel J. Levendowski, Philip Westbrook, Gene Davis, Michelle N. Lumicao, Richard E. Olmstead, Miodrag Popovic, Vladimir T. Zivkovic, and Caitlin K. Ramsey. «EEG quantification of alertness: methods for early identification of individuals most susceptible to sleep deprivation». In: ed. by John A. Caldwell and Nancy Jo Wesensten. Orlando, Florida, USA, May 2005, p. 78 (cit. on pp. 8, 9).

[24] Shani Waninger, Chris Berka, Amir Meghdadi, Marija S. Karic, Kimberly Stevens, Cinthya Aguero, Tatiana Sitnikova, David H. Salat, and Ajay Verma. «Event-related potentials during sustained attention and memory tasks: Utility as biomarkers for mild cognitive impairment». In: *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 10.1 (Jan. 2018), pp. 452–460 (cit. on p. 8).

[25] Maja Stikic, Robin R. Johnson, Daniel J. Levendowski, Djordje P. Popovic, Richard E. Olmstead, and Chris Berka. «Eeg-Derived Estimators of Present and Future Cognitive Performance». In: *Frontiers in Human Neuroscience* 5 (2011) (cit. on p. 9).

[26] Garmin. *Stress Tracking Technology*. `https://www.garmin.com/en-US/garmin-technology/health-science/stress-tracking/`. n.d. (Cit. on p. 9).