MASTER

Pervasive computing for early burnout detection

Dornostup, Y.

Award date:
2014

Link to publication
Pervasive computing for early burnout detection

Master Thesis

(Public version)

Yulia Dornostup

Supervisors:

Dr. Natalia Sidorova
Dr. Joyce Westerink

2014
Abstract

Burnout is a psychological term for the experience of a long-term mental exhaustion, physical fatigue, diminished interest, sleep and concentration problems that can occur irrespective of the type of profession. It is a serious problem affecting employees and employers worldwide. Therefore, there are a lot of researches investigating why people have a burnout syndrome, as well as the consequences of this phenomenon. At this moment, there has not been revealed any good potential biomarkers of burnout.

This work is a part of the big project aimed at finding a biomarker that allows predicting upcoming burnout. It is assumed that it is more likely to find a correlation between the intrapersonal changes of physical condition and a burnout level, rather than to find a biomarker of burnout following the way of the existing studies – by comparing snapshots of physiological measurements of different persons. In particular, the hypothesis is that a burnout level correlates with a heart rate dynamics, by analogy with overtraining that has a close resemblance to burnout.

The validation of the hypothesis requires a big user study, which should be carefully planned to avoid potential problems. This work is the preliminary study aimed at investigation of the application of pervasive technologies for validation of the hypothesis, and understanding how to set up the big user study.

In the thesis, the different kinds of the heart rate monitors were analysed according to the defined specific requirements of the project, and three small user studies were conducted allowing to reveal the potential risks of the big study and find the mitigation strategies.

In the first user study, reliability of several heart rate monitors was estimated by comparing their measurements with the measurements of a clinically validated heart rate monitor considered as ground truth. For the analysis of the results of this study, an algorithm for synchronisation of time series of two monitors was developed. The second study investigated the user behaviour with respect to using different devices for data recording. The software application for results visualizing in a form convenient for the analysis was created. The third user study was aimed at testing the settings for the big user study. The algorithm for automated discovery of the moments when a person falls asleep and wakes up was successfully developed. Additionally, the results of the study revealed possible problems that can appear in the big study.

The results of all three user study are accumulated and the recommendations for the big user study planning are provided.
Contents

1 Introduction ......................................................... 7
2 Problem definition and background information ........ 9
3 Heart rate monitoring devices .................................. 12
   3.1 Heart rate monitoring devices attributes and their impact on the study .......... 12
   3.2 Chosen devices .............................................. 15
4 Structure of the user studies ................................... 15
5 User study 1: Device comparison ......................... 18
   5.1 Objectives and test procedure .......................... 18
   5.2 Classes of the devices .................................... 18
   5.3 The second class of the devices ....................... 18
   5.4 The third class of the devices ......................... 18
   5.5 Discussion .................................................. 18
6 User study 2: iPod vs iPhone ................................. 19
   6.1 Objectives and test procedure .......................... 19
   6.2 Software application for data visualisation ........ 19
   6.3 Data analysis ............................................... 19
   6.4 Discussion .................................................. 19
7 User study 3: Pilot study ....................................... 35
   7.1 Objectives and test procedure .......................... 35
   7.2 Boundaries of a night ..................................... 35
   7.3 The amount of the obtained data ...................... 35
   7.4 Discussion .................................................. 35
8 General discussion and recommendations ............ 45
   8.1 Technology ................................................... 45
   8.2 People ......................................................... 45
   8.3 Time ............................................................ 45
9 Conclusion ......................................................... 47
Appendix A .......................................................... Error! Bookmark not defined.
Bibliography ......................................................... 48
1 Introduction

Burnout is a stress state characterized by symptoms of mental exhaustion and physical fatigue, detachment from work, and feelings of diminished competence [1]. It is a serious problem affecting employees and employers worldwide. According to the bureau “Statistics Netherlands”, the percentage of employees who indicated that they had burnout symptoms has increased in recent years from 11 percent in 2007 to 13 percent in 2011. The cost of burnout is high for both workers and employers. Maslach et al. point out to the fact that burnout leads in the first place to higher costs and financial losses because of higher absenteeism rates and more frequent sick leaves [2].

There are a lot of papers investigating the reasons for people burnout, as well as the consequences of this phenomenon. The results of the existing researches contradict to each other. Some of them found some correlation of burnout with some physiological measurements, while the others assert that there is no good potential biomarker of burnout. In our opinion, the main reason of that is comparing of measurements of burnout patients with that of controls, instead of comparing physical conditions of the same person when he has a burnout syndrome, and when he does not, because the physical conditions of different persons vary a lot regardless to burnout. The idea is that the changes in some physiological measurements of the same person may indicate the changes of his overall physical condition and may be used for predicting upcoming burnout.

Pervasive technologies are used in various spheres of human activity and become an integral part of a daily life. In particular, pervasive technologies are applied in healthcare and include smart homes, mobile and ubiquitous telemedicine to support medical diagnosis, treatment and patient care especially in rural areas, pervasive patient monitoring services, intelligent emergency monitoring, health aware mobile devices, pervasive life style management, and medical inventory management systems [1]. Therefore, pervasive sensor technologies can be used for unobtrusive monitoring of intra-individual changes in repetitive, daily measurements of the relevant parameters in order to find a biomarker of burnout.

This work is a part of the big project aimed at finding a biomarker that allows predicting upcoming burnout; in particular, a heart rate dynamic seems to be interesting in that context. It is focused on the investigation how pervasive computing can help in early diagnostic of burnout, as well as on planning a large user study aimed at the validation of some hypotheses on possible biomarkers of burnout. The big study is expensive and there are many decisions that should be made:

- Which technology should be used?
- How many participants should be involved in order to get a desired amount of data?
- How long the study should be?
- Which physical parameter should be measured? When? How often?
These choices should be made in a smart way; therefore we performed three small user studies for the estimating different aspects, risks and problems that can appear in the large study.

First of all, it appears that there is a variety of different kinds of the devises that can monitor heart rate. In order to choose the most appropriate one(s) for the project we have conducted a user study. In that study we have compared the accuracy of the measurements of several prototype devices and heart rate monitors from mass market with a medical device Nexus-10, which is accurate but obtrusive. The MIO Alpha wristband appears the most accurate heart rate monitor among others.

The main disadvantage of MIO is that it does not store the measurements and, thus, requires an additional device for receiving and storing of data. We considered two options of the receiver: a personal mobile phone and an additional device. We have conducted the second user study in order to estimate the difference of these two options in respect to users’ behavior.

The third user study was conducted as the pilot study of the big project study in order to estimate all potential risks. We have tested the preliminary settings of the big study on a few persons and reveal a lot of potential problems. The results of all these user studies together allowed giving the recommendations to the planning of the big study.

Some technical issues were revealed in the studies. For example, the first study shows that clocks of different devices are not only difficult to synchronize for some initial point of time, but they also run with different speed. Therefore, the algorithm for an automatically synchronization of two heart rate recordings was developed.

Important moments for measurements are moments of falling asleep and waking up. In order to capture these moments, the accelerometer sensor in addition to heart rate monitor was included, and the algorithm for an automatically recognition of the periods when a person is awake and when he is sleeping, was created.

The thesis is structured as follows: in Chapter 2 some background information together with the goal of this work are given; Chapter 3 provides the requirements to the heart rate monitors and the description of heart rate monitors that were considered in the project; in Chapter 4 a brief description of the user studies is given as well as the reasoning of the conducting all of them; Chapter 5, 6 and 7 describe the settings and the analysis of the results of each user study, respectively; finally, Chapter 8 summarizes the results of all the three study and gives recommendations for the big user study.
2 Problem definition and background information

Burnout

According to Maslach et al. [2], burnout is a psychological syndrome that involves a prolonged response to chronic interpersonal stressors on the job. The main symptom of burnout is emotional exhaustion [3], which leads to impaired functioning on the job [4] as well as an increase of sickness absence [5].

Burnout is not included in any psychiatric classification system. In particular, it is not considered as a disorder in the DSM-IV [6], although burnout is specified in the ICD-10 [7] as a "State of vital exhaustion" (Z73.0) under "Problems related to life-management difficulty" (Z73), but not considered as a "disorder" as well. That is why, firstly, mental health care institutions are often confused to diagnose the burnout [8] and, secondly, it is difficult to find statistical figures around the prevalence of burnout. Indeed, the symptoms of the burnout overlap some syndromes of such psychiatric disorders as depression and emotional exhaustion [9], chronic fatigue syndrome, and overtraining. However, Glass shows that burnout and depressive syndrome are closely related, but still are not just two different terms for the same state [10]. It means that physiological measurements used in the diagnostics of these medical conditions can be also relevant to burnout.

Heart activity measurements

Heart rate (HR) is one of the most important physiological measures to monitor. It is expressed in the number of heartbeats per a specific unit of time, usually one minute. Heart rate is not a constant value; it can vary because of changes in the state of body and mind. For instance, during sleep, the heart rate is usually lower than in daytime, while during the physical exercises, the heart rate may increase significantly. It is known that emotional states, for example, stress, influence heart rate as well. There is a baseline of measurements, which is resting heart rate. Resting heart rate is the heart rate measured when the person is at rest but awake, and it is individual for each person. Typically the resting heart rate of healthy adults varies between 60 and 80 beats per minute. However, individuals who exercise and have a good physical condition, such as athletes, can have a lower resting heart rate. Another important physiological measure is the recovery heart rate. It can be measured immediately after an exercise. Normally, the heart rate of a healthy person should drop and this drop should be more than 12 bpm for an adult [11]. Heart rate variability is one of the most popular indicators of a heart condition. It reflects the variation of time between two consecutive heart beats.

Heart rate and burnout

In the study described in [12], 22 burnout patients and 23 healthy controls were compared. The result shows that burnout subjects have a higher resting heart rate than controls on average. This is illustrated in Figure 1. The similar results were obtained in [13], but for the chronic fatigue syndrome; in this study 30 chronic fatigue syndrome patients were compared
with 38 controls, and on average the patients had a higher heart rate and a lower heart rate variability both during wakefulness and sleep in comparison to controls.

Figure 1. Mean heart rate in burnout patients and healthy controls during six phases of the laboratory session. BASE, baseline; PREP, speech task preparation; MA, mental arithmetic; SPEECH, speech task; REC1, first recovery phase; REC2, second recovery phase (the picture is taken from [11]).

However, there are other studies which fail to find significant difference between healthy people and burnout patients. For example, Danhof-Pont gives a systematic review of the studies which compare biomarkers of burnout patients and controls [14] and does not reveal any potential biomarkers for burnout. The reason of such results can be a short study period, a limited number of participants or very big variability in burnout patients’ severity. Indeed, a heart rate, for example, is a very specific health feature and is different for different persons. In fact, even though the difference between the mean heart rate of the people from the two groups in Figure 1 can be clearly seen, it does not mean that all burnout patients had a higher heart rate than a heart rate of anyone from the controls. Therefore, by measuring a heart rate of a random person we cannot decide to which group that person belongs and, thus, a one-off measurement of a heart rate cannot be used for diagnostics of burnout.

The existing literature has mostly focused on limited number of subjects monitored over short periods of time, and the majority of them compared the metrics of burnout patients and healthy controls. However, it is interesting to figure out if there is any difference of the heart rate of the same person when he is suffering from burnout and when he does not. For instance, an overtraining, that shows close resemblance to burnout, can be observed from the changes in HR and HRV. That brings us to the hypothesis that will be the basis of the case study in this work; it is related to the possibility of burnout prediction on basis of heart rate dynamics. In particular, the hypothesis is that a human, who is under the risk of burnout, has a higher heart rate in the morning (just awake), than he has in normal conditions, and this metrics is gradually increasing during the period of a high working load. A low morning heart rate indicates that the rest during the night is enough for recreation, while extremely high or extremely low morning heart rates means health problems.

The hypothesis can be validated in two-phases. This project is the first phase, a preliminary study, which is aimed at understanding of what exactly to measure and when, which
technologies to use to obtain desired data, etc. to build a model, which helps to predict burnout. Within the first phase three small user studies will be conducted. The result of that will be the settings of the real study (second phase), concerning the quality of the results expected from it, available technologies, budget and other limitations.
3 Heart rate monitoring devices

To test the hypothesis, a device for continuous heart rate monitoring is needed. A preliminary research of the market has shown a big variety of options for heart rate monitoring. In order to identify which device or a set of devices is the most suitable for the purposes of our study, we have defined a set of the key attributes of the devices and compare different options according to them.

The choice of the important attributes depends on the aim and restrictions of a particular study or other purpose of use. Particular characteristics of devices can be considered crucial for one study and completely not important for another. Attributes can be divided into groups which are common for most areas and can be further detailed for each purpose of use.

One of such groups is the Data attributes which describes reliability of data obtained with a device, precision of measurements, the way of storing and/or transmitting data, etc. Depending on a purpose of use, it can be necessary to consider the Time attributes of devices: for instance, how long a device can work without recharge, for which period it can store data, how long it can be used, how much time is needed for setup and fixing possible problems, etc. These two groups of the attributes are highly related to each other. Moreover, it can be difficult to refer some attributes to one particular group. For instance, the attribute “Sampling frequency” can be placed into the Time attributes group, because it describes the number of samples in one time unit. However it can be related to the Data attributes as well, because it describes how much data can be obtained during a fixed period of time and how accurately the monitored feature can be captured. That is why the Data attributes and Time attributes groups can be united into Quality attributes group.

If using a device involves people directly, the group of User attributes should be taken into account as well. There such characteristics as obtrusiveness of the device, intuitiveness and even its design can be specified, as well as all potential risks and possible harm.

Another group of device characteristics is the Cost attributes of the device. It can include many factors which form the final price of using a device, for example, the price of the device, cost of shipping, cost for setup, etc.

Different conditions of the studies require different attributes of the devices for comparing, therefore it is always useful to determine the groups of interesting attributes first and then detail them with the particular characteristics of the devices.

3.1 Heart rate monitoring devices attributes and their impact on the study

There is a variety of the devices for heart rate monitoring: for example, wristbands, video cameras, mattresses, chest belts, finger and ear clips, and many others. They use different technologies for heart rate calculation, should be attached to different body parts and have different attributes in technical specification. Therefore, it is not effective to compare the devices using only their technical specification, but the characteristics important for a particular study can be derived from it. The characteristics that are important for our study are listed and described below.
Data attributes

- **Measurement data**
  Some of the devices give heart rate as an output of measurements, while others can give data that have to be processed to derive heart rate values. Additionally, some devices measure not only heart rate, but also such characteristics as heart rate variability, physical activity level, etc.

- **Data storage**
  Since the user studies are supposed to be not in a laboratory but in the conditions of everyday live, the devices have to store the measurements. The capacity of the memory is a factor which can limit the time of continuous measurements. If a device cannot store data (or at least transmit it) it is impossible to get data for analysis and thus the device is useless in our study.

- **Data transmission**
  Some devices do not have internal memory but they transmit data to another device in real time. Data transmission to a mobile phone is very common for heart rate monitors designed for sportsmen. Usually such devices use Bluetooth or ANT+ protocols for data transmitting; and it is required that receivers use the same protocol.

- **Receiver**
  If a device requires a receiver, then the receiver’s characteristics should be considered as well. The receiver can introduce new limitations such as its battery life, its memory size, difficulty of the user interface, cost and many others.

- **Software application**
  A receiver and/or a monitor itself need a software application to handle the transmitted data. That can be an existing application, which is free or costs money, or written specially for a particular study. The application can allow storing raw data, or it can only output postprocessed data, or even only visualisation for the transmitted data.

- **Frequency of sampling**
  Some of devices give information about each heart beat, while another record heart rate each second or once per minute. However, there are some monitors that give one or several values of the heart rate for the whole period of monitoring: for example median, maximum and minimum heart rate that was measured during a monitoring session.

- **Reliability**
  To make the right conclusion about the hypothesis, we need a precise measurement. Although the manufactures of the devices claim some level of precision, the estimation of this level was mainly done in laboratory conditions, and we cannot expect the same quality of data in our studies.

Time attributes

- **Time of continuous measurement**
  We are especially interesting in night and morning heart rate measurements. To catch a moment just before a participant wakes up, we have to measure heart rate during the
whole night (or automatically start measurements at some point at the night time). That is why the time of continuous measurement is a very important characteristic of a device in our study. The value of this attribute depends on the battery lifetime of a device and a receiver device (if applicable) and memory capacity.

- **Time of recharging**
  Many of heart rate monitoring devices have autonomous power supply and only require periodical recharging. We need to take into consideration the time needed for recharging the device, because it may require pausing of monitoring. Time of recharging (if a device can be recharged) can be obtained directly from a technical specification of a device.

- **Measuring conditions**
  The ability of monitoring heart rate during the night time is not a technical characteristic of a device, but this attribute is extremely important for our study. Some of the devices cannot measure heart rate at night due to the lack of light, or for some other reasons, while other devices cannot measure heart rate in daytime, for instance, mattress sensors can conduct heart rate monitoring only while a user is lying in a bed.

**User attributes**

- **Unobtrusiveness**
  Our study involves direct user interaction with a heart rate monitoring device during a long period of time (up to 6 months), therefore the device must be as unobtrusive as possible. It is a subjective attribute which can be estimated taking into account such primary characteristics as body part needed for measuring, battery life which determine how frequent a user has to recharge it, its waterproofness and even the design of the device.

- **Intuitiveness**
  Due to long ambulatory usage of the device it is necessary that installation (meaning both “putting it on” and “turning it on”) and usage of it do not require much efforts and special skills. This attribute can be derived from the information about the usage of the device, the user manual (how easy to use a device, which language is used, how well the usage process described).

- **Type**
  It is important to keep in mind in which way a device interact with the body: which part of the body is needed for heart rate measuring, which conditions are required. This attribute is already included to Unobtrusiveness attribute, but it could be important to consider it separately as well. For instance, it can be helpful while choosing a combination of devices, because it allows easy determining which devices cannot operate together.

**Cost attributes**

- **Cost**
The cost attribute is an important attribute for the burnout study, because the budget is limited, while the number of participants needed for making good conclusion is big. Cost attribute includes the price of the device, costs of shipping, costs of the supporting equipment and a software application if needed.

- **Quantity of devices which we already have**
  It is not a characteristic of a device itself, but we should take into account that some devices are already available and thus do not have to be bought.

It can be clearly seen, that most of the technical characteristics become parts of groups Data and Time attributes.

### 3.2 Chosen devices

*This part is confidential.*
4 Structure of the user studies

The validation of the hypothesis stated in Chapter 1 is performed in the two big phases. The first phase is a preliminary one and it is described in the thesis. The main task for this phase is to find the optimal settings for the “Big burnout study” which is a part of a second phase of the validation, while the big study itself is beyond the bounds of this work.

The preliminary phase of the hypothesis validation is needed for reduction of the risks of the big user study and to insure researchers against some potential faults. Ideally for the validation of the hypothesis we would like to have a 24-hours heart rate track of a number of people who is going to burn out during several months, together with the responses on the MBI questionnaire. However, it is a very difficult task for both volunteers and the researches, besides it costs a lot of money because the heart rate monitors have to be bought. That is why it is extremely important to take into account all aspects which can have an influence on the study. For that we need to answer the following questions:

- How many participants should be involved? (In order to get needed amount of data, taking into account that some data will be lost)
- Which devices should be used for HR monitoring?
- How long the study should be?
- When HR should be measured (day/night/morning/evening/24-hour)?
- What are the reasons of loosing data and how to avoid them?

All these decisions allow do not fail the big user study and minimise its cost.

To be able to set the big burnout study we have conducted three small user studies. The first one is the comparison of the heart rate monitors, which allows us to choose from the list of devices (see Chapter 3) the most suitable device for the purposes of the big study. The aim of this user study is to compare the accuracy of the monitors in comparison with the Nexus, which is a medical device. The second study is aimed to investigate the user behaviour in the sense of preferences of the condition of wearing the heart rate monitors. Since some of the HR monitors require the additional device (receiver) for storing heart rate data, it can be an additional burden for the participants. That is why, in parallel to the first user study, we wanted to test how people behave with different devices as a receiver (iPod, iPhone). This user study also aimed to figure out the percent of data that we can expect to get in the big study, and give the explanations of some reasons why data can be lost.

After the analysis of these user studies we have conducted the Pilot study with chosen devices. The pilot study aimed to test the chosen settings of the big user study on a small number of people during a relatively short period. The whole structure of the user studies is depicted in Figure 9, the detailed description and analysis of the result of each study can be found in Chapters 5-7.
Figure 2: Structure of the user studies
5  User study 1: Device comparison

This part is confidential.
6 User study 2: iPod vs iPhone

6.1 Objectives and test procedure

The primary objective of the user study is to figure out whether there is a significant difference in the behaviour of the participants when they use MIO Alpha wristband together with iPhone and MIO Alpha with iPod. We assumed that for a person it could be easier to use its own mobile phone as a receiving device for MIO, than to carry an additional device for that.

The secondary objectives are to understand how much data can be obtained from an user study with MIO Alpha and how much data can be lost. Additionally we wanted to explain the possible reasons of loosing data: is it a user fault or device problems. We expected to see at what time and for what period the participants usually took off the devices in order to set that time as a time for recharging in the next user studies.

Four male Philips employees who had their own iPhone participated in this user study. They were instructed to wear MIO Alpha wristband during 2 weeks, day and night. They were also notified that we were especially interested in night and morning data, however the more data would be recorded, the better it would be. Two participants used iPhone as a receiver of MIO data during the first week of the user study and iPod during the second one, while another two participants vice versa, in order to exclude the impact of adaptation to the user study conditions. We did not specify when the participant should recharge the devices, but we asked to try to do it each day at the same time.

For recording the MIO data the “HeartRateLogger” software application developed by Philips was installed on iPods and iPhones. As soon as MIO Alpha finds heart rate it becomes discoverable for Bluetooth devices. When the connection between MIO and a receiver is established, the application creates a file with a current date and time as a name. The heart rate is recorded together with a timestamp when the value is received. If the connection between the receiver and MIO is lost, then the application stops writing the current file and creates a new one when the connection is established again. After the first pairing the devices, the application remembers the MIO and can automatically reconnect after the interruption of communication.

<table>
<thead>
<tr>
<th>Week 1</th>
<th>Group I</th>
<th>Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>MIO Alpha&lt;br&gt;iPod touch&lt;br&gt;iPhone (personal)</td>
<td>MIO Alpha&lt;br&gt;iPhone (personal)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week 2</th>
<th>Group I</th>
<th>Group II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 2</td>
<td>MIO Alpha&lt;br&gt;iPhone (personal)</td>
<td>MIO Alpha&lt;br&gt;iPod touch&lt;br&gt;iPhone (personal)</td>
</tr>
</tbody>
</table>
6.2 Software application for data visualisation

To understand how much data is obtained, when and possibly why data is lost, we wanted to visualise it. The way of representing the data was changing while we were analysing it. First, the idea was to represent each day of user study for each participant in form of horizontal line coloured in black and white depending on the data existence at each particular moment. In other words, a day is represented with a set of intervals with and without data. Additionally, we have introduced two colours for distinguishing data coming from iPod and iPhone. For iPhone data a green colour was chosen, and for iPod data orange. In Figure 28 you can see the data of one participant. There is a time line above the block of the day lines and the percentage of time with data is shown at the right.

![Figure 3: First way of representing data (Participant G234)](image)

6.2.1 Types of gaps

It can be clearly seen that the intervals without data (which are also called “gaps”) are of different length and possibly of different reasons. For example, if there is a gap of 2 minutes during a working day, we can assume that a person did not take off the MIO, but went far from a receiver so that MIO and iPod or iPhone lost the connection between each other. Another reason of gaps can be charging MIO. For full charging of the device 2 hours needed, but it is not necessary that a participant put on the wristband immediately after that. Finally, sometimes the participants could not wear the devices due to their personal reasons for a long time. It was decided to add this information to the previous picture, the result you can see in Figure 29.

The application was designed in such a way that it allows changing the length of these three types of gaps and immediately shows the results on the picture. We have tried different settings of the gaps length. Finally, it was set as follows:

- $1 \text{ minute} < \text{Length}_{\text{Short gap}} \leq 30 \text{ minutes}$
- $30 \text{ minute} < \text{Length}_{\text{Long gap}} \leq 180 \text{ minutes}$
- $180 \text{ minute} < \text{Length}_{\text{N/A}}$

Short gaps between intervals with data are represented with red colour and long gaps are represented with white colour. The gaps which are greater than 3 hours is called N/A and represented with grey colour. We assume that during this type of gaps the participants took the MIO Alpha off independently of which kind of the receiver was used, and thus these
periods should not be taken into account in the analysis of differences in user behaviour corresponding to wearing MIO with iPod or iPhone.

![Figure 4: Data visualization with different types of gaps.](image)

(a) Data of the participant G234. (b) Legend

### 6.2.2 Percent with data

As you can see in Figure 28 the user study was started on Monday with iPhone as a receiver, next Monday the iPhone was replaced with iPod, and the third Monday was the last day of user study. It is wrong to consider second Monday separately for two devices for two reasons. First, it makes misleading impression of existing two days instead of one, and second, it unreasonably decreases the percentage of time with data. To solve this problem some additional changes were made. For instance, the days when the receiving devices were swapped are represented now in one day line. This led to necessity of calculating the percentage time with data for iPod and iPhone separately. The way of calculating the percentage of time with data was changed as well.

It is interesting how much data we can expect to get when a person *is wearing* the devices, i.e. when the long periods of time without data are not taken into account. To compute percentage of time of wearing MIO with the “iPhone” and “iPod” (columns “iPhone” and “iPod” in Figure 29) the following formulas are used:

\[
\text{iPhone}(\text{Day}_c) = \frac{\sum_{i \in \text{Day}_c, \text{Type}(i) = \text{iPhoneData}} |i|}{|\text{Day}_c| - \sum_{i \in \text{Day}_c, \text{Type}(i) = \text{N/A}} |i|}
\]

\[
\text{iPod}(\text{Day}_c) = \frac{\sum_{i \in \text{Day}_c, \text{Type}(i) = \text{iPodData}} |i|}{|\text{Day}_c| - \sum_{i \in \text{Day}_c, \text{Type}(i) = \text{N/A}} |i|}
\]
The length of all days is the same and it equals to 86400 seconds. Applying these formulas to the days, when a person was wearing both iPod and iPhone, is incorrect, because in that case short and long gaps between intervals with data when a person was wearing MIO with iPod will be taken into account while calculating percentage for iPhone column. That is why the application first checks each day if it is a day with both iPod and iPhone, and if it is, makes following calculations:

- Finds the earliest start and the latest end of the intervals of type “iPhone data”: \( \text{iPhoneMin} \) and \( \text{iPhoneMax} \)
- Finds the earliest start and the latest end of the intervals of type “iPod data”: \( \text{iPodMin} \) and \( \text{iPodMax} \)
- Filters “N/A” intervals:

\[
\begin{align*}
\text{iPhone}_{N/A} &= \sum_{i \in \text{Day_c}, \; \text{Type}(i) = N/A \; \text{AND} \; \text{Start}(i) \geq \text{iPhoneMin} \; \text{AND} \; \text{End}(i) \leq \text{iPhoneMax}} |i|
\end{align*}
\]

\[
\begin{align*}
\text{iPod}_{N/A} &= \sum_{i \in \text{Day_c}, \; \text{Type}(i) = N/A \; \text{AND} \; \text{Start}(i) \geq \text{iPodMin} \; \text{AND} \; \text{End}(i) \leq \text{iPodMax}} |i|
\end{align*}
\]

- Calculates the percentage of time with iPhone and iPod data separately:

\[
\begin{align*}
\text{iPhone}(\text{Day}_c) &= \frac{\sum_{i \in \text{Day}_c, \; \text{Type}(i) = \text{iPhoneData}} |i|}{(\text{iPhoneMax} - \text{iPhoneMin} + 1) - \text{iPhone}_{N/A}}
\end{align*}
\]

\[
\begin{align*}
\text{iPod}(\text{Day}_c) &= \frac{\sum_{i \in \text{Day}_c, \; \text{Type}(i) = \text{iPodData}} |i|}{(\text{iPhoneMax} - \text{iPodMin} + 1) - \text{iPod}_{N/A}}
\end{align*}
\]

The values from the “Total” column are calculated as follows:

\[
\begin{align*}
\text{Total}(\text{Day}_c) &= \frac{\sum_{i \in \text{Day}_c, \; \text{Type}(i) = \text{iPhoneData} \; \text{OR} \; \text{Type}(i) = \text{iPodData}} |i|}{|\text{Day}_c| - \sum_{i \in \text{Day}_c, \; \text{Type}(i) = N/A} |i|}
\end{align*}
\]

As you can see in Figure 29 the numbers in the first three columns are quite big even if a participant was wearing the devices only a half of a day. For example, on Sunday first week the participant started recording the data only at 2 p.m., but in the “iPhone” and “Total” columns one hundred percents are indicated. Indeed, there were no pauses while a person was wearing the devices that day, but in fact we do not have the recording heart rate data for whole day. That is why the last column (“With data”) was added:

\[
\begin{align*}
\text{WithData}(\text{Day}_c) &= \frac{\sum_{i \in \text{Day}_c, \; \text{Type}(i) = \text{iPhoneData} \; \text{OR} \; \text{Type}(i) = \text{iPodData}} |i|}{|\text{Day}_c|}
\end{align*}
\]
To summarise, the “With data” column shows how much data is obtained in fact, while the “Total” column gives an impression about the amount of data which is lost while a participant was monitoring his heart rate, i.e. how much data is lost during short and long gaps.

6.2.3 Zero data
Analysis of the data showed that during the first week of the user study the iPod of one participant received heart rate data from MIO almost without any gaps (see Figure 30(a)). That was very strange because the battery of MIO allows contentious working during ~30 hours, and after that it needs to be recharged. At the same time, MIO cannot monitor heart rate while it is charging due to the design of the wristband. We have contacted with the participant and figured out that during first week he did not turn off the heart rate monitoring mode of MIO when put it on charging. We realized that even if we have a data for some period it can be that this data is meaningless. Indeed, when a MIO cannot find pulse it does not break the connection with a receiver, but continues sending zeros instead of a heart rate.

It allows as creating a new type of intervals while reading data files. This new types are called “iPhone Zero” and “iPod Zero” depending on the receiver and coloured with light green and yellow correspondingly. The intervals of “iPhone Zero” and “iPod Zero” types are considered the same was as long and short gaps while the percents of time with data are calculated. Later, it turned out that data of all participants have zero values in heart rate.
Figure 5: Zero data representation. (a) Data representation without zero data detection. (b) Data representation with zero data detection. (c) Legend

6.3 Data analysis

6.3.1 Analysis of the behaviour of the participants
Two participants coded with G234 and K590 were wearing MIO Alpha with iPhone the first week and with iPod the second one. A lot of night data of the participant G234 were lost
because he took off the devices late evening for more than 3 hours even though he was instructed that night and morning data are especially important for the study (Figure 31). He also did not wear the devices from Tuesday to Thursday of the second week because of the personal reasons. When he was wearing the devices there were a few short periods without data. Probably he took off the MIO for water procedures in the morning, around 6-8 a.m. There are two “iPhone Zero” and one “iPod Zero” intervals in the evening. It seems that he put his MIO on charging and did not turn off the heart rate monitoring mode. That is why we have relatively big numbers in the “Total” column (around 90%), but quite small numbers in the “With data” column, which shows absolute percent data obtained during each day. We cannot see any differences in the patterns of wearing MIO together with iPhone and iPod.

In Figure 32 the difference between first week and second one is clearly seen. There is a pattern of wearing MIO with iPhone during the first week of the user study: each day except Saturday the participant took off the MIO for a period of 1-2 hours, as a rule he did not turn off the monitoring mode and that is why we can see on the picture “iPhone Zero” intervals instead of “Long gap” intervals, which are supposed to represent the time of recharging. The average percent of time with data when the participant was wearing the devices is above 90%. The situation was dramatically changed when a participant started to use iPod as a receiver. There are a lot of short gaps (shorter than 30 minutes) during day and night time. There are also long gaps during day time between 12 a.m. and 4 p.m. which are depicted with white colour. The participant explained that he kept iPod in his bag an iPhone in his pocket. That means that during working hours wherever he goes he has the iPhone with him and leaves the iPod in his office; that is why the connection between iPod and MIO was broken much often then the connection between iPhone and MIO. That can explain the gaps during day time, but it does not explain the short gaps during night time. Obviously, the reason of these gaps is not the behaviour of the participant, but a bad connection of MIO with iPod.
The next two participants who are coded as R612 and X217 first used iPod as a receiver and then iPhone. The participant with code R612 was extremely responsible. As a result, we can see the most clear pattern of the wearing of the devices in Figure 33. During both weeks, he took the MIO approximately for 2 hours between 1 p.m. and 10 p.m. Since there is no other long gaps, it can be concluded that the participant recharged the MIO during that time. The percent of time that the participant was monitoring and recording heart rate is extremely high for all days independently of the receiver.

The participant coded with X217 had his holidays in contrast to other participants who had their standard working days. That is why Figure 34 differs from the previous ones: there are a lot of N/A intervals, when the participant did not wear devices at all, and it is hard to distinguish the overall pattern of the behaviour. There is a lot of zero intervals starting from...
Friday and ending on Saturday of the first week. According to the sequence of the intervals with normal heart rate data and zero heart rate data those day, we can assume that the participant did not took off and then put on MIO again, but MIO lost the heart rate while it was on the participant’s wrist.

Figure 9: Data of the participant X217

In Figure 35 the percent of the time with data is grouped by the participants and the receiver. Figure 35(a) shows the summarised result without distinguishing zero heart rate, and in Figure 35(b) the zero heart rate is subtracted. Both pictures do not include intervals which is called N/A, i.e. the figures are drawn on the basis of values is the “Total” column of the previous pictures. It is clearly seen that subtracting zero data significantly decreases the average percent of the time with data for the participant G234 (second week), because the recording of the last day of the user study of this participant contains mostly zero heart rate.
Figure 10: Percent of the time with data grouping by participant and the type of a receiver. (a) Zero data are considered as normal data. (b) Zero data are considered as intervals without data.

In Figure 36 the information is grouped by the participants. Handling zero data as a lack of data makes violin plots to be more stretched. In Figure 36 the values from the “Total” column are used, representing the amount of time with data when the participants were actually wearing the devices or put them on charging, but there were no gaps in data longer than 3 hours. In Figure 37 such long gaps are taken into account and considered as intervals without data. Independently of the way of handling zero data the participant R612 shows higher average percent of wearing the devices, while three other subjects show almost the same results with mean equalled to 60%.
Figure 11: Percent of the time with data grouping by participant. N/A intervals are not included. (a) Zero data are considered as normal data. (b) Zero data are considered as intervals without data.

Figure 12: Percent of the time with data grouping by participant. (a) Zero data are considered as normal data. (b) Zero data are considered as intervals without data.

We have also grouped the data of all participants by the type of the receiver: iPhone and iPod. In Figure 38(a) it is clearly seen that deviation of the percents with iPod is greater than that of data with iPhone. However, if the data of the participant K590, where iPod lost the connection with MIO very frequently, is excluded then the difference between iPhone and iPod is eliminated. That means that if a person has an iPod in this pocket instead of his bag, than it does not matter what to use as a receiver.
Figure 13: Percent time with data grouping by the type of a receiver. Zero data are considered as normal data. N/A intervals are not included. (a) All participants. (b) Participant K590 is not included.

We have changed Figure 38 by subtracting zero data (see Figure 39). In comparison to Figure 38 the mean of the iPod data in Figure 39 is less than 100 percent. Mainly it is because the participant R612 did not turn off the MIO while was charging it (see Figure 33). The orange violin is stretched down because of the three outliers, but more than half of values are above 70% in both Figure 39 (a) and (b).

Figure 14: Percent time with data grouping by the type of a receiver. Zero data are considered as intervals without data. N/A intervals are not included. (a) All participants. (b) Participant K590 is not included.

6.3.2 Interval length

To investigate the nature of the intervals with data and gaps between them (their lengths, the time when they start, the time when they end, their order) we have presented them in scatter plot (see Figure 40). The horizontal axis represents the time line staring from 12 p.m. The vertical axis indicates the length of the intervals or gaps in seconds.

In Figure 40 (a) and (b) the data of the participant G234 are shown. There are a few intervals and gaps in whole. There are only gaps in the data recorded with the iPhone which started in the night from 12 p.m. till 12 a.m. as well as intervals from the iPhone started from 9 p.m. till 8 a.m. The intervals with heart rate data are rather long, approximately 30 000 seconds, which
is equal to 8.33 hours, and the longest intervals started at lunch time. In general, the length of the gaps is less than that of intervals with data both for iPod and iPhone.

The data of the participant R612 are depicted in Figure 40 (c) and (d). There is a significant difference between the length of the intervals with data and the gaps between them: the intervals are much longer than gaps. There is no gaps which is started at night from 12 p.m. till 8 a.m. There are several clots of the point on the Figure 40 (d) around 3 p.m., 7 p.m. and 11 p.m. It is holds for both iPod and iPhone data and shows that the participant is very stable in his chose of the time for charging the devices.

There are a lot of very short intervals and gaps in iPod data in Figure 40(e) and (f), which depicts the data of the participant K590. The intervals of iPhone data are longer, but there are considerable fewer of them. Finally, there is no any activity, i.e. starting of an interval with data or starting of a gap (which is actually the end of the previous data interval), early morning from 6 a.m. till 8 a.m. both in iPod and iPhone data.
Figure 15: Start and length of the intervals with data and gaps, grouped by participants.
Participants: G234 - (a), (b), R612 - (c),(d), K590 - (e),(f), X217 - (g), (h).
Finally, Figure 40 (g) and (h) show the data of the participant X217. There are long intervals and gaps between them as well as short ones. Intervals with data, which start after 6 p.m., are mainly longer than those which start in the morning. From 2 a.m. till 8 a.m. there is no start of the intervals with data. The rational reason is that the participant sleep at that time and cannot turn on the devices. However, exactly during this period the longest gaps start, especially in iPod data. As we expect that the subject was sleeping during this time, it could be that the battery of MIO became empty and it remained off till the morning.

On the whole, there is a time when no one interval starts, or no one gap starts, or no any activity at all. Nevertheless, the figure of the participant K590 significantly differs from the other, because of the bad connection of the MIO with iPod, which has led to producing a lot of short intervals and gaps.

Figure 41 (a) and (b) show the data of all the participants together. The intervals of the participant K590 stand out from the rest and make it difficult to analyse the remaining data. That is why the data of this participant are excluded in Figure 41 (c) and (d). It is clearly seen that, generally, gaps are shorter than intervals with data, however in the night some very long gaps start. The activity during the night time is generally less than during day time (“activity” means turning off and on the device or loosing and recovering the connection). For instance, there are no any starts of intervals from 2 a.m. till 7 a.m.

Figure 42 shows the distributions of the lengths of the intervals with data and gaps, grouped by the type of the receiver. Again in Figure 42 (b) the data in the participant K590 is excluded because it differs from the rest data due to inappropriate using of the iPod. It is clearly seen
that in average the length of gaps is less than that of intervals. However, the means of the intervals lengths for both iPod and iPhone is approximately 1000 second, which is only 16.6 minutes. At the same time, in Figure 42 (b) there is no difference between iPod and iPhone data.

Figure 17: Distribution of the lengths of the intervals and gaps. (a) All participants. (b) Participant K590 is excluded.

6.4 Discussion

Generally, the participants were used iPod the same way as iPhones. The only one participant K590 had his iPod in his bag, because there was no place for the device in his pocket. As a result his iPod frequently lost the connection with his MIO for short periods. In the next user studies, in which MIO is used, it is important to be sure that the potential participant is going to carry the receiver with him, not in his bag.

Before the analysis of the heart rate data it is needed to check the correction of it. For example, if a MIO Alpha does not sense the heart rate, it continues sending “0” instead of a heart rate. Thereby zero values can appear in heart rate data in two cases: when a person take off the MIO and does not turn it off, or when a MIO looses heart rate while being on the wrist of a person. The last case is very common for participants with hairy arms.

When the participants were wearing the MIO or putting it on charging, more than 90% of the data were recorded. However, taking into account that the participants left the device off for a long period of time (more than 3 hours), the total percent of time when the data was recorded is about 60%. That means that to obtain more data we need to engage the group of people who are more responsible (as the participant R612) or motivate the participant by some benefits, not necessary financial.
7 User study 3: Pilot study

7.1 Objectives and test procedure
The primary objective of the pilot study is to test the chosen settings of the big user study on a small number of people during a relatively short period, because the mistakes in the big study are very expensive.

The secondary objective is to check the correlation between the HR and the answers on the questionnaires.

According to the results of the first and second user studies, the MIO wristband was chosen as a HR monitor, together with iPod as a data receiver and the “HeartRateLogger” software application installed on it. We have asked 6 volunteers to monitor their HR during 4 weeks, day and night. We have also included the DTI-2 wristband which has an accelerometer and, thus, can be useful in catching the moment of falling asleep and waking up. As in the previous study, we did not specify when the participant should recharge the devices, but we asked to try to do it each day at the same time.

Besides of wearing the two wristbands, the participants were asked to fill in the questionnaires about their psychological and physical state: 5 questions each morning, 4 questions in the evening, and once per week (on Friday) the Burnout questionnaire. The software application “StressDataCollection” developed in TU/e for the Stress@Work study was use in this pilot study. All the questionnaires were presented in the electronic form, and the participants could choose the answers by slicking on them. Additionally, the application allows automatic copying the data from the DTI-2 and sending it together with the questionnaire data to the server, when the device is plug in to the computer through a USB port for charging. In Figure 43 you can see the questions of the morning and evening questionnaires: the participants could click on one of the pictures from each row that describes his condition the best.
For the estimation of the burnout level of the participants, the Maslach Burnout Inventory General Survey (MBI-GS) was used. It is the main questionnaire used for research on burnout [16, 17]. It defines burnout as a 3-dimensional syndrome, characterized by:

- energy depletion (exhaustion),
- increased mental distance from one’s job (cynicism),
- reduced professional efficacy.

MBI-GS consist of 16 questions each of which is scored on a 7 point scale, ranging from 0 (never) to 7 (daily). The scores are converted into 3 subscales according to the definition of burnout: exhaustion, cynicism and efficacy.

The analysis of this user study is divided into two parts. First we describe some analysis that we were able to do, and then describe the amount of data that was obtained. Such structure is chosen because the attempt of the analysis discovered the additional problems: the real amounts of the data that can be used in further analysis can differ from the amount of the collected data. In particular, some gaps in the recording can make impossible to calculate the important features and therefore reduce the amount of data samples available for analysis.

### 7.2 Boundaries of a night

While planning the user study we supposed that the HR slightly decreases while a person is sleeping. In Figure 44 (a) and (b) you can see the example of the HR of two participants during the night period. The HR in figure recorded with MIO and then averaged for each 5 minutes, because such accuracy is considered as sufficient in the study. We can notice in Figure 44 following interesting facts:

- the night HR pattern is different for different persons;
- however for the same person the night HR pattern repeats each night;
• the HR slightly decreases during the night, as we expected.

Figure 19. Night HR of two participants. The nights are matched manually. The black lines are the trend lines.

In order to simplify the notion of a night HR pattern the following 5 points of the HR line were chosen as the main HR measurements (see Figure 45); in all cases a “point” means the average HR on 5 minutes:

• $hr_1$ - the HR just before a person falls asleep;
• $hr_2$ - the HR just after a person falls asleep;
• $hr_3$ - the minimal HR during a night;
• $hr_4$ - the HR just before a person wakes up;
• $hr_5$ - the HR just after a person wakes up;
Figure 20. The main points of the night HR

To be able to extract these five HR points, the moments of falling asleep and waking up must be known. It is possible to automatically recognize the time of falling asleep and waking up using the DTI data, in particular, the accelerometer data. The frequency of the DTI for this study was set as 2 samples per second. In Figure 46 you can see the fragment of the DTI data. We can see that a participant moved actively till approximately 1 a.m.; from 1 a.m. till 8 a.m. a person moved his arm approximately twice per hour. We can conclude that the period 1 a.m. - 8 a.m. is a period while the participant was sleeping. Very similar figures were obtained for other participants, thus, we can also conclude that the period of sleep does not characterized by the lack of any activity; however the frequency of movement is dramatically decreased in such period.

Figure 21. An example of the DTI accelerometer data

For automatically detection of the moments of falling asleep and waking up, in other works the “start” and “end” of the night, the following four steps were applied (in all the steps the same example of the DTI data is used as the example of the result of the step as in Figure 46):

**Step 1**

It is clearly seen that the coordinates of all three axes change a lot while a person is awake and change much less frequent while he is sleeping. That means that during the night time there are intervals with the stable value of X, Y and Z, while during the day time the variability of them is very high. That is why the first step of the algorithm of the night boundaries detection
is finding the sample standard deviation of the values within each 10 seconds for each axis using the formula:

\[
\sigma_i^X = \sqrt{\frac{1}{n} \sum_{j=i+n}^{(i+1)n-1} (x_j - \bar{x})^2},
\]

where \(i\) - is a number of a 10-seconds interval, \(n\) - is a number of samples within that interval (\(n = 20\), since the frequency of the DTI is 2 samples per second), \(x_j\) - is a \(j\)-th sample and \(\bar{x}\) is an average of the samples within that interval.

After that we need to find the biggest sample standard deviation amongst all the axes \(\sigma_i = \max(\sigma_i^X, \sigma_i^Y, \sigma_i^Z)\). In Figure 47 you can see the result of this step.

---

**Figure 22. Sample standard deviation of an example of the DTI accelerometer data**

**Step 2**

Now we need to take a median of the sample standard deviation samples on the interval of 5 minutes (5 minutes is in enough for representing HR trend).

\[
m_k = \text{median}_{i=k+l}^{(k+1)l-1}(\sigma_i)
\]

where \(k\) - is a number of 5-minutes interval, of \(\sigma_i\) - is a sample standard deviation for \(i\)-th 10-seconds interval, \(l\) - is a number of such elements within that 5-minutes interval (\(l = 30\)).
Figure 23. Median of the sample standard deviations of the DTI accelerometer data for each 5 minutes

Figure 48 shows the result of the second step. We easily distinguish the periods when a person was active from the periods when he was calm. Moreover, $m_k$ value is low and extremely stable during the period, when a person was calm.

**Step 3**

Now we can set $m_0$ as boundary of different types of activity:

- if at a moment $k$ the value $m_k \leq m_0$, then we consider that a person is *Calm*;
- if at a moment $k$ the value $m_k > m_0$, then we consider that a person is *Active*;
- if at a moment $k$ the value $m_k$ is lost for some reason, then we do not have information to detect the type of activity and mark it as *NA*;

The Algorithm 6 allows labeling time intervals according to the described rules:

1. if $(A[0].M < m)$
   2. type ← “calm”
2. else if $(A[0].M < NAvalue)$
   3. type ← “action”
4. else “NA”
5. last ← A[0]
6. i ← 0
7. while $(i \neq \#A)$ do
   8. if $(A[i].M < v \ AND \ type \neq \text{“calm”})$
      9. Activity ← Activity ∪ {<last.T, A[i].T-1, type>}
     10. type ← “calm”
     11. last ← A[i].T
   12. else if $(A[i].M < NAvalue \ AND \ type \neq \text{“action”})$
      13. Activity ← Activity ∪ {<last.T, A[i].T-1, type>}
     14. type ← “action”
     15. last ← A[i].T

40
```plaintext
17    else if (A[i].M = NAvalue AND type ≠ “NA”)
18        Activity ← Activity ∪ {<last.T,A[i].T-1,type>}
19        type ← “NA”
20        last ← A[i].T
21    i ← i + 1
22end do
```

Algorithm 1. Labeling the intervals with the activity types

**Step 4**

The last step of this procedure is the simplest one: the “start” and the “end” of a night can be found as the start and the end of the longest *Calm* interval. Actually, if the approximate time of going to sleep and waking up are known, it may be needed to add some constraints that make the full algorithm of finding nights boundaries even simpler.

In Figure 25 the DTI data (median of the sample st.dev.) are presented together with the HR data for the same period, the grey area shows the automatically defined boundaries of the night. It can be noticed that in the morning (around 9 a.m.) there is a gap in HR data and a flat Calm interval in accelerometer data. It is very likely that the devices were taken off for this period (10 minutes) while a person took a morning shower.
7.3 The amount of the obtained data

It appeared that the obtained amount of data that can be used in further analysis was much less than we had expected when planning the user study. We assume that data for a particular night is present if there are:

- evening questionnaire before the night (all answers);
- morning questionnaire after the night (all answers);
- DTI recording that allows calculating the boundaries on the night;
- MIO data for the whole period within the night boundaries (in order to be able to calculate the minimal HR) and in points $hr_1$ and $hr_5$.

Table 6 shows that the first four participants took part in the user study during 4 weeks, and the last one participant agreed to do that for a longer period. In fact, the behavior of the participant 5 was very different from others: he has eagerness to be helpful in the study, reporting all tiny problems and scrupulously following the user study instructions, as a result we have 46 nights of measurements together with filled in evening and morning questionnaires before and after that nights correspondently. From each of the rest participants we have got the data for almost a half of the nights.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Length of the user study in nights</th>
<th>Number of lost nights (different reasons)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Measurements</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Participant 1</td>
<td>28</td>
<td>15</td>
</tr>
<tr>
<td>Participant 2</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td>Participant 3</td>
<td>33</td>
<td>17</td>
</tr>
<tr>
<td>Participant 4</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td>Participant 5</td>
<td>53</td>
<td>20</td>
</tr>
</tbody>
</table>

*6 nights without data or with big gaps, and 14 nights where only $hr_5$ point is lost

Table 1. The number of nights with and without data needed for analysis.

The major reason of the impossibility of the analysis of the rest of the night is the lack of the measurements data. The measurements can be lost due to different reasons:

- lost HR (not wearing MIO at all or gaps at some interesting points);
- lost accelerometer data (not wearing DTI or gaps that do not allow discovering night boundaries);
- or lost both HR and DTI data together.

For example, there is one night of measurements and there is full recording of the DTI that allows an automated discovering the time moment of falling asleep ($t_{st}$) and waking up ($t_e$). If there is any gap in HR data between $t_{st}$ and $t_e$, then it is impossible to find a minimal HR value ($hr_3$ in Figure 45). However the most unexpected fact is that it is appeared that almost half of the nights of the participant 5 have a gap in the HR data, when a person just wake up ($hr_5$). A point $hr_5$ represents the average HR for a 5-minuters period. It is very likely that a
person take a shower in the morning and that is why take off the devices. There two options, how to avoid such problems and do not throw this point out of the consideration. The first one is to ask a participants do not do take off the devices in the first 5 minutes after awake; the second one is to shorten the period of averaging of HR.

Since we have defined the list of the data that we want to have for each night, the lack of one of these issues leads to considering whole night of measurements as lost. It is very unlikely, that participants were wearing MIO and DTI wristbands separately, because they were instructed that the data must be collected together at the same time, however, it does not mean that the HR and accelerometer data are always absent and present together. For instance, there was a case when a participant has occasionally dropped his DTI, which is not waterproof, into the water, and it took some time to replace it. The other problem of using the DTI is that the date/time settings can be reset to defaults for no apparent reason. It also takes time to reveal the problem and reset the current time, or replace the device. As a rule, the real date and time of the recording after that type of the device error cannot be recovered, and the measurement cannot be used. Therefore, using the DTI wristband requires a permanent control of its recording and the possibility to fix date/time resetting problem shortly.

Filling everyday questionnaires appears to be less obtrusive than wearing two wristbands which measure HR and accelerometer data; however, it turned out, that it is impossible to find any correlations between the answers on the questionnaires and HR for two reasons:

- there are very few samples to find intrapersonal dependencies between changes of HR and so called “mood” (a cumulative value of the questionnaire results) for the participants 1-4;
- there is almost no variation in the answers of the participant 5.

The Burnout questionnaire (MBI-GS) should have been filled in each Friday after a working week. Normally, we expected to get 4 filled in questionnaires from each participant; however, the actual numbers of them are 4, 4, 2, 3 and 7 for each participant correspondently. Since we were restricted to involve only healthy people, who do not suffer from a burnout syndrome, and the duration of the user study is short enough, the participants’ MBI results did not change a lot, and therefore it is impossible to check the hypothesis.

7.4 Discussion

The pilot study showed a lot of important outcomes. Some of them are quite positive, while the others - the disappointing ones - can be considered as even more significant, because they suggest how to avoid mistakes in the big study.

The positive outcome of the study is that it turned out to be possible to automate the process of recognizing the boundaries of a night: the moment when a person falls asleep and wakes up. The algorithm developed in this project uses the accelerometer data recorded with DTI wristband and can be used in the big user study.
The most disappointing outcomes of the study are related to the amount of data that can be lost. The total percentage of the nights for which there are HR data, accelerometer data and filled in questionnaires is small - about 50% of all the nights of the study. It should be noticed that such big amount of data is lost not just because the participants do not follow the instructions, but also because of other reasons, for example, the malfunction of one of the devices.

The specific of MIO may also be a reason of data loss. Since MIO cannot store data onboard, it requires permanent connection with a receiver: if the connection is lost, then HR rate data are lost, and then the whole night can be considered as lost (if such gap is too big or it is in crucial moments).

Another problem of using the MIO is periodical loosing the Bluetooth connections with the receiver for a very short period: each 7 seconds with measurements are followed by 10 seconds without data. According to our observation such behavior is caused by Wi-Fi interference. Even though more than a half of data is lost in such a case, it does not influence on the accuracy to within a minute. The more important problem is that in such case a lot of small files are produced and they occupy much more space on a disc than if it would be one file. Therefore, after a long period of recording (for instance, a week) the process of uploading data from an iPod/iPhone to the computer is problematic, causing of freezing both computer and the receiving device.

The results of the burnout questionnaires did not change enough during the study for each participant. The first possible reason of that is that all participants were healthy at the beginning of the study and the length of the study is too short for observing changes in the state of a person (positive or negative). The second possible reason may be the fact that the participants have memorized the questions together with their answers and tended to repeat them in order to be consistent. Indeed, at least one of the participants has indicated that he has remembered the questions. Therefore, filling in the questionnaires should be done rarely than once a week.
8 General discussion and recommendations

In this work, three user studies were conducted and their outcomes were analyzed. In this chapter the results of these user studies are interpreted in terms of recommendations for the set-up of the big user study. The recommendations are divided into three blocks: devices, people and time.

8.1 Technology

For the devices tested, we found that MIO gives reasonably reliable information about the heart rate. It also has the advantage of being much less obtrusive compared to other devices. That is why, when only interested in measurement of HR, the use of MIO is recommended.

When using MIO, an iPod or another recording device is needed to gather the measurements. It appeared that considerable gaps in data occur over the period of measurement. In some cases people do not want or forget to take their recording device with them, in other cases the Bluetooth link gets temporarily broken for some reasons. It does make some difference whether the recording device is the participants own iPhone or a separate, additional iPod. Thus, especially for daytime measurements, when the recording device is not on the bedside table anyhow, we recommend to use the user’s iPhone for data collection. Even better it would be to get a MIO-version with data storage on it. That will also eliminate the risk of loosing data because of Bluetooth problems.

In order to distinguish the sleeping phase from the active phase the accelerometer data are needed. Actually, MIO has an accelerometer, but currently there is no way to read this data from the device, it is only used internally. That is why an accelerometer should be included into the study. Including an additional device makes the participation in the study more uncomfortable and leads to increasing of the probability of loosing data (see Section 7.4).

Ideally, it should be one device which monitors HR, records the measurements onboard and makes possible to recover the start and the end of night (has accelerometer data open). If such device is currently unavailable, then we recommend improving the application “StressDataCollector” used in the third experiment (see Section 7.2) so that it would collect both DTI and MIO data, and automatically check it (existence, date/time, gaps) in order to be able to react to any problem in time.

8.2 People

Finding the biomarkers of burnout requires having a variation in person’s burnout level. In the pilot study we were restricted to only healthy people. The level of burnout of these people did not change much during 4-5 weeks. We recommend in the big user study involve participants from one or both of two groups:

- people who has a burnout syndrome and recovering from it (in order to get a positive changes in burnout level);
- or people assumed to be the most predispose to burnout by virtue of their professions (however, there is no guarantee that a participant’s burnout level will change during a study);
The number of people should be chosen taking into account that a big percentage of the data can be unavailable for the analysis for different reasons. We assume that the number of nights suitable for analysis according to our requirements will be much bigger than 50% of the length of the study, if the study will be planned according to the recommendations, which are given in the sections 8.1-8.3.

In order to mitigate the risks of losing the data connected with user behavior the following actions should be taken:

- Involve people who are really interested in the study, motivated to follow the instructions and have an opportunity for that. In experiments 2 and 3 there were two different people whose behaviors were different from the other participants in a good way. They strictly followed the instructions, reported about problems, and as a result, collected much bigger amount of the data than the rest of the participants.
- Do not ask too much. In the second study (devices comparison) we asked participants to wear the devices only for one night, and we got the data. In the third experiment the participants were asked to wear the devices during about 30 days and nights, and we lost a lot of the data. The other example is the questionnaires. The participants skipped filling the morning or evening questionnaires, but filled in the burnout questionnaire once a week.

8.3 Time
In order to see the changes in the burnout level we advice to prolong the study, since there were no significant changes in it during the period of the study. At the same time, filling questionnaire should be done less often than once a week, because the participants indicated that they remember the questions and their answers.

It is also possible to limit the measurements by only night time, instead of a whole day, because the activity during the daytime can be very different.


\section{Conclusion}
In this work, we have studied the application of the pervasive technologies for predicting burnout. Validation of the hypothesis that HR monitoring allows early detecting upcoming burnout requires a big user study which is difficult, expensive and should be carefully planned beforehand. That is why this work was the preliminary study including three user studies aimed at the understanding how to set up the big study.

We have considered a variety of the devices that can continuously monitor HR and have chosen those that appeared to be the most suitable for our study, according to the characteristics defined as the most important in the context of the project. To be able to find the balance between the unobtrusiveness of a device and its reliability, we have conducted a small user study: 8 participants have been wearing a set of the HR monitors for one night at their own homes. The Nexus-10 ECG sensor signal was used as the ground truth to estimate the accuracy of the devices. Based on the study results, the MIO Alpha wristband turned out to be both the most unobtrusiveness and the most reliable HR monitors among the considered ones. Its main disadvantage is the need of an additional Bluetooth Smart device to store the measurements.

The second study was aimed at the investigation if there is any difference whether the recording device is an iPhone (belonging to the participant) or a separate, additional iPod. It involved 4 people who were asked to wear the MIO Alpha during 4 weeks whole day and night: 2 weeks they used their own iPhone for receiving MIO data, and 2 weeks an iPod. The user study shows that there are considerable gaps in data collected with both iPhone and iPod; however, the probability of not getting measurements with iPod is higher, especially during daytime.

The last study was aimed at testing the presumptive settings of the prospective big study. In that study 5 participants were involved. They were asked to wear the MIO Alpha together with the DTI-2 wristband during 4 weeks whole day and night, as well as to fill in the Burnout questionnaire (MBI-GS) once per week, and everyday questionnaires each morning and each evening. The DTI-2 was used to monitor movements with 3D accelerometer, and it was included into the user study in order to distinguish the periods when a person was awake from the periods when he was sleeping. Indeed, it appeared to be possible to discover the moments when a person falls asleep and wakes up automatically based on the accelerometer data. However, it also turned out that a lot of the data were lost due to various reasons: for example, the participants did not wear one or both waistbands for some periods, they skipped the filling of the questionnaires, the connection between MIO and iPod was lost.

The summarized results of the user studies allowed us to formulate the advices for the planning of the big study mitigating the discovered risks. Taking into consideration the above, it can be concluded that the formulated objectives have been reached and the project has been finished successfully.
Bibliography


