

How accurately can sitting and the intensity of walking and cycling be classified using an accelerometer on the waist for the purpose of the “Global recommendations on physical activity for health”?

Diederik van der Valk

EWI

Biomedical Signals and Systems – Telemedicine

**Examination committee:**

Prof. Dr. Ir. H.J. Hermens

Ir. S.T. Boerema

Ir. T.M. Tönis

**Document number**

BSS – 15 - 09

06-07-2015

## Preface

This study has been written for my bachelor assignment of the bachelor Electrical Engineering. With the knowledge gathered during the bachelor I was glad to finally apply this in practice. I'm satisfied with the result and I hope the results can be used in the future.

I would like to thank my supervisors Prof. Dr. Ir. H.J. Hermens, Ir. T.M. Tönis and Ir. S Boerema for having offered me the opportunity to perform this study in such a manner, that I could have a look at all the aspects of the research, from literature research via creating my own dataset to writing an analysis program.

Furthermore I would like to thank everybody who cooperated in this study, especially all the participants who voluntarily let me order them around, doing all sorts of activities while wearing an accelerometer on their belt. I could not have done the research without them.

## Abstract

The society has become more sedentary and has developed a lack of physical activity, therefore increasing health risks. Feedback is needed to change these behaviours. For this feedback, first accurate monitoring is needed: sedentary behaviour must be classified as well as the intensity of physical activity.

In this report a State of the Art analysis is performed to compare different classification techniques and finally two methods, both using an accelerometer on the waist, are worked out. These methods are Integral of the Modulus of the Accelerometer (IMA) classification and a machine learning technique (MLT): support vector machine (SVM). These methods are then applied in a laboratory experiment to study their quality. A measurement setup is made to create a dataset of the following activities: standing, sitting, lying, walking (2.4 - 7.5 km/h) and cycling (10.1-19 km/h). This dataset (n=15) is analysed and classified using Matlab for both methods.

The IMA method was unable to monitor sedentary behaviour, but could classify the physical activity (PA) intensity with an accuracy of 66%. The SVM method within subjects was able to monitor sedentary behaviour with an accuracy of  $91\pm 20\%$  and the classification of the PA intensity has an accuracy of  $94\pm 5\%$ . For between subjects the accuracies decrease to  $71\pm 13\%$  for PA intensity accuracy and  $45\pm 35\%$  for the sedentary behaviour classification.

IMA was implemented in the old feedback system, monitoring the overall daily amount of physical activity, but can significantly be outperformed by replacing it with the current SVM implementation. At this moment however, SVM can only be used to improve the old system, it cannot yet be used to create new additions to the feedback system, such as the implementation of the feedback of the sedentary behaviour and specific physical activity intensities. New methods and properties of the current SVM system have been found that might increase the accuracies of the between subject analysis and therefore might enable SVM to become applicable for the extra feedback options.

# Content

<b>PREFACE .....</b>	<b>2</b>
<b>ABSTRACT.....</b>	<b>3</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>5</b>
<b>1. INTRODUCTION .....</b>	<b>6</b>
<b>2. STATE OF THE ART .....</b>	<b>8</b>
2.1. GENERAL MEASURES OF PHYSICAL ACTIVITY INTENSITY .....	8
2.2. GUIDELINES OF PHYSICAL ACTIVITIES.....	8
2.3. MONITORING PHYSICAL ACTIVITY, DIRECT EE OR INDIRECT .....	9
2.4. MEASURING PHYSICAL ACTIVITY IN PRACTICE.....	10
2.5. POSITION OF THE ACCELEROMETER .....	11
2.6. PROCESSING PHYSICAL ACTIVITY DATA.....	11
2.7. IMA .....	11
2.8. INDIRECT METHOD (MLT).....	11
2.9. SVM .....	14
2.10. FEATURE CALCULATION.....	15
<b>3. METHOD .....</b>	<b>18</b>
3.1. DESIGN .....	18
3.2. PARTICIPANTS.....	18
3.3. ETHICS .....	18
3.4. MATERIALS .....	18
3.5. PROCEDURE .....	24
3.6. DATA PROCESSING MATLAB PROTOCOL .....	25
3.7. OUTCOME MEASURES.....	30
3.8. STATISTICAL ANALYSIS.....	30
<b>4. RESULTS .....</b>	<b>30</b>
4.1. PARTICIPANT CHARACTERISTICS.....	30
4.2. CALIBRATION.....	30
4.3. DATASET .....	30
4.4. MATLAB SCRIPT .....	31
4.5. DATA SYNCHRONISATION & MISLABELLING DATA .....	33
4.6. FEATURES .....	34
4.7. STATISTICAL ANALYSIS .....	35
4.8. TYPE OF ACTIVITY .....	36
4.9. IMA .....	38
4.10. COMBINING THE RESULTS.....	39
<b>5. DISCUSSION .....</b>	<b>42</b>
5.1. DATASET .....	42
5.2. DATA PROCESSING.....	43
5.3. GENERAL .....	47
<b>6. CONCLUSION .....</b>	<b>49</b>
<b>7. REFERENCES .....</b>	<b>50</b>
<b>8. APPENDIX.....</b>	<b>51</b>
A: INFORMED CONSENT .....	51
B: ANALOGUE ANNOTATION TOOL .....	51
C: MATLAB CODE .....	52
D: DETAILED RESULTS.....	52

## List of abbreviations

Below a summary is given of the used abbreviations in this study.

ANOVA	Analysis of Variance
EE	Energy Expenditure
IMA	Integral of the Modulus of the Accelerometer
LOO	Leave One Out
MET	Metabolic Equivalent of Task
MLT	Machine learning technique
PA	physical activity
RRD	Roessingh Research and Development
SD	Standard Deviation
SMA	Signal Magnitude Area
SMV	Signal Magnitude Vector
SVM	Support vector machine
WHO	World Health Organisation
WS	Within Subject

## 1. Introduction

Our current society has become a more sedentary society, causing people to become “more inactive”. This inactivity can cause health problems and therefore it is recommended that this inactivity is reduced[3]. Besides reducing the inactivity also moderate intensity physical activity is needed to maintain good health[1]. To achieve these two goals, people must be aware of their situation before they can try to change this behaviour. In order to help them achieve this awareness a monitoring and feedback system can be used.

Within Roessingh Research and Development (RRD) research is done to design such a feedback system. Accurate monitoring however is first needed to get information to improve the feedback system. For this monitoring, the type of activity should be classified to feedback on the sedentary behaviour and the intensity of the activity. This classification provides data that can be used to add features to the feedback system, such as information on the amount of sedentary behaviour as well as an extended feedback on the activities themselves, such as the hint to walk faster in order to execute the activities with the recommended intensities. Currently a system has been developed to determine the physical activity using a single 3D accelerometer[4], which uses Integral of the Modulus of the Accelerometer (IMA) analysis, but is not yet able to classify the types of activities. Furthermore the accuracy of classifying the intensities of activities is still dependent on the type of activity, cycling is often underestimated, etcetera. The main goal of this research is therefore to improve the current monitoring system. The recommendations used to implement in the feedback system are the “Global recommendations on physical activity for health” and since this research is done in the Netherlands, the “Nederlandse Norm Gezond Bewegen” is also kept in mind. In order to achieve the goal of improving the current monitoring system, the following research question has been proposed:

***How accurate can sitting and the intensity of walking and cycling be classified using an accelerometer on the waist for the purpose of the “Global recommendations on physical activity for health”?***

Following from this question the following sub questions rise:

1. How can the “Global recommendations on physical activity for health” be operationalized?
2. What is the state of the art in methods to classify physical activity from 3D accelerometer data?
3. Which physical activity intensities should be distinguished and what type of physical activities are relevant for the border between these intensities?
4. How accurate can the physical activity be classified using the Integral of the Modulus of the Accelerometer (IMA)?
5. How accurate can the physical activity be classified using “type of activity classification” when applying the most appropriate Machine Learning Technique (MLT)?

The state of the art of physical activity classification in general will be looked up to give a brief review of classification options. This in order to find the answer to sub questions 1 and 2. Furthermore the state of the art of the classification of physical activity using a (tri-axial) accelerometer on the hip is researched as mentioned in sub question 3. This gives a summary of different methods and the corresponding accuracy at the moment of writing and explains the IMA method and the analysis using support vector machine (SVM) in more detail.

With this knowledge an experiment is set-up to evaluate the quality of the SVM method described in the literature part compared to the currently used IMA method. The study protocol will be developed based on best practices from literature and data analysis requirements regarding number of subjects, number of repetitions and duration of activities.

The main difference between the SVM and IMA method is the aggregation level of the data. The IMA method uses a simple integral, therefore losing a part of the data, compared to the SVM method which looks at the raw data itself to calculate new features, for instance the IMA. The results are analysed and discussed in order to find the accuracy of the used methods regarding between- and within-subject variability. This is used to answer sub questions 4 and 5.

The accuracy of the methods that have been tested in practice will be compared to the accuracy of the methods found in literature, thereby answering the main research question.

## 2. State of the Art

### 2.1. General measures of physical activity intensity

For this study various subjects are discussed, but all use some basic definitions, these are worked out below. First the basic definitions are stated, then worked out.

The study is focussed on classifying the intensity of Physical Activities (PA). Physical activity is defined following the World Health Organisation as “any bodily movement produced by skeletal muscles that requires energy expenditure”[1] In literature PA is often defined, besides this Energy Expenditure (EE), as Metabolic Equivalent of Task (MET) or as intensity levels expressed in light-, moderate- and vigorous physical activity.

- The EE is given in ( $\text{kcal} \cdot \text{min}^{-1}$ ), this is an absolute intensity of PA and represents the energy used by the body per minute. This expression is different for all people, since body weight etcetera are not incorporated.
- The MET is a relative measure, relating the current intensity to the average metabolic rate of a person in rest. If the intensity of an activity is for instance tripled compared to rest, the MET value is 3. The MET is an average of a large group of people and 1 MET is defined by convention to  $3.5 \text{ ml O}_2 \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ . This value can be used as a common definition for all people.
- The MET values are used as cut-off points to indicate different PA intensities. PA intensity is divided in the light-, moderate- and vigorous physical activity intensity levels. These levels are often used in health recommendations.
- In order to determine the absolute energy used, different methods have been developed, such as doubly labelled water (DLW) and indirect calorimetry.
  - In the case of DLW, the subject receives a dose of water consisting of isotopes of hydrogen and oxide. By keeping track of the concentrations of these elements in the body, often measured from the concentration in the urine, the amount of oxide converted to  $\text{CO}_2$  can be calculated. The amount of  $\text{CO}_2$  corresponds to the amount of energy converted, this gives the average daily metabolic rate.
  - Indirect calorimetry also measures the amount of  $\text{CO}_2$  generated by the body, but in this case using the concentration in the air. This is for long term often done using a respiration chamber in which the subject perform any activity without hinder. For short term, breath by breath, also a mask can be used, but this might influence the activities slightly.
  - With these tests, the EE can be calculated as an absolute measure, for various time frames, weeks, hours or breath by breath.

Sedentary behaviour is defined, following the Sedentary Behaviour Research Network as “any waking activity characterized by an energy expenditure  $\leq 1.5$  metabolic equivalents and a sitting or reclining posture.” [2]

- For this sedentary behaviour the posture, whether the subject is sitting or in another posture, such as lying or standing, must be recognised. Following the definition the PA intensity must be classified as well.

### 2.2. Guidelines of physical activities

Sedentary behaviour is recognised as a problem for physical inactivity by the World Health Organisation (WHO), as well as the need for moderate intensity physical activity. They have set a global target to reduce physical inactivity, including for instance sedentary behaviour, by 10% by 2025 [3] and made recommendations to stimulate the amount of moderate intensity physical activity. The Dutch government has separately recognised the contribution of physical activity to health as well as the problems caused by sedentary behaviour and also made recommendations for those to decrease the sedentary behaviour and increase the PA [4].



For this research the focus has been on adults and a part of their corresponding recommendations on physical activity for health. The recommendation that is looked at is the recommendation of doing at least 150 minutes of moderate-intensity PA. Since the research is done in the Netherlands, besides the international Norm, also the Dutch norm is used. The international and Dutch norms being respectively: “to do at least 150 minutes of moderate intensity physical activity per week” and “move at least 30 minutes a day for five days a week with moderate intensity physical activity”. Also the definition of moderate intensity physical activity is slightly different, being between 3.0 and 5.9 METs for the international and between 4.0 and 6.5 METs for the Dutch recommendations. A section from the recommendations can be read below [4, 5].

Age group: 18-64 years old

1. Adults aged 18-64 years should do at least 150 minutes of moderate-intensity aerobic physical activity throughout the week, or do at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week, Or an equivalent combination of moderate- and vigorous-intensity activity.
2. Aerobic activity should be performed in bouts of at least 10 minutes duration.
3. For additional health benefits, adults should increase their moderate-intensity aerobic physical activity to 300 minutes per week, Or engage in 150 minutes of vigorous-intensity aerobic physical activity per week, or an equivalent combination of moderate- and vigorous-intensity activity.
4. Muscle-strengthening activities should be done involving major muscle groups on 2 or more days a week.

Moderate-intensity physical activity:

On an absolute scale, moderate intensity refers to the physical activity that is performed at 3.0–5.9 MET. On a scale relative to an individual’s personal capacity, moderate-intensity physical activity is usually a 5 or 6 on a scale of 0–10.

Vigorous-intensity physical activity:

On an absolute scale, vigorous intensity refers to physical activity that is performed at 6.0 or more MET for adults and typically 7.0 or more MET for children and youth. On a scale relative to an individual’s personal capacity, vigorous-intensity physical activity is usually a 7 or 8 on a scale of 0–10.

Figure 1 Selection of "Global recommendations on physical activity for health" [5]

Volwassenen (18-55 jaar): In totaal minstens een half uur per dag minimaal matig intensieve lichamelijke activiteit (tussen de 4 en 6,5 MET), op minimaal 5 dagen per week. Matig lichamelijk actief betekent voor volwassenen bijvoorbeeld stevig wandelen (5 km/uur) of fietsen (16 km/uur).

Figure 2 Quote of the Nederlandse Norm Gezond Bewegen [4]

### 2.3. Monitoring physical activity, direct EE or indirect

In order to monitor the physical activity, different methods can be used varying in feasibility and validity. Diaries and questionnaires are very feasible, but they have low validity in classifying the time of PA intensities performed during the day. Absolute measurements such as double labelled water (DLW) and indirect calorimetry are obtrusive and/or costly and therefore not feasible to implement on large scale and therefore not in this research. In between these options, there is the possibility of measuring using **body worn sensors**, such as accelerometers and pedometers. These are the middle ground for the feasibility and validity. Furthermore, looking ahead, the accelerometer has the possibility to be improved compared to their current state of the art usage. These considerations are plotted, showing the current situation as a filled dot and the potential situation regarding feasibility and validity with an open dot in Figure 3.

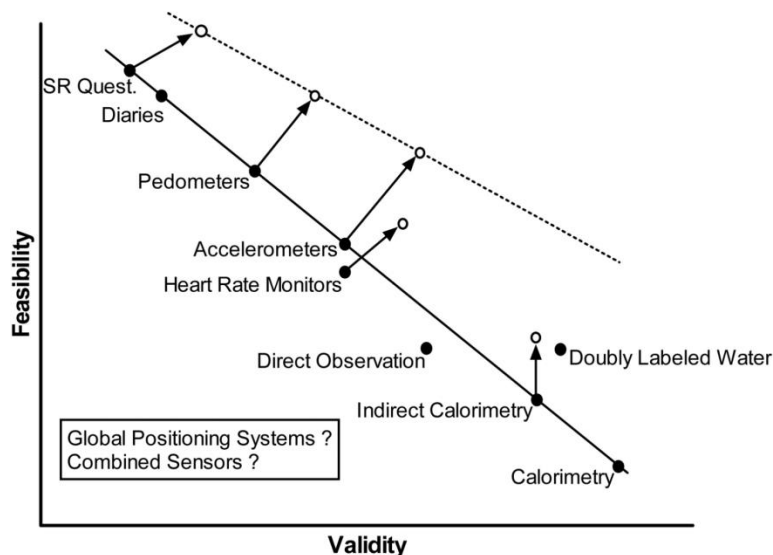


Figure 3 Activity classification methods general [6]

## 2.4. Measuring physical activity in practice

In order to determine - using indirect methods - the physical activity, various possibilities are available. Comparing different technologies, the **accelerometer** is chosen since it is a technology that can register a large range of potential behaviours, is rather unobtrusive and is sensitive for small body changes. The disadvantage however is the difficulty to determine the behaviour from a single sensor, but this is one of the challenges to be overcome in this research. Other body worn sensors are limited in their possibilities, such as a footswitch, or are more error prone in their results, such as a gyroscope. A summary of often used body worn sensors is visible in Table 1, stressing the relevance of the accelerometer. A side note must be made, that in order to be able to distinguish the mentioned different movements, a tri-axial accelerometer is needed to keep track of these movements in all directions.

TECHNOLOGY	RELEVANT BEHAVIOUR OUTPUTS	ADVANTAGES	DISADVANTAGES
ACCELEROMETER	Step count, distance, stance/swing time, energy expenditure, limb/body orientation, acceleration, active/sedentary time	Unobtrusive, large range of potential behaviours monitored	Difficult to determine behaviour from single sensor
GYROSCOPE	Limb/body orientation, angular velocity	Unobtrusive, large range of potential behaviours monitored	Difficult to determine behaviour from single sensor, require careful calibration, susceptible to drift over time
MAGNETOMETER	Orientation	Unobtrusive	Orientation alone is of limited use, require calibration
FOOTSWITCH	Stance time, swing time, stride time, step count, distance walked	Unobtrusive	Limited to information about gait
BAROMETRIC PRESSURE SENSOR	Altitude	Unobtrusive	Altitude alone is of limited use

Table 1 comparison of different indirect body worn methods to determine physical energy [7]

## 2.5. Position of the accelerometer

The position of the accelerometer also has quite some influence on the type of data gathered. Ellis, K., et al. stated: “The wrist accelerometer proved more useful in predicting activities with significant arm movement, while the hip accelerometer was superior for predicting locomotion and estimating EE” [8]

Often used positions are: the head, torso, waist, wrist, and ankle. The **waist** is chosen since it is compared to the other positions: low obtrusive; and has for small movements and postures high accuracy.[9-11]

## 2.6. Processing physical activity data

Accelerometers can be used to determine directly the PA intensity or indirectly the type of activity, which can be used to determine the PA intensity. In this study, the focus of this analysis is on often occurring activities such as standing, sitting, lying, walking with different speeds and cycling at different speeds. Methods analysed in this study to determine the PA intensity of these activities are respectively the determination of the **Integral of the Modulus of the Accelerometer (IMA)** and using a **machine learning technique (MLT)** to determine the type of activity, which can then be converted to a PA intensity value. Comparing these methods globally, the IMA method is the simple version, requiring the least processing, but also the least accurate method, processing only the acceleration of the accelerometer itself, therefore underestimating activities such as cycling. The method using MLT is the more advanced method, therefore also requiring more processing power, but using this more sophisticated analysis enabling the possibility to recognize activities and compensating for the underestimation of the activities, therefore achieving a higher accuracy.

## 2.7. IMA

For the determination of the PA intensity using IMA, the acceleration in three orthogonal dimensions is summed and integrated for a certain time period called an epoch length, this integral is called an IMA value.

$$IMA = \int_{t_0}^{t_0+epoch} (|acc_x| + |acc_y| + |acc_z|)dt$$

These IMA values can then be used as an input to a classification system which outputs the corresponding PA intensity using simple thresholding of these IMA values. The value of this thresholding can be determined from ROC curves, finding the value with the highest combination of sensitivity and specificity. For an explanation of the specifications of this classification, such as the values for the thresholding, see the work of A. Selderhuis[12].

As shortly mentioned before, a disadvantage of this method is the fact that certain activities are underestimated. “Waist-worn accelerometers underestimate the EE of free-living individuals [...]from arm activity, standing posture, vertical work [...], carrying extra weight[...], bicycling [...] swimming. [...]cannot detect increases in EE that occur at running velocities over 9 km·h<sup>-1</sup>.”[13]

## 2.8. Indirect method (MLT)

MLTs use a set of features extracted from the dataset of the accelerometer to determine different outcomes, in the case of this study the type of activity (sitting, standing, lying, walking and cycling). Determination of the type of activities can be done using various MLTs. These types of activities can then be related to the PA intensity. Below first the type of activities and the correspondence to PA intensity is explained, in general and for this study specific, afterwards the MLT analysis itself is worked out.

### 2.8.1. Type of activity

Specific types of activities can be related to specific PA intensities. In order to find this relation one of the absolute measurements of PA intensity such as double labelled water (DLW) and indirect calorimetry can be used. These measurements are used to determine the EE from the amount of oxygen used, in respectively a time period of days, minutes and breath by breath. Also the MET value can be determined if the mass and size of the subjects is known. If activities are performed during these absolute measurements, the found MET values can be linked to those specific activities. This relation is an average, since each person has slightly different energy expenditure for each activity. An **activity compendium** has been made of these

“intensity – type of activity” pairs.[14] A selection of these mentioned MET-activity pairs potentially relevant for the Global Physical Activity Recommendations is given in Table 2.

Code	METs	Activity
07011	1.3	lying quietly, doing nothing, lying in bed awake, listening to music (not talking or reading)
07021	1.3	sitting quietly, general
07023	1.8	sitting, fidgeting feet
07040	1.3	standing quietly, standing in a line
07041	1.8	standing, fidgeting
17085	2.5	bird watching, slow walk (no literature confirmation)
17170	3.0	walking, 2.5 mph (4.0 km/h), level, firm surface
17190	3.5	walking, 2.8 to 3.2 mph (4.5 to 5.1 km/h), level, moderate pace, firm surface
17200	4.3	walking, 3.5 mph (5.6 km/h), level, brisk, firm surface, walking for exercise
17220	5.0	walking, 4.0 mph (6.4 km/h), level, firm surface, very brisk pace
17230	7.0	walking, 4.5 mph (7.2 km/h), level, firm surface, very, very brisk
17231	8.3	walking, 5.0 mph (8.0 km/h), level, firm surface
01018	3.5	bicycling, leisure, 5.5 mph (8.9 km/h)
01019	5.8	bicycling, leisure, 9.4 mph (15.1 km/h)
01020	6.8	bicycling, 10-11.9 mph (16.1-19.2 km/h), leisure, slow, light effort
01030	8.0	bicycling, 12-13.9 mph (19.3-22.4 km/h), leisure, moderate effort
01040	10.0	bicycling, 14-15.9 mph (22.5-25.6 km/h), racing or leisure, fast, vigorous effort

Table 2 List of common activities with corresponding MET values [15]

When looking at these PA intensities combined with the recommendations<sup>1</sup> relevant for this study, the following analysis can be drawn: lying, sitting and standing are all part of the light category. Walking and cycling both start as a moderate activity, but at a specific speed become a vigorous activity.

Plotting the MET values versus the speeds of walking and cycling, an approximate linear behaviour can be found ( $R^2 = 0.95$ ). The recommendations consist of values not directly given in the compendium; therefore these are interpolated using the previous assumed linearity.

From literature both walking and cycling are categorised as moderate or vigorous, but walking slower is possible, therefore being categorised as a light activity. Cycling slower than 8.9 km/h, in order to reach the light category for the international recommendation, is not taken into account, as people might fall over at this speed, moreover, these lower speeds are not mentioned in the activity compendium. A list of the different speeds including the new calculated speeds to include the intensity boundaries is visible in

Table 3.

MET	2	3	3,5	4	4,3	5	5,8	6	6,5	6,8
-----	---	---	-----	---	-----	---	-----	---	-----	-----

<sup>1</sup> A short summary of recommendations. For international norm: light till 3.0 MET, moderate between 3.0 and 5.9 MET and vigorous if higher, for the Dutch norm: light till 4.0 MET, moderate between 4.0 and 6.5 MET and vigorous above.

<b>WALK SPEED (KM/H)</b>	2,4	4	4,8	5,3	5,6	6,4	--	6,9	7,1	--
<b>CYCLE SPEED (KM/H)</b>	--	--	8,9	10,1	--	--	15,1	15,6	16,9	17,65

Table 3 MET values selected from the compendium including new calculated values

For the sedentary behaviour, one of the requisites was the activity requiring a MET value  $\leq 1.5$ . In this study this requirement is left out, allowing sitting activities to be classified as sedentary behaviour, neglecting the PA intensity. This might classify activities, such as “sitting, fidgeting feet”, with MET 1.8 as sedentary behaviour, so with this assumption the classification might contain too many activities compared to the official definition.

The activities of this study are summarised below in Figure 4 with their corresponding MET values, also including the analysis for the linearity and specifying the PA intensity boundaries from the recommendations.

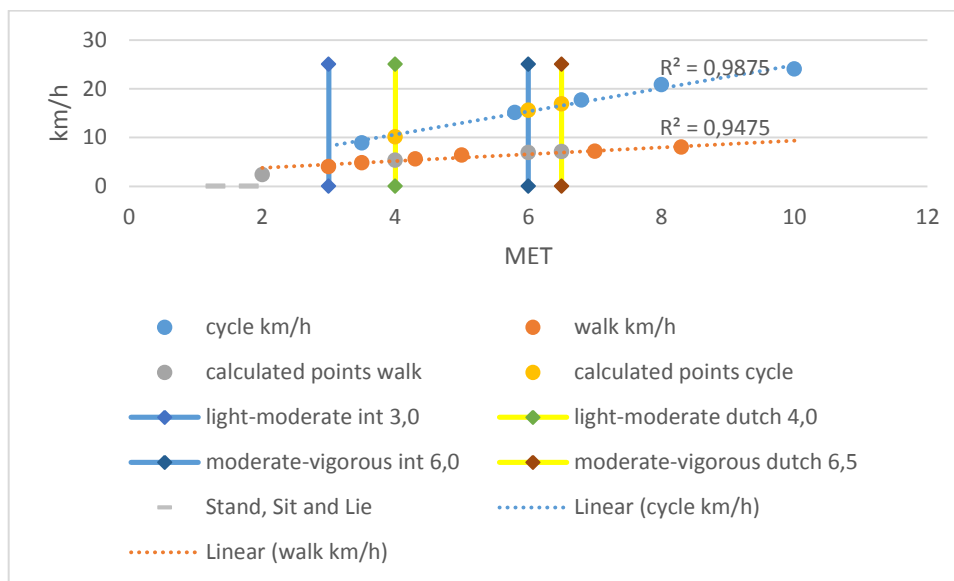


Figure 4 Relevant MET values for postures, walking and cycling according to the WHO PA recommendations (3.0 and 6.0 MET) and the Dutch PA recommendations (4.0 and 6.5 MET). Including linear trendlines, the PA intensity boundaries and interpolated compendium points

### 2.8.2. MLT analysis

In order to use this indirect method, the type of activity must first be determined. As mentioned, this can be done using MLTs. The state of the art consists of a lot of different techniques, such as **Decision Trees**, Neural Networks and **Support Vector Machines (SVM)**. The SVM method has been chosen, after comparing different studies and keeping some constraints in mind, worked out next. At first, for this study the decision tree and the support vector machine were already selected, such that the choice would be smaller. The first series of constraints were the following: the techniques had to work with just a single accelerometer on the waist, be simple enough to be implemented within limited time frame while keeping in mind that it might be used later in smartphones, this was valid for both techniques. When looking at processing power, an estimate was made that the decision tree would need more processing, since it processes the data at the moment itself, while the support vector machine has to process beforehand – with more data –, but during the actual measurement only has to compare the data. When looking at Table 4, it can be seen that for the usage of the accelerometer just on the waist - following Ravi et al - the accuracy of the support vector machine is highest compared to the decision tree. Taking these considerations in account the SVM method has been chosen to implement as MLT.

Publication (number of subjects)	Activities (number of activities)	Accelerometer placements	Inter-subject classification accuracy
Bao and Intille (2004) (20 subjects)	Walking, sitting, cycling, running, vacuuming, folding laundry and more (20)	Shank, thigh, upper arm, wrist and hip	Decision tree (84%) kNN (83%) Naive Bayes (52%)
Parkka <i>et al</i> (2006) (16 subjects)	Lying, sitting, walking, Nordic walking, rowing, cycling and more (8)	Chest and wrist	Decision tree (86%) Hierarchical (82%) Neural network (82%)
Maurer <i>et al</i> (2006) (6 subjects)	Sitting, standing, walking, ascending/descending stairs and running (6)	Wrist	Decision tree (87%) Naive Bayes <sup>a</sup> (<87%) kNN <sup>a</sup> (<87%)
Pirttikangas <i>et al</i> (2006) (13 subjects)	Typing, watching TV, drinking, walking upstairs, cycling and more (17)	Both wrists, thigh and necklace	Neural network (93%) kNN (90%)
Ermes <i>et al</i> (2008) (12 subjects)	Lying, sitting, walking, Nordic walking, rowing, playing football and more (9)	Hip and wrist	Neural network (87%) Hierarchical (83%) Decision tree (60%)
Ravi <i>et al</i> (2005) (2 subjects)	Standing, running, sit-ups, vacuuming, brushing teeth, walking and more (8)	Waist	Naive Bayes (64%) SVM (63%) Decision trees (57%) kNN (50%)
Lester <i>et al</i> (2005) (2 subjects)	Walking, driving, jogging, ascending/descending in an escalator and more (10)	Shoulder	Naive Bayes (67%) HMM (47%) HMM and binary classifiers (95%)
Allen <i>et al</i> (2006) (6 subjects)	Sitting, standing, lying, walking and four postural transitions (8)	Waist	Gaussian mixture model <sup>b</sup> (91%) Hierarchical (71%)

<sup>a</sup> No data were presented on classification accuracy.

<sup>b</sup> Some subject-specific training was used for this classifier.

Table 4 Comparison different classifiers including accuracy [16]

## 2.9. SVM

SVM is one of the MLTs which can be used as a classifier to link specific inputs to certain classes. In the case of this type of activity classification it outputs the type of activity using the accelerometer data as input. The SVM is trained using labelled features calculated from the raw accelerometer data. These features can for instance be the mean, the standard deviation and peak values. During the training these features are coupled to specific classes (for instance walking, cycling, etc.). After this training the SVM can be tested by putting generated features of other labelled data in to see whether they are classified correctly. One of the advantages of SVM is that the support vectors must only be calculated once. The features must be calculated each time, but the processing of these features, the SVM training, is only done once. This SVM training is a high processing task, creating support vectors from the labelled features. Afterwards during the real time classification itself, new features – found with simple calculations - only have to be compared to the already calculated support vectors to find the corresponding class.

### Simple basic theory

The basis of the SVM theory can be explained using Figure 5. It shows an worked out example of a dataset containing objects with 2 features each. It plots these data points (in this example of versicolor and virginica) with one feature on the x-axis and one on the y-axis (mathematically this can of course be extended to more dimensions). These labelled data points are input for the SVM training, which generates support vectors, these are vectors pointing to the separation region. From these support vectors a hyperplane can be calculated to separate the data. In the figure it is clearly visible that the line (hyperplane), calculated using SVM, separates the different classes (versicolor and virginica).

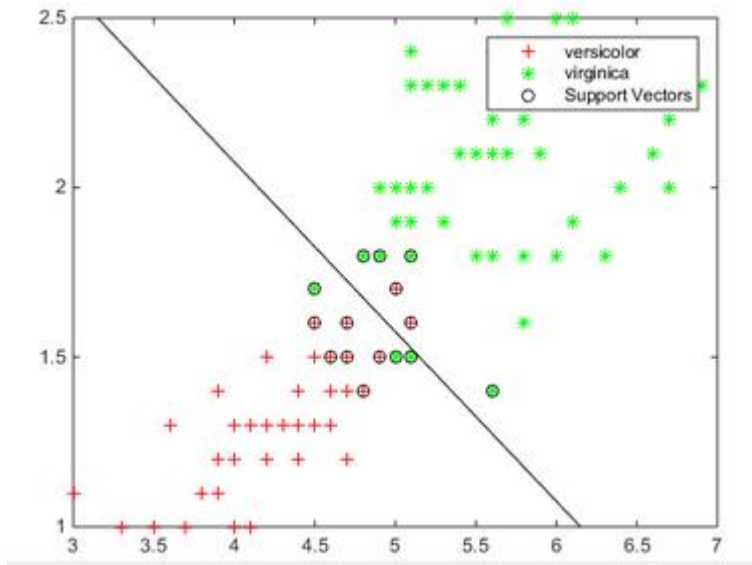


Figure 5 Example SVM with resulting support vectors [17]. The axis values are not important in this case, they do show the differentiation of the data

### Extensive theory in practice

SVM is more extensive than just being able to separate two classes. It can separate any amount of classes with as many features as wanted. There are several functionalities of the SVM that can increase the speed and performance of the classification. These functionalities are the following: the type of kernel function; parameters depending on the kernel function and the choice between one-vs-one or one-vs-all.

- The **kernel functions** are used to extend the dataset to a higher dimension, separating the data further in these extra dimensions, examples of these kernel functions are linear, polynomial and Gaussian Radial Basis functions. All functions have advantages and disadvantages, but due to the limited time for this study, these are not elaborated in this study. In general the most used functions are polynomial and Gaussian Radial Basis functions. With the different functions different parameters are included. An example parameter is the overlap parameter, which gives a value how many data points may coincide.
- Another important choice is **one-vs-one** or **one-vs-all**. In case of one-vs-all, all data points are taken and the support vectors are calculated between one class and all other classes, which are then seen as one class. In this way limited comparisons have to be made to find the current class, but if extra features are added, all support vectors must be recalculated. In case of one-vs-one all classes are compared to each other, this generates more support vectors and therefore more comparison is needed to find the result of the classification, but new features can be added without having to recalculate the previous support vectors. The system can easily be expanded with new data.

### 2.10. Feature calculation

For the SVM features are needed for the classification. Before these features are calculated, some design choices have to be made first, such as the kernel function, the one-vs-one or one-versus-all choice or the new option of the type of window. One feature must be calculated per window.

The features must be calculated per a pre-defined time frame, this can be via normal, sliding, event or activity windows. The difference is visible in Figure 6. It shows a random accelerometer data section and the corresponding windows, depending on the window choice, from top to bottom: normal; sliding; event; and activity.

In case of the **normal window** the time is fixed and the data is separated in equal time windows, spaced evenly. In case of the **sliding windows**, the length of the window is also set, but it is shifted one sample point per time, therefore creating many windows in the time the normal window only has created one window. An



**event window** first determines whether an event has occurred (for instance a change in activity or a sudden change), then a new window is started. The last option is the use of **activity windows**, in this case the pre-processing already separates the different activities (and must therefore already be able to classify the transitions), determining the windows in this manner. Since this study is done with a mobile application in mind, the choice for the normal window has been made, since this option requires no pre-processing and only classification once per given time window.

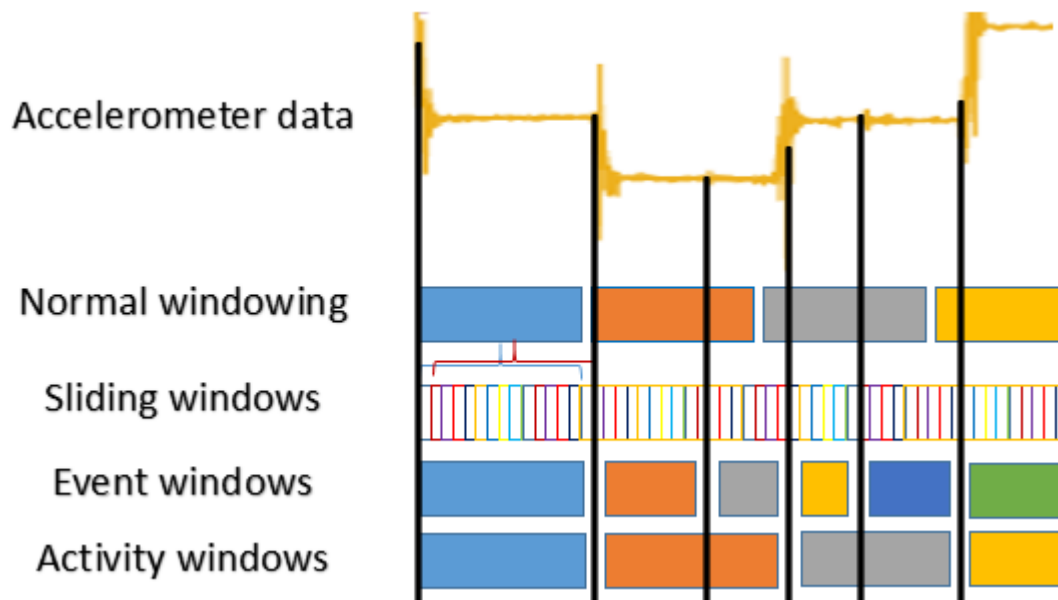


Figure 6 Different windows. At the top a part of accelerometer data is plotted, below it the 4 types of windows: normal, sliding, event and activity. Each new activity window is started with another colour.

Since the device is planned to be used real time, the activities should be recognised as fast as possible. The accuracy of the classification changes with length of the window sizes, but has a maximum at around 6 till 14 seconds[11] for all activities with use of SVM. Further, in order to compensate for the overshoot of the IMA method after changing orientation - due to changing postures - a period of at least twice the IMA's epoch length is chosen. The overshoot after change of orientation is caused by the IMA value being calculated from the sum of the different axis of the accelerometer, these separate parts are changing value and therefore the total value. The count for the duration period is started as soon as the person has finished the transition between activities, which is determined by the observer.

The IMA method has its own feature calculation. It integrates all accelerometer data over a given Epoch length, being 10 seconds. Options for this calculation can be to choose whether the x, y and z axis are integrated separate or combined.

For the SVM method the input is one feature per time window. These features are used to separate the classes and different variations can be used. The different types of features that can be used, can be separated in the time domain, frequency domain and discrete domain. These domains can have further subdivisions. Depending on the type of activity different feature groups (and therefore features) are found to be optimal. A visualisation of this hierarchy and a few examples of features are visible in Figure 7. The most common known **features** are the **mean**, the **standard deviation** and the **dominant frequency or FFT Peak**.



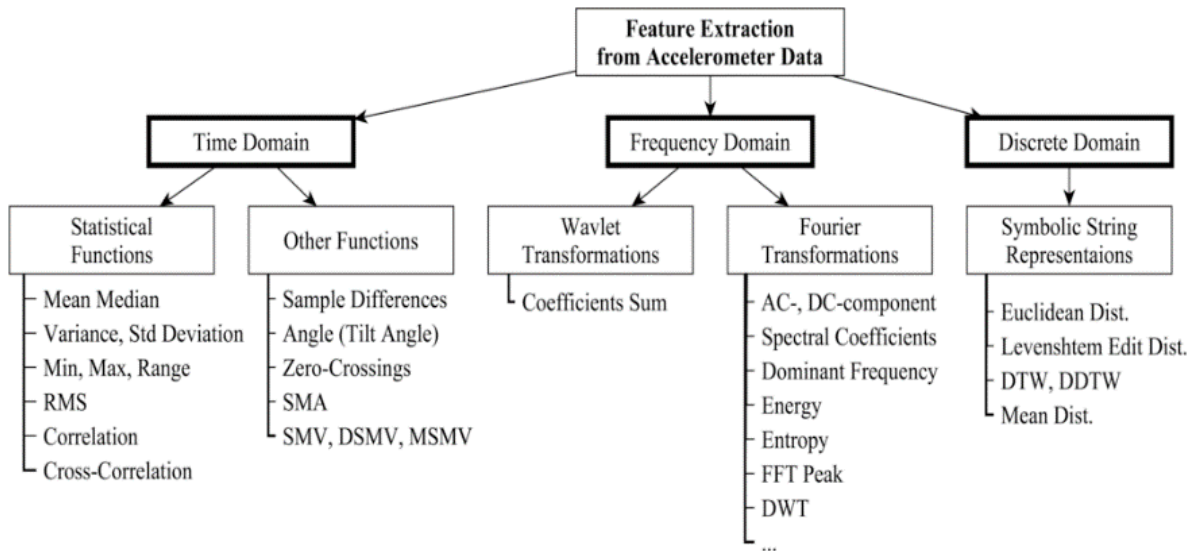


Figure 7 List of possible features for SVM classification [11]

A list of features with common activities that can be classified using those features can be seen in Table 5. Most features have different types of activities that they specifically help to classify.

Feature	Type of activity
<b>Mean</b>	Sitting and standing
<b>RMS</b>	Walking patterns
<b>Energy/entropy</b>	Activities of Daily Living (ADL)
<b>Standard deviation (SD)</b>	Activity classification in general
<b>Signal Magnitude Area (SMA)</b>	Differentiation between activity and inactivity
<b>Signal Magnitude Vector (SMV)</b>	Movement intensity

Table 5 List of features with corresponding types of activities that can be classified best following literature [11]

$$IMA = \int_{t_0}^{t_0+epoch} (|acc_x| + |acc_y| + |acc_z|) dt$$

$$SMA = \frac{IMA}{nr \text{ of samples in time window}} \quad SMV = \int (x^2 + y^2 + z^2) du$$

## 3. Method

### 3.1. Design

A classification study was conducted with office workers and students from RRD and University of Twente. The protocol for this research is given in Figure 8. Using this protocol a gold standard dataset for the classification of activities is set up, consisting of about 1 hour 10 minutes per participant. This dataset is analysed and different classification methods have been applied, validating the self-written algorithm within different participants and between different participants.

### 3.2. Participants

The participating subjects are healthy adults between 20 and 35 years old. In order to get enough data, also enabling cross validation, 15 participants are measured. For comparison within the group: All activities are estimated to be needed for at least 12 participants to achieve a result representing the chosen population and three participants to validate. Other similar research has used 3 till 20 participants [12, 18, 19]

### 3.3. Ethics

All participants were provided with an informed consent, which they signed at the start of the measurements. Approval from the Ethics Committee was not needed, since the study did not focus on health outcomes, but rather on activity. Furthermore it was conducted as a pilot. The informed consent is available in Appendix A: Informed consent.

### 3.4. Materials

#### Location and equipment

A part of the measurements was done in a controlled area. This was inside an air-conditioned room at room temperature. Equipment used is a treadmill and a treatment table. Another part of the measurements is done outside. The parking lot in front of the RRD building is used. It is approximately 180m long and on one part paved except a small part asphalt at about 2/5<sup>th</sup> of the distance and parallel to it about 180 m long with only asphalt. Two city bikes are available with speedometers (Hema bicycle computer Beetle 2) with three equidistant magnets on the wheel to accurately record low speeds. Furthermore for the measurement itself, the participant has an elastic band around his/her waist with the receiver (a reprogrammed HTC Desire) of the accelerometer data. A promove3D accelerometer (Inertia Technology, Enschede, The Netherlands) is attached to the participant using the supplied clip. An analogue annotation tool is used as well as a digital annotation tool by the observer, both are explained later.

#### Protocol activities

For this study, the sedentary behaviour and the PA intensity are to be classified. For this study, the PA intensity was limited to the activities walking and cycling. In order to make this classification system, a dataset with gold standard is needed to train the system and to test it. In order to create this dataset a new measurement protocol has been written to create this dataset.

For the recommendation of the sedentary behaviour, the posture and the intensity should first be classified, in this study however, the intensity of sitting is neglected, assuming all sitting positions to be sedentary behaviour. Sitting can be seen as the in-between step between standing and lying. In order to prevent sitting from being classified as one of those, the different postures stand, sit and lie are chosen to be included in the dataset.

For the PA intensity the activities were limited to walking and cycling. From the compendium, different speeds have been selected, corresponding to the borders between the light, moderate and vigorous PA intensity. This has been worked out before in Figure 4. From this plot the boundary speeds between the different PA intensity levels can be retrieved. The values are shown in Table 6.

	WHO Low-Moderate (3 MET)	Dutch Low-Moderate (4 MET)	WHO Moderate-Vigorous (6 MET)	Dutch Moderate-Vigorous (6.5 MET)
Walk (km/h)	4.0	5.3	6.9	7.1
Cycle (km/h)	XX	10.1	15.6	16.9

Table 6 Summary walking and cycling speeds corresponding to the different activities. The XX corresponds to having no valid value.

- Walking at four different speeds was sufficient to find all four boundaries, not all are used however. The 6.9 and 7.1 are expected to be too close to each other to accurately classify the difference. They are combined in the speed of 6.9 km/h. Additional to these three walking speeds, also a upper and lower limit is chosen to be able to see whether the behaviour continues at these speeds. As a lower limit 2.4 km/h had been chosen (corresponding to the round number 1.5 mph) and as an upper limit 7.5 km/h has been chosen since it was expected that people were able to still walk this speed without starting to jog. This information of the lower and upper limit might be used to see that and the intensity boundaries are not at the same point where the MET corresponding to the activity changes suddenly or where the activity starts to show other feature behaviour. The speeds chosen for the walking protocol are therefore 2.4, 4.0, 5.3, 6.9 and 7.5 km/h.
- Cycling only has three different speeds in table 5. It also has at 3.5 MET a speed of 8.9 km/h, but since this is not an intensity boundary (of the recommendations between light and moderate), it is ignored. Again in this case the 15.6 and 16.9 km/h are expected to be too close to each other to get a significant difference, they are therefore combined in a value in between the original values, namely 16.0 km/h. Again upper and lower limits are looked at. The lower limit in this case is however left out since the estimate was that for significantly lower speeds, the chance of falling over became too large or that people might start swinging sideways to keep the balance. In order to keep a safe experiment with constant behaviour, so no swinging added, the lower limit is left out. The upper limit is chosen to be 19.0 km/h, which is expected to be significant higher than the 16.0 km/h and is a common cycling speed. The speeds chosen are thus 10.1, 16.0 and 19.0 km/h.

When looking at types of activities, an estimate has been made that it is also possible to use the different transitions between activities to help classify the different activities coming afterwards. Therefore also relevant transitions are looked for between the different activities. Not relevant transitions for instance are a transition from cycling to lying, since this will not happen unless someone tumbles over or passes through other transitions like standing, etcetera.

### Protocol type of activities

The needed activities have been determined, but there are still several ways to gather the data, for instance for the cycling part: using a home trainer for measurements; a bike outside with orders to cycle a certain speed for measurement; and letting subject cycle themselves and register at that moment the speed and classify accordingly. As T. Bastian said: "Automatic identification methods should be developed using data acquired in free-living conditions rather than data from standardized laboratory activity sets only, and their limits carefully tested before they are used in field studies." [20]

A balance is chosen between lab- free-living conditions, while still being able to gather the needed data, efficiently and accurately.

- For the postures this consisted of a standardised surrounding and a standardised timing and order of activities. The material consisted of a treatment table, which was height adjustable, which could be used as a bed and as a chair. The subjects were asked to stand, sit and lie down various times and were allowed to sit, lie or stand as they liked. This variation for instance enabled people to shift their weight from one leg to another each activity, therefore including small alterations subjects might make in their postures. The choice of one treatment table has been made for its simplicity.

- For the walking two choices have been made. Since the speed of walking is expected to be difficult to control at low speeds as well as hard to keep constant, a treadmill is used. However, in order to represent free-living, a treadmill is not preferred, therefore also an outside part, with a level surface, is used. The treadmill data gives accurate results, which might make the classification easier. The outside walking data gives more free-living results; the speed however is still imposed.
- For the cycling part the choice has been to only create a dataset of cycling outside, since cycling on a home trainer leaves the balancing part out, therefore not representing the free-living as much as wanted. The cycling is therefore just done outside, but the cycling speed however is again imposed.

Due to these choices four different parts for a protocol have been designed; 1) postures, 2) walking on a treadmill, 3) walking outside and 4) cycling outside.

Furthermore when looking at the amount of repetitions needed; for comparison within participants, all activities are estimated to be needed at least three times per participant to classify these activities and one more of this activity to validate it within that person, to reduce errors due to measurement errors.

In order to implement the previous findings about the types of activities, the different measurement setups and the requirement of using at least 3 measurements per activity, a protocol has been designed. This protocol – with timing which is explained later – is visible in Figure 8. In order to get as many variations in sitting, lying and standing, the posture cycle has been repeated 4 times instead of 3 times. Due to the order all different transitions are incorporated. For the walking part on the treadmill also the transitions between the different consecutive speeds are recorded. For the walking part outside a difference is made between the first two cycles and the third cycle to incorporate the transitions. This is explained later. The same is valid for the cycling part.

### **Protocol timing**

In order to get the required amount of data for the classification, all activities must be performed at least three times. This is also shown in the protocol. The time period of all activities was said to be a multiple of 6 and at least 2 times the IMA epoch length of 10 seconds, therefore being 24 seconds per activity. These 24 seconds deliver at least 3 full windows of 6 seconds if the start is slightly too soon or too late, therefore satisfying the requirement of having at least 3 measurements per activity. The transitions between these activities are wanted to be as short as possible, but normally executed (not hurried) to get representable transitions. The transitions have been tested in practice to get an idea of the time needed per type of transition: sit to stand, sit to lie, changing speed of the treadmill etc.

This timing is also implemented in the protocol in Figure 8. At the top of the Figure a summary of the time is given, stating that a total measurement takes 1 hour and 8 minutes.

[illegible]

### Gold standard: digital annotation tool

participant what activities it has done before. The same reason for the choice of the body worn sensors above camera surveillance and questionnaires is applicable here. The camera system is too expensive to set up, plus the analysis must be done manually afterwards, even though the advantage of registering all data. The questionnaire in this case is not possible for the accuracy of seconds needed (on a total of 24 seconds per activity). The chosen method in this case is a new developed application that enables an observer to label on the spot. It used direct observation, directly logged everything to a PDA, was portable and due to the implementation had an accurate time registration and synchronisation. When looking at validity, the quality of the data is dependent on the observer; how accurate he is able to recognise the activities; his timing; and whether he presses the right buttons. When looking at feasibility; an observer has to be present – which he would have been all the same, in order to explain the experiment – and directly creates the final file which needs no further processing (in some cases minor edits of single activities).

Specific advantages are the possibility of adjustments, such as a slightly longer or shorter measurement time; the order of the activities is not strict, if this order is changed, it is directly annotated; errors can directly be corrected by marking this (on a separate sheet) and restarting that activity with correct annotation.

This application is designed specific for this experiment and from now on referred to as the *digital annotation tool*. It creates labelled timestamps at the moment the observer selects the next activity.

The tool is run on a reprogrammed smartphone (HTC Desire). It shows one screen with a timer and all the different activities as single buttons. One button is selected at all times, indicating the current activity as a check for the observer. The observer every time presses the corresponding button at the start of the activity or at the start of a transition.

The timer gives a beep every time it has reached a multiple of 24 seconds – which was the standard time for the activities – in order to help the observer get a correct timing. If a new activity is selected, the timer resets. A screenshot of this application is visible in Figure 9. Due to an old version, the application still displays old speeds, namely 13.7 and 19.3 km/h for the cycling part, but these are later used as respectively 16.0 and 19.0 km/h. Tread and C sit stand for standing on treadmill, to specifically label standing on a treadmill after an activity and C sit to label sitting still on a bike. These activities are later excluded from the data, since they were thought of being too specific.

The digital annotation tool has been designed by the researcher for this study and was programmed by MSc Dennis Hofs.



Figure 9 Screenshot of digital annotation tool

The data is saved locally, including time stamps in milliseconds since the start of the day. A sim card is inserted, in order to make time synchronisation possible for the measurement and the control device. After a complete measurement of a participant, the data must be removed from the device using a cable, or via a wireless connection to a secure database.

#### **Gold standard: analogue addition and valid labels**

Besides this digital annotation tool also annotations on paper are made to mark special events, such as repositioning of the measurement device, wrong timing and wrong button presses on the digital annotation tool and failed activities, for instance due to interference of passing cars. The speeds of the outside activities are also marked, namely the minimum and maximum speed. The structure of the analogue annotation is visible in Appendix B: Analogue annotation tool. It also helps the observer to know which activity is to come.

This analogue annotation paper is mostly used to mark the (in)correct digital annotations of activities. If for instance an activity was performed shorter than the described time of 24 seconds it can be noted that the activity was invalid (or that it just was too short). Manual corrections afterwards can be used to remove or change these data if needed, since this is not possible real-time during the measurement.

In this study a validity check has been designed. The different measurement sections have been chosen to have different focus points for the validity of the data, due to the different characteristics of these parts.

- For the posture section, the activities are seen as valid if the time is minimal 24 seconds and the participant has not moved significantly during the measurement.
- For the treadmill section, the activities are seen as valid if the time is minimal 24 seconds and the transitions have been consistent.
- For the walk section outside, the activities are seen as valid if the time is minimal 24 seconds, it has been on a straight part of the road, (not in a curve) and the speed has not differed more than 0.5 km/h from the provided speed.

- For the cycle section outside, the activities are seen as valid if the time is minimal 20 seconds, it has been on a straight part of the road, (not in a curve) and the speed has not differed more than 1.5 km/h from the provided speed.

In all sections focus is put on the fact of a stable activity. The classification is designed to be made using the pure activities, and other behaviour, like transitions are noted differently, such that they can be added afterwards if needed.

### 3.5. Procedure

All participants got an elastic waistband with a HTC Desire attached (as receiver of the data) and a promote3D accelerometer (Inertia Technology, Enschede, The Netherlands), which was already clipped in vertically, which he or she could position on the side of the hip on either side of the body.

The accelerometer was set to gather tri-axial accelerometer data at a frequency of 100 Hz. All accelerometer data points are saved, but only every 1000 points a timestamp is added. At this same time, the IMA value is calculated. The digital annotation tool saves the selected label together with the timestamp at the moment of the selection.

The observer operates the digital annotation tool: he checks the analogue annotation tool to see the activities to perform and after the time period (24 seconds) of the previous activity, he asks the participant to perform the next activity, at this same time he selects the transition button and the timer restarts. The participant continues to the next activity and if the observer judges the activity stable, he presses the corresponding activity to start the labelling of this activity.

In case of the treadmill part, after the transition is started, the observer changes the speed and after the transition is done, he starts the labelling of the next activity.

In case of walking or cycling outside; he starts the transition, changes his imposing speed (explained later) to the next activity and if he judges it stable, he starts the next activity.

The speeds inside were imposed using the treadmill to give the exact speed. For the speed measured and imposed to the participants in the outside parts, the observer had to walk and cycled along, while monitoring the speed.

The observer walked next to his bike (with speedometer) for the walking part or cycled on it for the cycling part. He imposed the required speed by moving at this speed. The participant was instructed to follow the speed of the observer.

Before the measurement the observer and participant moved at the different speeds to get an idea of the speeds and the accelerations needed to change between the speeds as soon as possible but with as less overshoot as possible. Using these test rounds the observer could give hints to the participant before the real measurements to walk and/or cycle faster or slower.

If an activity was started, the observer tried to find the constant speed and if achieved started the measurement on the digital annotation tool, after 24 seconds (and the transition had been started) the minimum and maximum speed were noted, this was done for two rounds (from minimum speed to maximum speed and back to minimum is one round). The third round the speeds were put in sequence to get the transitions between the different speeds. In the curves of the road, the measurement was set to transition and after the curve the previous speed was restarted to get a valid labelling.

After the treadmill part the participant was offered some water, due to the effort and inside climate.

After all the measurements were done the equipment was taken back and the data was saved from the devices to the computer, after which the devices were reset.



### 3.6. Data processing Matlab protocol

With the dataset with gold standard complete, the classification can be done. A schematic of the total processing procedure can be seen in Figure 10. All the steps are worked out below. Since this Matlab code is an intermediate step towards the final results, problems that might be encountered are stated and solutions given if found.

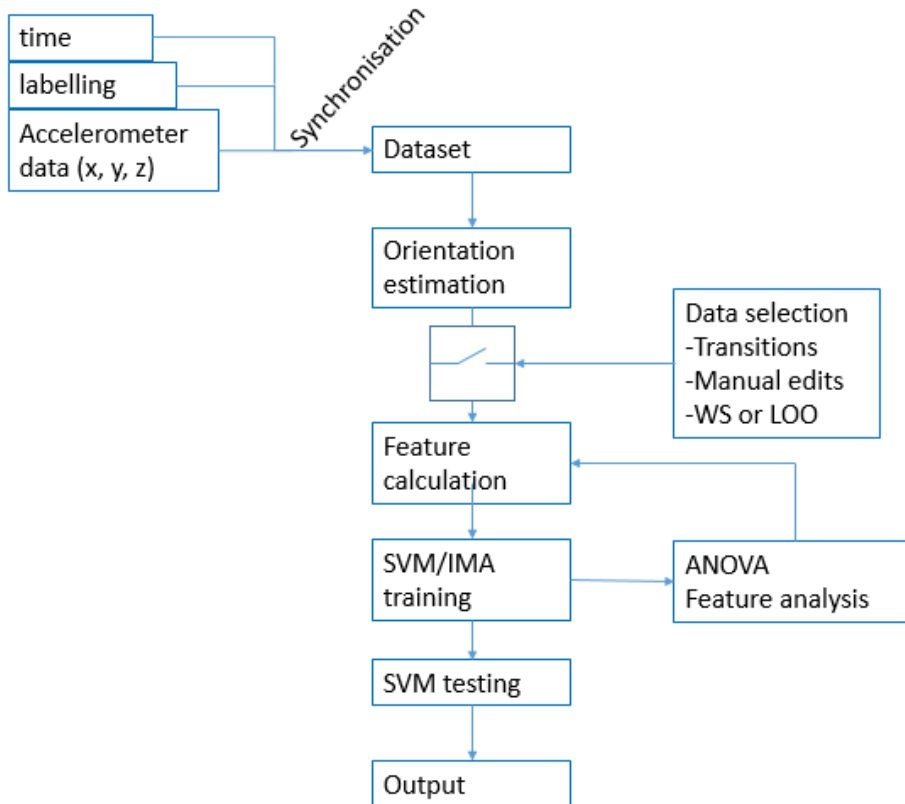


Figure 10 Flow chart of Data processing

**Synchronisation:** The raw data from the measurements must first be imported and converted to a standard dataset. The structure of this dataset is visible in Figure 11. It shows the option to add more data, such as gyroscope data, magnetometer or to be added data. The activitymatrix is created from the digital annotation tool. The start and stop time combined with each activity are given in this variable. The time variable contains the time stamps of the accelerometer. The timenew variable is a vector calculated from the time stamps, taking the start and end point and interpolating using the sample frequency. The acc variable represents the accelerometer data synchronised to this new time axis, adding NaN at the spots where the time between time stamps contained a gap in accelerometer data, due to wireless connection errors. The IMA can directly be imported from the device. The activity variable is the values of the activitymatrix assigned to the new time axis. Since the activitymatrix is generated on a different device than the time used here, a check has been built in to change the activity to invalid at points where these different time frames do not overlap.

datacomb.PP02	
Field ▲	Value
gyr	'nvt'
mag	'nvt'
activitymatrix	185x3 double
time	499x1 double
timenew	499233x1 double
acc	499233x3 double
IMA	499x1 double
activity	499233x1 double

Figure 11 Example of structure of data after importation and synchronisation into Matlab

In the Matlab script all activities have been converted to numbers. The renaming of the activities can be seen in Figure 12. The renaming has been done in groups, keeping the postures together at 1-3, the walking activities at 11-15 and the cycling activities at 21-23. Other assignments used are: 0, which is a temporary feature indicating a label still needs to be assigned; 40, indicating a transition and 50, indicating a label that will be ignored per definition in the calculation.

0	1	2	3	11	12	13	14	15	21	22	23	40	50
Unassigned	Stand	Sit	Lie	Walk 2.4 km/h	Walk 4.0 km/h	Walk 5.3 km/h	Walk 6.9 km/h	Walk 7.5 km/h	Cycle 10.1 km/h	Cycle 16.0 km/h	Cycle 19.0 km/h	Transition	Mislabelled data

Figure 12 Renaming table for assignment of numbers to activities.

The minimum time for an activity was set at 14 seconds for activities to be counted as valid and a minimum time of 1 second was set for transitions, otherwise it was recognised as a typing error and therefore marked as 50. One of the imported datasets is plotted in Figure 13. From left to right the measurement can be seen to contain the different measurement parts. First the postures, then walking on the treadmill with no rest in between, walking outside and cycling outside. As a detail, number 10, corresponding to tread and number 20 corresponding to C sit can still be seen in the plot. Furthermore at the beginning and at the end the label is 50 due to synchronisation between the different device and as well 50 at a few moments in between where wrong data was found (too short periods, therefore recognised as wrong button presses) and corrected.

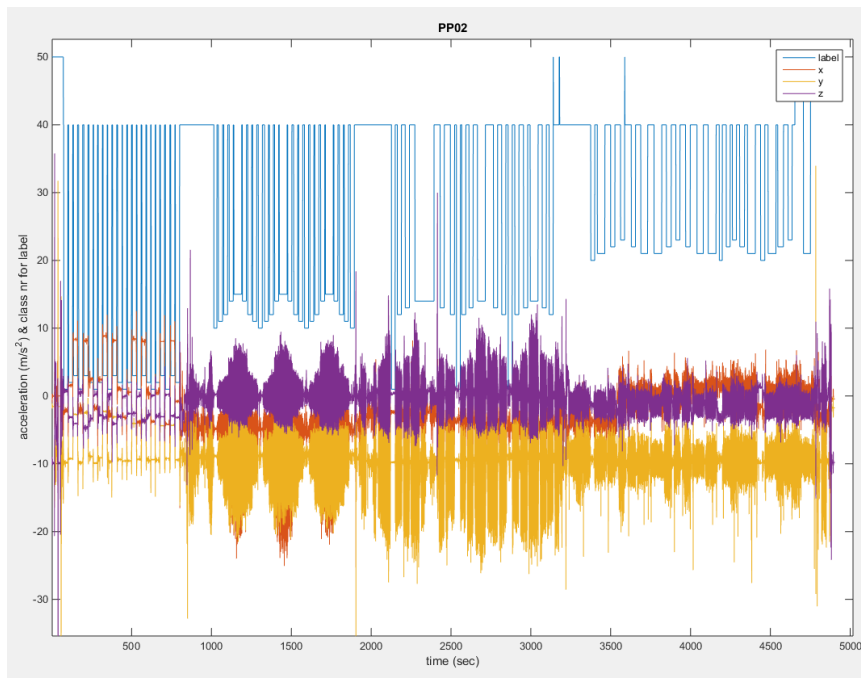


Figure 13 plotting of imported dataset of participant 2 into Matlab, including synchronisation. The different measurement sections can be recognised in the data, till about 800 sec the postures (with a shift in orientation of the x, y and z for each posture), till 2000 sec the walking on the treadmill part (with an increase in fluctuations for higher speeds), till 3300 seconds the walking part outside (with also an increase in fluctuations for higher speeds, but also breaks between the speeds) and till the end the cycling part which appears to be quite constant. The label is seen to correspond to the labels assigned in Figure 12.

**Orientation Estimation:** The participant was free to choose where to place the accelerometer: left or right hip. Furthermore it can be mounted in 4 different orientations. In this part the data can optional be corrected for the left and/or right hip and if needed for the different orientations. The standard orientation of the data is chosen to be the following (when looking forward): x-axis forward-backward, y-axis up-down, z-axis right-left. The average can be taken to find the y-axis, from the gravitational force. In this research, this is always the same, since the device was clipped in with the same orientation each time.

The side of the hip has not been registered for all participants, only PP14 was noted to have placed the accelerometer on the left hip. When looking at the data a potential difference between the left and right hip can in most cases be seen in the data, especially when looking at the cycling part. When cycling, the person is on average sitting with the hip slightly bent back. The x-axis registers this angle. In Figure 14 in the top row participants are selected with similar behaviour in the cycling part as participant 14 who had the accelerometer on the left side of the hip. In the bottom row a few participants are shown which seem to have had the accelerometer on the right side of the hip, when looking at the cycling part (most right set with label 20-23). 14, 05 and 01 appear to have an average of x above the average of z, while in the other cases this is the other way around. In the last case the division is not any more the division between top and bottom row, so there must be more variables. Further analysis is needed, but not executed in this study. For the dataorientation, if it is used, the average of the x is found and inverted (multiplied with -1), to make sure the data is equal (at least at the cycling part).

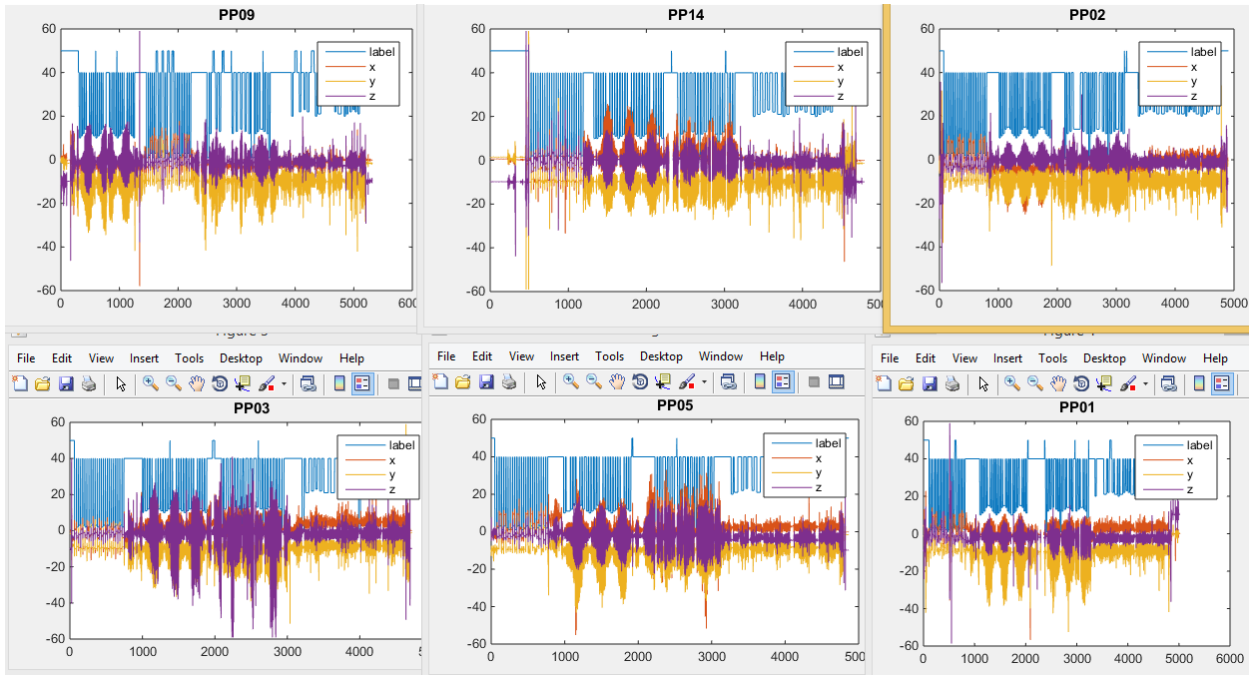


Figure 14 Different participants, pointing out potential left-right hip analysis. PP 14 is left hip. On the x-axis the time is plotted in seconds and on the y axis, the acceleration in  $m/s^2$  is plotted for the x, y and z and the labelling is plotted to its own assigned numbering

**Data Selection:** Incorrect data can be removed from the dataset, for example start-up and stopping regions and wrong data labelling. Furthermore transitions can be chosen to be implemented in the training or excluded. Manual changes from the analogue annotation tool are due to lack of time only partly implemented, but wrong button presses etc. have been corrected. In this Matlab code two definitions of transitions are used: the ones from the gold standard and the windows containing multiple activities. The transitions from the gold standard are defined too broad to be used, they include the wanted transitions, but also include not wanted data, such as the walking between activities. The transitions calculated during the windowing are potentially more interesting, since these transitions show the labels containing two or more activities. If these are then labelled it is interesting to see whether the first, the second or another activity is assigned. This sort of adds noise to the signal and enables to check the performance of the classification. Due to overlap of these two definitions all transitions are combined and mostly contain wrong data and are therefore not used in general.

**Switch:** The data must be split in training and testing parts depending on whether Within Subject (WS) analysis is used or Leave One Out (LOO) analysis. For the within participant selection, a complete participant is loaded and a part ( $1/10^{th}$ ) – chosen with a random function – is left out, the remainder of the data is trained and afterwards tested with the left out selection. This is repeated 10 times with the same results from the random function, in order to compare all 10 parts with the rest. For the LOO, one complete participant is left out, then all the other participants are combined and this combination is trained. This trained combination is then tested with the participant left out.

**Feature Calculation:** Features must be calculated from the raw accelerometer data, for instance: mean, peak value and standard deviation. For the features first a window is needed. The length of the windows can be set to 6 or 10 seconds to represent respectively the length for the highest literature accuracy or the 10 seconds epoch length of the previous method. The average sample frequency is calculated, in practice it is on average 101.79 Hz, despite the setting of the device of 100 Hz. Using this average and the given window length, the amount of samples is calculated needed for one window. For each window the feature could now be calculated. Specific analysis of the created features is worked out later in Features (4.6).

**SVM training:** For the actual training of the SVM the used variables are: one-vs-one and as Kernel function the “Gaussian Radial Basis”, with as extra parameters a 'BoxConstraint' of  $2e^{-1}$ . One-vs-one since it makes

clear which activities are misclassified to which activities and the Kernel function since in other mobile applications, similar to this type of activity recognition also used this function.[21] The outcomes are the trained models, representing the support vectors calculated.

*SVM testing:* In this block the testing data is classified with the trained data and the left out training data. This comparison is a standard implemented function of Matlab.

*ANOVA and post-hoc:* The analysis of the calculated features combined with the corresponding labels can help to find useful features. More details about the method are visible in Statistical Analysis (3.8). The data is first trained using SVM as worked out in the paragraphs before. This can separate the data using the WS or LOO method, depending on the preferred analysis; finding features for WS, LOO or both. The resulting data is used to determine the ANOVA. The ANOVA is checked manually whether it satisfies the requirement of  $p < 0.05$ . The results are directly plotted by the Matlab function. For detailed analysis, the results are also used as input for the post-hoc analysis using the Bonferroni method with  $\alpha < 0.001$ . Some extra code has been written to find all features that are satisfying the condition of  $\alpha < 0.001$  and summarise this amount of useful features per calculated class pairs (sit versus stand, sit versus lie, etc.). A result from the ANOVA and post-hoc analysis for within subject is visible in Figure 15. A list with the summary of the amount of features satisfying the Bonferroni analysis is visible in Appendix D: Detailed results.

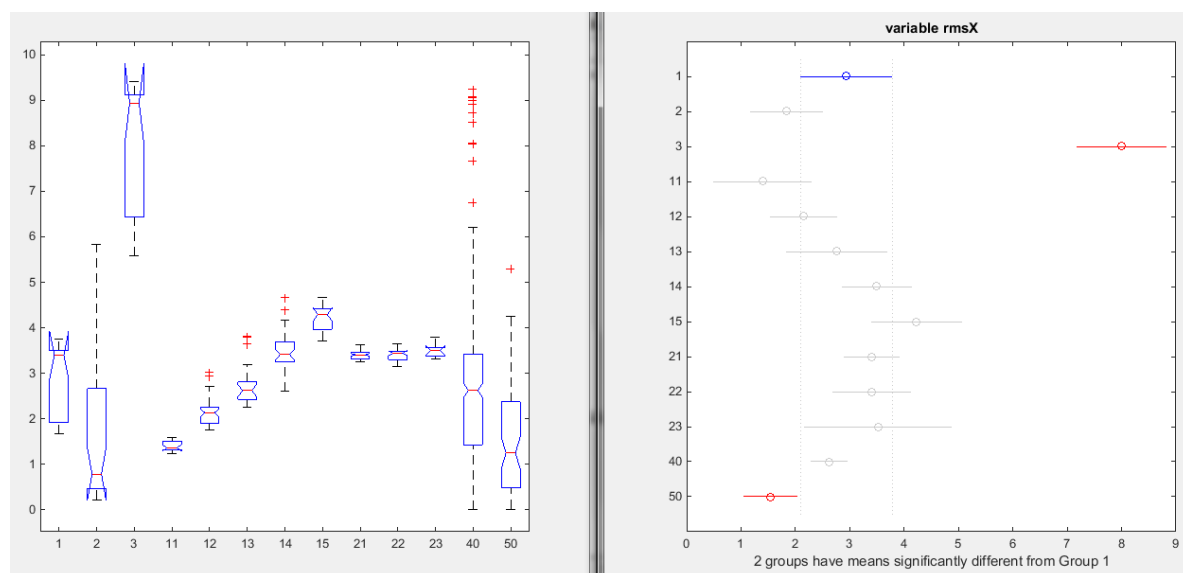


Figure 15 ANOVA and post-hoc analysis, on the left side ANOVA, showing detailed information and on the right side the post-hoc analysis showing the different classes that can be separated and to what degree. On the left side on the x-axis the different classes are plotted with their own numbering and on the y-axis the rms of the acceleration in  $m/s^2$ . On the right plot, the y and x-axis are switched

*IMA classification:* In order to calculate the results of the IMA settings, the IMA data is first needed. From the measurement system already IMA values are imported, but due to changes in synchronisation, these are no longer valid. The original c-code used in the measurement system is now applied to the new synchronised system to recalculate the IMA values. Due to limited time, the original method of the ROC curves to find the thresholding is not used, but the IMA values are seen as a feature as input of the SVM. The SVM kernel function has been set to linear in order to find the calculated thresholding values, but this has not been worked out completely to retrieve these values. The parameter used for the training is a 'BoxConstraint' of  $2e^{-1}$ , like for the SVM. Furthermore a hack had to be implemented to prevent synchronisation errors, since the amount of features had to be the same as the amount of labels. This hack adds NaN to make them of equal length, but sometimes a shift of 5 labels occurs, corresponding to a shift of  $5 \times 6 = 30$  seconds shift. It is not analysed which datasets are correctly implemented. The results seem still valid, but the accuracies can be higher if this shift would not have occurred.

*Output:* The results are processed and visualised using confusion matrices and different tables. The outcomes will be worked out in Outcome Measures (3.7).

### **3.7. Outcome Measures**

The primary outcomes were the dataset with gold standard and the classification system with the corresponding results. The classification results consisted primary of: the accuracies in confusion matrices for the IMA and SVM method; the results using different settings, such as 6 or 10 seconds; as well as the choice of different features that have been calculated.

Secondary outcomes were the quality of the dataset measurements. Furthermore, the classification data was processed using a custom self-adjusted script (Matlab, R2014b (8.4.0.150421)). The program trained support vector machine models and checks their validity, as well as calculated the IMA values and the validity of these.

### **3.8. Statistical Analysis**

One way Anova is used to check the different features used to determine the classification using the SVM method ( $p < 0.05$ ). A post hoc analysis using Bonferroni is used for further analysis to find good separating features. A value of  $\alpha < 0.001$  is used to separate the 15 features.

## **4. Results**

In the following section the results of the measurement protocol and the results of the IMA and SVM classification are stated.

### **4.1. Participant Characteristics**

15 Healthy participants have been measured: 9 men, 6 women. The participants were 22-35 years old with length ranging from 1,60m till 1,95m. Three participants were measured in two sessions (postures and treadmill at one go and later the walking and cycling part). One was measured in reversed order, starting with the walking and cycling outside and ending with the postures and walking inside. One had the posture and treadmill part switched.

### **4.2. Calibration**

The speedometers of the bikes have been calibrated. With both bikes with speedometer at the same time a complete round has been made. The speeds were within 0.5 km/h from each other and the final distance travelled at the end was the same. The round was approximately 400m. A single speedometer has been calibrated using a distance wheel (device that gives a “click” each meter travelled): the wheel was used to set out exactly 170m, at this distance the speedometer also precisely indicated 170m.

After all the experiments, the participant bike had travelled over 65 km. The total distance needed for the treadmill section was three times 340m, therefore approx. 1km/participant.

### **4.3. Dataset**

All the measurements have been labelled digitally and analogue. The classification has been done digital, but in the analogue version mistakes and extra details have been noted. For each activity (class) the amount of valid labels has been noted as well as the total amount of valid labels that were needed. For the speeds of walking and cycling outside the minima and maxima have been noted. The speeds were maximal 10% off compared to the supposed speeds, there are no overlapping speeds. The results are visible in Table 7. A side note must be made that the speeds given are the speeds of the observer, the speeds of the participant might have been slightly different, but following the method this difference was seen to be small following the observer, also using the calibration round at the beginning of the measurement. Walking speeds of 0.5 km/h off compared to the supposed speed also have been marked as invalid and cycling speeds of 1.5 km/h off compared to the supposed speed also have been marked as invalid.

ACTIVITY	AMOUNT OF VALID LABELS (24 SECONDS)	AVERAGE SPEED (KM/H) (OUTSIDE)
SIT	114 / 120	---
STAND	117 / 120	---
LIE	111 / 120	---
WALK 2.4 KM/H	88 / 90	---
WALK 4.0 KM/H	172/180	3.8-4.4
WALK 5.3 KM/H	86/90	4.9-5.8
WALK 6.9 KM/H	128/135	6.4-7.2
WALK 7.5 KM/H	86/90	---
CYCLE 10.1 KM/H	153/180	9.5-10.8
CYCLE 16.0 KM/H	64/90	14.0-16.9
CYCLE 19.0 KM/H	21/45	18.4-20.1

Table 7 Results from manual labelling, such as valid time windows and reached speeds

During the outside measurements if it was noted that an activity was measured wrong, it was redone in order to get the data correctly. For the first few participants sometimes the curves were included in the measurements, this has not always been annotated, so this cannot be separated any more, except by choosing the first eight (to be certain) participants and remove their cycling data. In order to keep enough classification data this action is not performed. To prevent this unclear annotation after the unknown amount of participants, the measurements were set to transition per definition in the curves.

The amount of valid time windows used in the Matlab calculation can also be found per category. The total amount of valid windows is the same for both Leave One Out and Within Subject methods if all participants are used and all data. For instance the sum of sitting and lying is 389 valid windows of six seconds. The valid windows from the protocol are given per participant per 24 seconds. For the processing the used windows are all six seconds, therefore this is chosen as variable to compare the results. The amount of valid labels of Matlab is divided by the amount of participants to find the average number of windows of six seconds. The results from the protocol are multiplied with four for the conversion from 24 to six seconds. For this comparison see Table 8.

ActivityNr	1	2	3	11	12	13	14	15	21	22	23
(Practice) valid windows (6 seconds)	565	389	389	299	601	295	501	325	576	378	196
(Practice) valid windows (6 seconds) per participant	37,6	26	26	20	40	19,6	33,6	21,6	38,4	25,2	13,2
(protocol ) valid windows (6 seconds) per participant	32	32	32	24	48	24	36	24	24	24	24

Table 8 Comparison amount of valid time windows from protocol with practice.

#### 4.4. Matlab script

A structure of the Matlab scripts shows the different sub-functions used to calculate the needed data. Twelve separate function files have been used and another 7 internal sub-functions. Internal functions are marked with brackets []. Using parameters all execution options can be set from the ExecuteTest file. The createdatacombsinglePP function is used to import the accelerometer data, the IMA values and the labelling data from the original files and synchronise this data. The resulting data can be visually checked using plotting. The orientation estimation is only implemented on the lowest level in DetOrient, since it changes



the data. A part of the data selection also occurs in this section by marking mislabelled data dependent on time and marking manual corrections from the observer. The switch is implemented by choosing to execute the LeaveOneOut or WithinSubject function, splitting the data corresponding. Inside these functions the features are calculated via the function calculate\_features2. After this calculation the Anova and multicompare can be executed, but for separate analysis this is also possible to execute directly through [AnovaPlotting], called from ExecuteTest. Inside LeaveOneOut and WithinSubject after datasplitting and feature calculation the SVM training is done and directly tested with the splitted data. With parameters a selection can be made how much data must be used for this testing. The results are given in a confusion matrix with corresponding total accuracy of the diagonal. The structure of the Matlab scripts is visible in Figure 16. It shows the hierarchy used in the scripts, showing which functions are called from which functions.

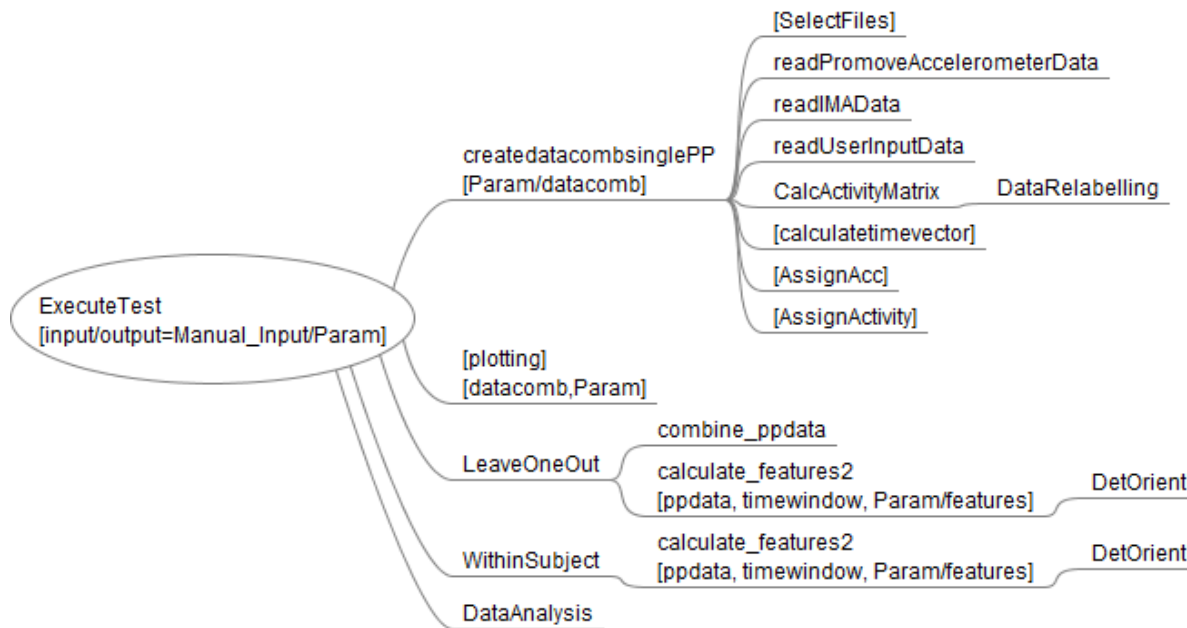


Figure 16 Matlab script structure, sub-functions – parts of other script files - are marked with brackets []

The Matlab scripts have been designed such that from one window all wanted settings can be executed, such as a visual plotting of the raw data, executing the LeaveOneOut and Within Subject analysis with the different settings and performing data analysis with different settings, such as ANOVA plotting and post hoc analysis and potential automatic processing of those data.

Some layout and design choices have been summed below

- The manual corrections of the annotation have not all been implemented, a short visual check has been done to remove wrong annotation data from wrong button presses.
- Automatic layout for Figures has not been implemented yet, such as correct axis labels.
- All classes are added hard coded to prevent size mismatch if not all classes are in a dataset. In case of new classes, these must be added manually.

The resulting work and authors are summarised in Table 9. Most of the code has been written as automatisation and the corresponding debugging. Some datasets contained no cycling activities, having caused the program to crash, but these problems are mostly solved. The program can now be run analysing all participants with WS and LOO analysis for different settings, such as six or ten seconds windows, IMA or SVM analysis, etc.



FUNCTION	WORDS	ORIGINAL AUTHOR	REMARKS
CALCACTIVITYMATRIX	150	D. van der Valk	Own script
CALCULATE_FEATURES2	842 (204)	T.M. Tönis	Functioning the same, some additions and foolproof/debugging made for automatisisation.
COMBINE_PPDATA	89	T.M. Tönis	No changes
CREATEDATACOMBSINGLEPP	719	D. van der Valk	Own script
DATAANALYSIS	119	D. van der Valk	Own script
DATARELABELLING	374	D. van der Valk	Own script
DETORIENT	90	D. van der Valk	Own script
EXECUTETEST	1126	D. van der Valk	Own script
IMAINERTIA	149 (114)	T.M. Tönis	Resampling added for synchronisation.
LEAVEONEOUT	572 (250)	T.M. Tönis	Changed for Automated analysis + needed debugging/making foolproof for automatisisation
READIMADATA	221	T.M. Tönis	No changes
READPROMOVEACCELEROMETERDATA	281	T.M. Tönis	No changes
READUSERINPUTDATA	315	T.M. Tönis	No changes
WITHINSUBJECT	534 (217)	T.M. Tönis	Changed for Automated analysis + needed debugging/making foolproof for automatisisation
<b>SUMMARY</b>	5581	T.M. Tönis has written the separate scripts and D. van der Valk has combined them in one program and added different features.	The scripts were used as a basic structure around which were executed and extended using the rest of the code.

Table 9 Overview of division design Matlab code, the amount of words to give an idea of the size of the script and "large changes?" to indicate the size and influence of the changes made to the scripts made by other authors. The number within brackets in words represents the original amount of words.

#### 4.5. Data synchronisation & mislabelling data

Since a wireless communication system is used, connection failure was expected. At these points the time scale continues linearly, but in the gaps NaN is added to the accelerometer data. The labelling is synchronised by connecting to the time scale using the corresponding time stamps. Mislabelled data, for instance if within a second 2 buttons are pushed are marked wrong labelled (50). A selection of the measured data from participant 01, showing gaps and mislabelled data is visible in Figure 17.

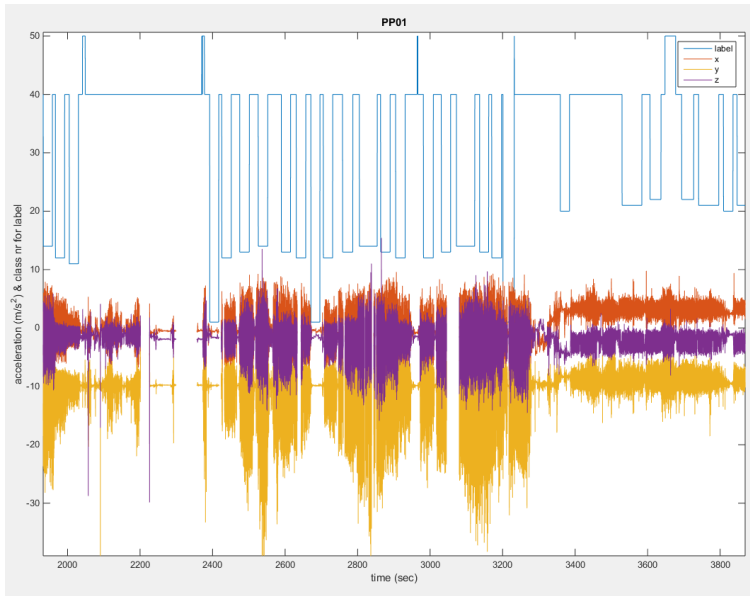


Figure 17 Selection of data of one participant plotting the accelerometer data and the labelling versus time. Gaps in the accelerometer data represent the NaN added for synchronisation and the peaks in labelling to 50 represent mislabelled data.

#### 4.6. Features

Different features have been analysed and tested. Orientation independent and orientation dependent features have been used. The features were first tested using ANOVA analysis to see whether they met the requirement of  $p < 0.05$ , an example of a feature not satisfying this requirement is visible in Figure 18. It had a  $p$  of 0.06 while other features often have a  $p$  between  $1e^{-50}$  and  $1e^{-130}$ .

After the ANOVA analysis was done, further specific analysis was done using the post hoc analysis. The post-hoc analysis gave plots. These plots had the value of the comparing feature on the x-axis and the different classes on the y-axis. The mean was marked as well as the 95% confidence interval. By selecting a class (or group) the classes that are significant different were marked by the program. A selection of these plotted features is visible in Table 10. These features are each time calculated within one participant, except if specifically stated.

A few orientation independent features looked at are:

- mean of the sumvector
- standard deviation (SD) of the sumvector
- root mean square (RMS) of the sumvector
- dominant frequency detection using Fourier analysis

This last frequency feature has not been implemented due to miscalculations.

For the orientation dependent features first the orientation is determined (x forward-back, y up-down and z left-right) and then the features are calculated:

- mean in y
- SD of x,
- RMS of x & y
- difference between x & y (x-y)
- peak to peak values of x & y

The z-direction is found to be varying too much between subjects, therefore lowering the classification accuracy, this can also be seen in Figure 20.

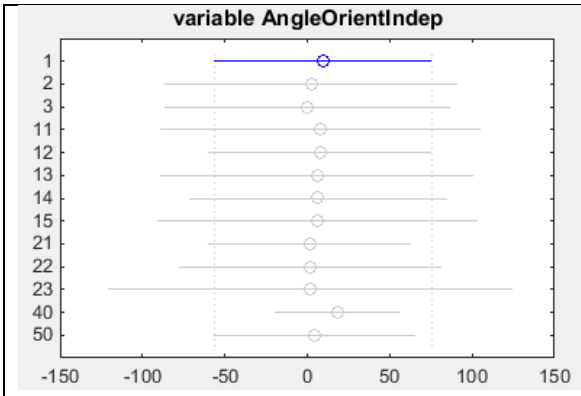


Figure 18 Example of a feature not satisfying an ANOVA of  $p < 0.05$ . on the x-axis the relation between the x and y direction of the accelerometer data and on the y-axis the own numbering of the classification

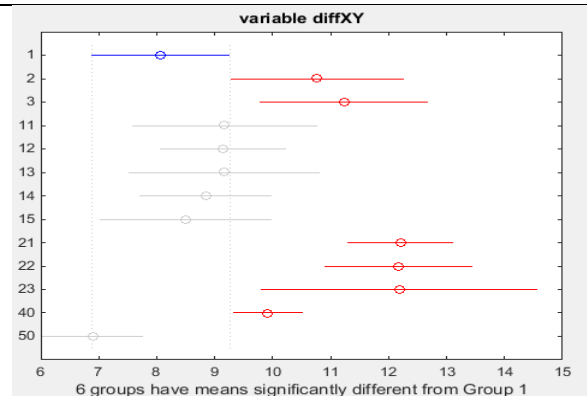


Figure 19 post-hoc analysis showing separating behaviour for standing within the postures and further separating the cycling from the walking. On the x-axis the difference is given in  $m/s^2$  and on the y-axis the own numbering of the classification is given

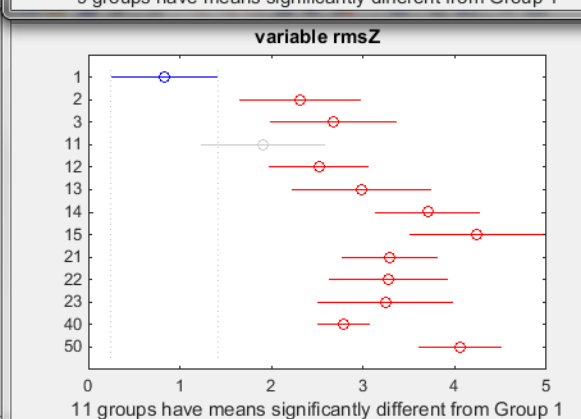
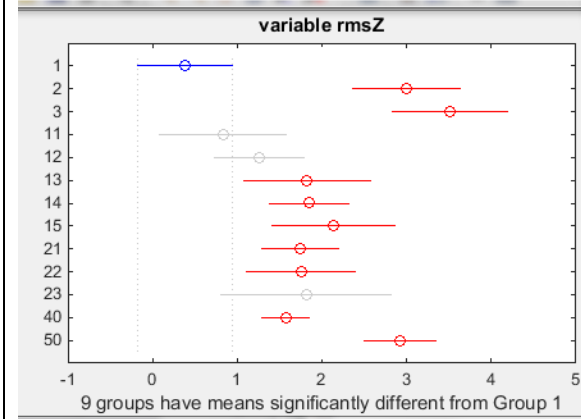
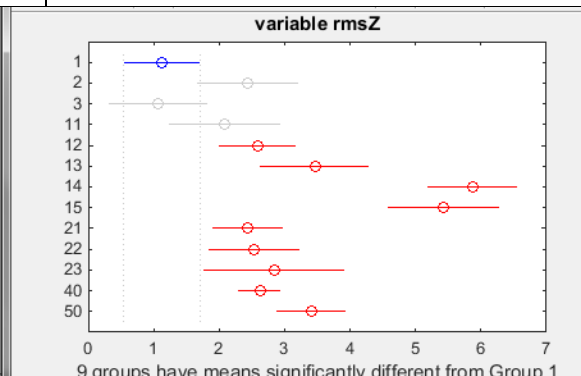
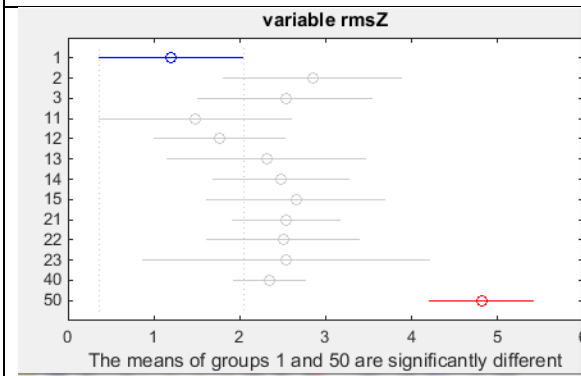


Figure 20 Example of a feature with significant difference between different participants. Each plot corresponds to the same feature, but a different participant. On the x-axis the RMS is given in  $m/s^2$  and on the y-axis the own numbering of the classification.

Table 10 Different features, changing in validity and accuracy, with specifying captions.

#### 4.7. Statistical analysis

The results of the ANOVA and post-hoc method have been compared. In case of the Within Subject analysis the ANOVA plotting corresponds quite well to the post hoc (Bonferroni) plotting. This can be seen in Figure 21 In case of the Between Subject analysis the standard deviation of the post hoc analysis has decreased, therefore showing a better separation between the different classes, (which is not true in practice), this is visible in Figure 22.

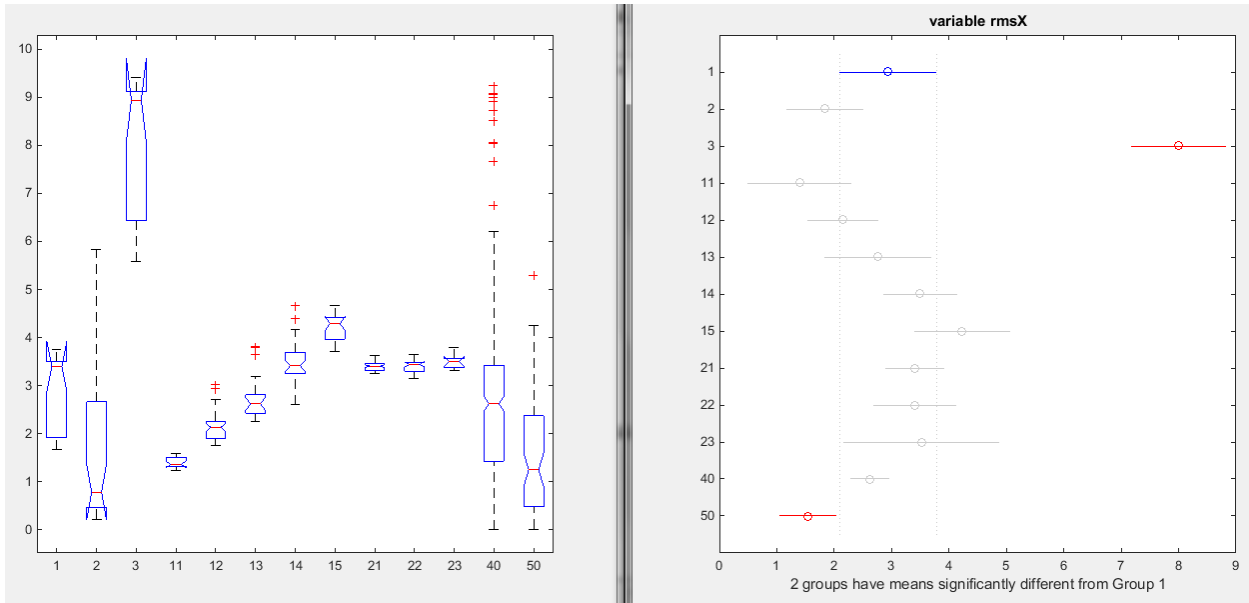


Figure 21 Analysis of Within Participant 1, ANOVA output on the left side and post hoc (Bonferroni) output on the right side. On the left side on the x-axis the different classes are plotted with their own numbering and on the y-axis the RMS of the acceleration in  $m/s^2$ . On the right plot, the y and x-axis are switched

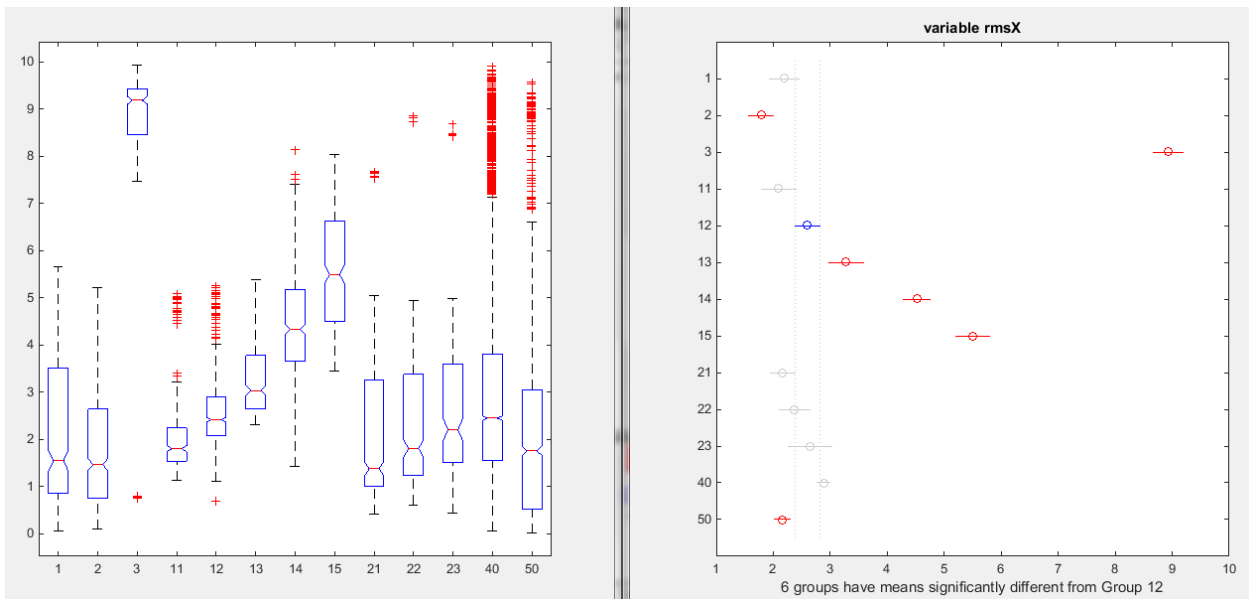


Figure 22 Analysis of Leave Participant 1 out, ANOVA output on the left side and post hoc (Bonferroni) output on the right side. On the left side on the x-axis the different classes are plotted with their own numbering and on the y-axis the rms of the acceleration in  $m/s^2$ . On the right plot, the y and x-axis are switched

#### 4.8. Type of activity

For the type of activity classification, the Matlab script has generated confusion matrices to show the results per labelled class with respect to the predicted classification. Below in Table 11 and Table 12 the results are visible of the two SVM classification techniques: Within Person and Leave One Out. On the vertical axis are the correct labelled classes and on the horizontal axis are the predicted classes. All activities on the diagonal of the matrix are classified correctly. All activities that have been predicted “50: Miss labelled” had connection problems and were assigned during classification to this value. The transitions are left out.

Using different colours, sections have been marked, in the top right corner of each section a numbering (as superscript) is added as legend. It can be seen that most activities are located around the diagonal within the own section: postures together, walking together and cycling together (1). There are two sections showing

the overlap between walking and cycling (2). Furthermore most activities have been predicted as standing a few times (3), the same for cycling with 19.0 km/h (4). The previous mentioned assigned value of 50 is also marked (5).

LABEL\PRED	1	2	3	11	12	13	14	15	21	22	23	50
1: STAND	<b><u>511</u></b>	27	3 <sup>1</sup>	5	6	2	2 <sup>3</sup>	0	1	3 <sup>3</sup>	5 <sup>4</sup>	0
2: SIT	25	<b><u>322</u></b>	23	0	0	0	0	0	0	0	1	0
3: LIE	4	30	<b><u>351</u></b>	0	0	0	0	0	0	0	1	0
11: W2.4	0	0	0	<b><u>275</u></b>	10	1 <sup>1</sup>	0	0	2	5 <sup>2</sup>	6	0
12: W4.0	0	0	0	4	<b><u>566</u></b>	17	1	0	0	1	2	1 <sup>5</sup>
13: W5.3	0	0	0	0	2	<b><u>272</u></b>	12	3	0	0	1	2
14: W6.9	4	0	0	0	4	6	<b><u>428</u></b>	41	0	1	8	1
15: W7.5	0	0	0	0	0	0	19	<b><u>306</u></b>	0	0	0	0
21: C10.1	0	0	0	1 <sup>2</sup>	0	0	0	0	<b><u>433</u></b>	92	29 <sup>1</sup>	0
22: C16.0	5	0	0	14	6	0	1	0	32	<b><u>203</u></b>	108	0
23: C19.0	0	0	0	4	13	3	0	0	0	40	<b><u>121</u></b>	0
50: MISLAB.	0	0	0	0	0	0	0	0	0	0	0	<b><u>0</u></b>

Table 11 Confusion matrix SVM WS Method 15 participants, excluding transitions, with a total accuracy of 86.0%. The diagonal is marked bold and underlined. Groups of predicted data are marked with different colours and a superscript at the top right of each area. The superscripts are used to mark the different sections, 1:postures, walking and cycling; 2 walking/cycling misclassification; 3 standing misclassification; 4 cycling 19.0 km/h misclassification; and 5 connection error misclassification.

LABEL\PRED	1	2	3	11	12	13	14	15	21	22	23	50
1: STAND	<b><u>426</u></b>	107	8 <sup>1</sup>	10	7	3 <sup>3</sup>	0	0	0	2 <sup>3</sup>	2 <sup>4</sup>	0
2: SIT	121	<b><u>214</u></b>	34	1	0	0	0	0	0	1	0	0
3: LIE	90	47	<b><u>248</u></b>	0	0	0	0	0	0	0	1 <sup>4</sup>	0
11: W2.4	0	0	0	<b><u>262</u></b>	17 <sup>1</sup>	0	0	0	4	9	7 <sup>2</sup>	0
12: W4.0	0	0	0	19	<b><u>450</u></b>	39	13	0	0	0	70	1 <sup>5</sup>
13: W5.3	0	0	0	0	20	<b><u>217</u></b>	42	11	0	0	0	2
14: W6.9	4	0	0	7	8	21	<b><u>330</u></b>	122	0	0	0	1
15: W7.5	0	0	0	0	0	3	131	<b><u>191</u></b>	0	0	0	0
21: C10.1	0	0	0	27	1 <sup>2</sup>	0	0	0	<b><u>402</u></b>	104	21 <sup>1</sup>	0
22: C16.0	3	2	0	23	23	1	0	0	67	<b><u>121</u></b>	129	0
23: C19.0	0	0	0	15	31	3	2	0	1	36	<b><u>93</u></b>	0
50: MISLAB.	0	0	0	0	0	0	0	0	0	0	0	<b><u>0</u></b>

Table 12 Confusion matrix. SVM LOO Method all 15 participants, excluding transitions, with a total accuracy of 66.71% The diagonal is marked bold and underlined. Groups of predicted data are marked with different colours and a superscript at the top right of each area. The superscripts are used to mark the different sections, 1:postures, walking and cycling; 2 walking/cycling misclassification; 3 standing misclassification; 4 cycling 19.0 km/h misclassification; and 5 connection error misclassification.

#### 4.9. IMA

The IMA results found using the IMA values as input for the linear SVM model, keeping in mind that the IMA values might be shifted due to timing, are visible in Table 13 and Table 14.

Using different colours, sections have been marked, in the top right corner of each section a numbering (as superscript) is added as legend. It can be seen that most activities are located around the diagonal within the own section: postures together, walking together and cycling together (1). There are two sections showing the overlap between walking and cycling (2). Furthermore most activities have been predicted as standing a few times (3). The walking section is a few times predicted as a posture (4). In case of the LOO analysis sitting is classified most often of the different postures (5). The areas are less strict defined compared to the SVM and are less grouped. The amount of misclassification is higher and the postures and walking are now mixed, which were separated mostly for SVM. In case of LOO the standing and lying postures are, when looking at significance, not classified at all. Apparently the classification seems to have put all the postures together in one group, since no difference can be made.

LABEL\PRED	1	2	3	11	12	13	14	15	21	22	23
1: STAND	<b><u>225</u></b>	148	160 <sup>1</sup>	9 <sup>4</sup>	1	1	1 <sup>3</sup>	0	11	2	7 <sup>3</sup>
2: SIT	119	<b><u>93</u></b>	149	7	0	0	0	0	3	0	0
3: LIE	89	102	<b><u>187</u></b>	8	0	0	0	0	0	0	0
11: W2.4	0	0	0	<b><u>239</u></b> <sup>1</sup>	0	0	0	0	46	8	6 <sup>2</sup>
12: W4.0	0	1 <sup>4</sup>	0	6	<b><u>395</u></b>	10	2	0	27	28	123
13: W5.3	0	2	0	0	0	<b><u>274</u></b>	10	2	1	0	3
14: W6.9	2	1	3	7	6	28	<b><u>332</u></b>	111	0	1	2
15: W7.5	0	0	0	0	0	0	14	<b><u>311</u></b>	0	0	0
21: C10.1	0	0	0	171	12 <sup>2</sup>	0	0	0	<b><u>303</u></b> <sup>1</sup>	50	19
22: C16.0	0	3	1	59	33	2	0	0	76	<b><u>117</u></b>	78
23: C19.0	0	0	0	16	39	8	0	0	7	35	<b><u>76</u></b>

Table 13 Confusion matrix, WS total 6 sec 15 participants IMA with a total accuracy of 57.6% The diagonal is marked bold and underlined. Groups of predicted data are marked with different colours and a superscript at the top right of each area. The superscripts are used to mark the different sections, 1:postures, walking and cycling; 2 walking/cycling misclassification; 3 standing misclassification; and 4 walk/posture misclassification

LABEL\PRED	1	2	3	11	12	13	14	15	21	22	23
1: STAND	<b><u>0</u></b>	545 <sup>5</sup>	1	4	1	1 <sup>3</sup>	0	0	6	2	5 <sup>3</sup>
2: SIT	0	<b><u>365</u></b>	0	3 <sup>4</sup>	0	0	0	0	3	0	0
3: LIE	0	383	<b><u>0</u></b>	3	0	0	0	0	0	0	0
11: W2.4	0	15	0	<b><u>20</u></b>	1 <sup>1</sup>	0	0	0	178	42	43 <sup>2</sup>
12: W4.0	0	1	0	0	<b><u>125</u></b>	96	4	0	37	67	262
13: W5.3	0	2	0	0	64	<b><u>154</u></b>	24	44	1	1	2
14: W6.9	0	11	0	0	2	110	<b><u>126</u></b>	234	2	1	7
15: W7.5	0	0	0	0	0	1	120	<b><u>204</u></b>	0	0	0
21: C10.1	0	17 <sup>5</sup>	0	9 <sup>2</sup>	0	0	0	0	<b><u>350</u></b>	132	47 <sup>1</sup>
22: C16.0	0	8	0	0	7	1 <sup>2</sup>	0	0	123	<b><u>88</u></b>	142

Table 14 Confusion matrix, LOO IMA analysis 6 sec with a total accuracy of 34.5% The diagonal is marked bold and underlined. Groups of predicted data are marked with different colours and a superscript at the top right of each area. The superscripts are used to mark the different sections, 1:postures, walking and cycling; 2 walking/cycling misclassification; 3 standing misclassification; 4 walk/posture misclassification; and 5 sitting misclassification

For the implementation of the IMA, the SVM method has been used, using the IMA value as the only feature. The SVM has been implemented with the linear kernel in order to prevent the values from being changed and higher dimensions being used, leaving the path of the original IMA analysis methods. In Figure 23 the results of an IMA analysis for one participant WS is given, showing the IMA values to have been between 0 and 100. Since only one feature is used, it is the same as looking at one single value, or in other words, thresholding is applied in this method. Using thresholding can be useful to see the different walking speeds, which seem to increase linearly, but cycling cannot be clearly distinguished. The different postures cannot be separated.

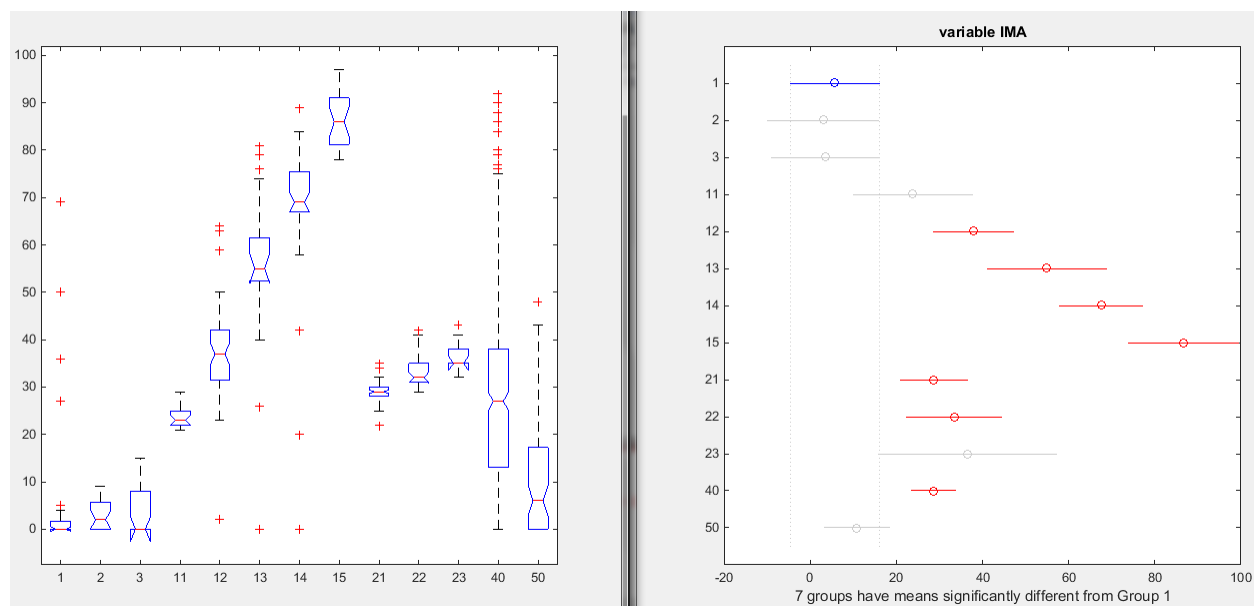


Figure 23 ANOVA and post hoc analysis of the IMA method of one participant with WS. On the axis the classes are plotted versus the IMA values, changing between 0 and 100. On the left side on the x-axis the different classes are plotted with their own numbering and on the y-axis the IMA in  $m/s^2$  On the right plot, the y and x-axis are switched

#### 4.10. Combining the results

The classification can be done with different settings. These settings can be used to find (with the current system) the optimal overall result. Different types of features can be chosen, only a selection of the data can be used, the IMA or SVM method can be used. A list of the most used settings is the following:

- In- or exclude all transitions
- A selection of data used for the classification.
  - Merging the different walking activities to one walking class and merging the different cycle classes to one cycling class.
  - Remove all classes, except for the walking data.
  - Remove all classes, except for the cycling data.
  - Remove all classes, except for the posture data.
  - Keep all classes, but only merge the cycle data to one cycling class.
- In- or exclude orientation dependent features in the classification.
- The IMA method or the SVM method.
  - Using different Kernel functions: linear or Radial Gaussian Function
- A time window of six or ten seconds.

- Calculate the sitting accuracy; this can be done for any of the above choices.

The selection of data enables a closer inspection of certain classes. In order to analyse these different settings, a script was made to put all the different settings after each other, creating one execution to calculate most settings. This run takes about 6800 seconds, corresponding to a small 2 hours. Below the different results will be shown.

The differences between the WS and LOO for both IMA and SVM for 6 seconds and for the different selections of data are shown in Table 15. Different results are visible:

- WS always has a higher accuracy than LOO (except for sitting for IMA).
- The standard deviation(SD) of SVM WS is always lower than SVM LOO (except include transitions).
- The SD of IMA is sometimes higher for WS, sometimes for LOO
- SVM always has a higher accuracy than IMA.
- SVM WS always has a lower SD than IMA WS (except include trans and sitting accuracy)

SETTINGS \ METHOD	SVM WITHIN SUBJECT (%±SD)	SVM LEAVE ONE OUT (%±SD)	IMA WITHIN SUBJECT (%± SD)	IMA LEAVE ONE OUT (%± SD)
<b>NORMAL</b>	86.1±7.8	61.1±12.6	62.8±14.2	35.6±9.8
<b>INCLUDE TRANS</b>	55.0±9.1	41.3±6.9	27.2±7.4	15.2±3.8
<b>MERGE WALK AND CYCLE</b>	97.0±4.1	81.7±12.2	68.9±12.5	61.4±11.2
<b>EXCLUDE ORIENTATION</b>	85.0±9.8	64.1±14.2	XXX	XXX
<b>JUST POSTURE</b>	94.0±10.2	83.2±14.4	41.0±14.4	24.6±11.8
<b>JUST WALK</b>	92.2±8.9	69.4±18.6	91.0±11.4	54.7±18.6
<b>JUST CYCLE</b>	71.4±15.5	56.1±20.0	65.5±19.4	59.2±14.8
<b>ONLY MERGE CYCLE</b>	94.1±5.3	70.7±13.4	66.1±14.6	44.6±13.2
<b>SITTING ACCURACY (NORMAL)</b>	90.7±19.6	44.7±34.7	14.6±17.7	23.2±11.0

Table 15 Overall performance of different settings, SVM versus IMA and WS versus LOO (time window is 6 seconds, Kernel function is rbf for SVM and IMA). IMA has no exclude orientation option, such that this is left out using XXX.

In order to validate different time windows, six and ten seconds are both analysed for IMA and SVM, the results are visible in Table 16. The following statements can be made:

- A period of six seconds always has a higher accuracy than the ten second windows for SVM.
- A period of six seconds always has a higher accuracy than the ten second windows for IMA, except “merge walk and cycle”, “just cycle” and the sitting accuracy.
- The SD is always higher for the ten seconds window than for the six seconds window, except for include trans of IMA.



SETTINGS \ METHOD	SVM WITHIN SUBJECT (%±SD) 6 SECONDS	SVM WITHIN SUBJECT (%±SD) 10 SECONDS	IMA WITHIN SUBJECT (%±SD) 6 SECONDS	IMA WITHIN SUBJECT (%±SD) 10 SECONDS
<b>NORMAL</b>	86.1±7.8	84.5±11.2	60.7±14.1	59.8±17.0
<b>INCLUDE TRANS</b>	55.0±9.1	50.0±10.8	26.3±7.5	20.7±6.9
<b>MERGE WALK AND CYCLE</b>	97.0±4.1	95.5±6.4	60.8±12.4	65.9±17.5
<b>EXCLUDE ORIENTATION</b>	85.0±9.8	83.4±12.5	XXX	XXX
<b>JUST POSTURE</b>	94.0±10.2	91.5±13.3	39.5±13.6	36.3±22.2
<b>JUST WALK</b>	92.2±8.9	90.3±14.2	90.7±12.2	89.7±14.7
<b>JUST CYCLE</b>	71.4±15.5	66.7±28.0	65.4±19.6	65.7±21.8
<b>ONLY MERGE CYCLE</b>	94.1±5.3	92.2±8.6	64.8±15.0	63.5±18.3
<b>SITTING ACCURACY (NORMAL)</b>	90.7±19.6	79.5±35.2	11.9±18.4	15.2±28.7

Table 16 Comparing overall performance of system with 6 versus 10 seconds for within subject for SVM with kernel rbf and IMA with a linear kernel. IMA has no exclude orientation option, such that this is left out using XXX

If these overall performance measurements are recalculated, the percentages change so far noticed maximal ±1%.

In literature ambulation is often 93% accurate and posture is 99.63% accurate[11]

The sitting accuracy has been calculated for the normal setting and for the only merge cycle setting, the results are visible in Table 17. It can be seen that the accuracies are for multiple settings about the same value, especially when looking at the high SD.

SETTINGS	SVM, WS, 6 SEC ( %±SD )
<b>MERGE CYCLE AND WALK</b>	89.9±21.7
<b>NORMAL</b>	90.7±19.6
<b>ONLY MERGE CYCLE</b>	87.9±23.6

Table 17 Different sedentary behaviour accuracies for different settings

## 5. Discussion

The original goal was to create a dataset with gold standard, to create a classification system and to find the best features and the resulting overall classification performance. The dataset has been made, finding some improvements on the way; the classification system has been made as expected, with of course some side notes; the features have had less attention, so no extensive research of good features has been done; but the overall performance was despite the simple features used performing at an interesting accuracy. Quite a few interesting facts have been found that might be used or further analysed in later research on this topic.

### 5.1. Dataset

When looking at the speeds that were used during the outside measurements, it could be seen that these corresponded quite well to the supposed speed. The speeds found for the outside experiments show that even though the speeds had to be provided manually, it was still possible to achieve an accuracy with only a 10% error of the speed. This accuracy enabled the different speeds to be kept separated and overlap of the data was prevented.

When looking at Table 8, it can be seen that the amount of labels from the protocol is not equal to the amount used in Matlab. In general there are less valid labels found in the Matlab script than there should be following the protocol, this is mainly caused by synchronisation and the fact that transitions are left out in the script. The synchronisation is visualised in Figure 24. The Dataset window shows a part of a protocol and below potential alignment of the Matlab windows is visible. If the time windows in Matlab are not exactly synchronised to the original dataset window, only three time windows of six seconds are recognised within the time window of 24 seconds. If namely in a Matlab window (6 seconds) two different activities are seen (in the case of the red blocks in the window) they are marked as transition and not used in the Matlab calculation. The differences are explained below:

- With the transitions removed, a part of potentially valid data is removed, therefore reducing the total amount of valid labels. There are however not exactly  $\frac{3}{4}$  of the amount left, since activities often are slightly longer than the 24 seconds, enabling four Matlab windows to fall within this larger window, therefore a higher amount than this  $\frac{3}{4}$  is found.
- Standing has more valid labels, since in the protocol between different walking cycles also standing is labelled.
- If cycling was measured shorter than 24 consecutive seconds, the measurement of that activity was done again, without removing the old data, so if this was for instance 18 seconds, three extra windows are found.

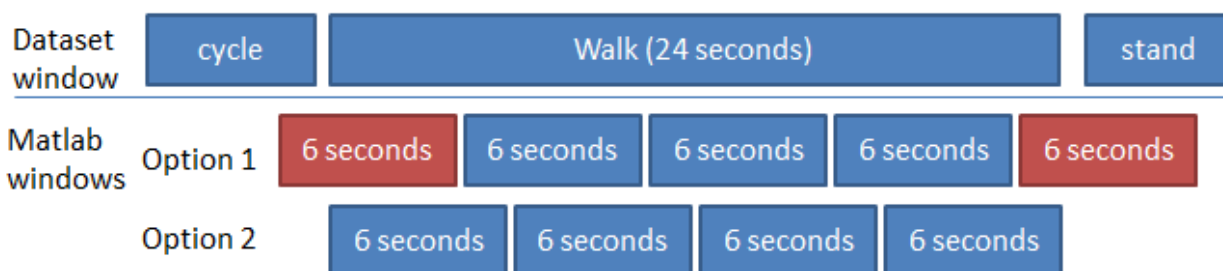


Figure 24 Showing the synchronisation of the Matlab windows compared to the windows from the dataset

For all the measurements the observer had to participate and join in order to show the correct activities. Since he was joining all the activities he could also have used a measurement system, therefore creating a larger (labelled) dataset at the same time. Focus must of course be paid to the difference in performance the observer might deliver due to the observation. During the walking he had to bring the bike, therefore changing his walking pattern and during the postures and treadmill section he did not participate, but during the cycling he could have joined in the measurements, keeping in mind that he often had to cycle with one hand due to the observation, but it is expected that this should not influence the measurement at the waist.

## Transitions

Due to the double definition of transitions, valid data can be removed. Due to the manual labelling of transitions extra class changes are found in the data and therefore more windows are removed from the classification, since they are now extra marked as transitions. If in the dataset a difference can be made between transitions and not labelled data, less data has to be removed for the classification.

## 5.2. Data processing

### Matlab execute file

The Matlab script has been designed such that all types of analysis can be performed from one screen with certain robustness build in. This robustness build in implies protection against problems due to inconsequent data, for instance cycling analysis with measurement data that does not contain cycling data.

### Datasynchronisation and datalabelling

When looking at the data synchronisation and data labelling part, it seems to be working fine. However when looking in detail, for instance the plot in Figure 17, it is visible that in case of large gaps due to wireless connection errors, the labelling is continued. This continuation of labelling implies that less data is available to classify that activity. A threshold should be found to find the minimum amount of data for correct analysis. If the gap is larger than this threshold, the corresponding label could be set to 50/wrong labelling, since the classification might give problems.

### Features

The currently used features are simple calculations requiring low processing power (minima, maxima, mean, etc.). The more processing power requiring possibility of frequency analysis has not been implemented yet, but seems promising when looking at the plotted acceleration data itself. When looking at Figure 25, the repetition of peaks of for instance the y-axis can be seen during this zoom of the walking activity. The expectation is that the classification will be more accurate if this frequency behaviour is included.

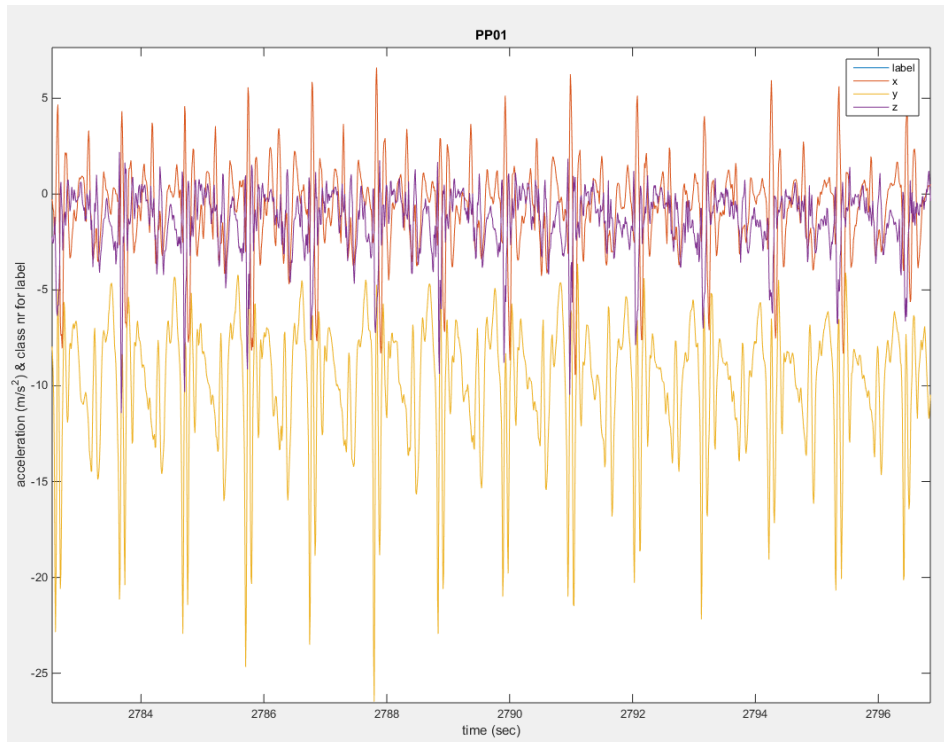


Figure 25 Zoom of walking data, pointing out the frequency behaviour that might be used in future analysis. The time is given in seconds after the start of the measurements

When looking at the used features, different conclusions can already be drawn:

- Different classes can be separated clearly for WS analysis.
  - The postures can be separated using standard deviations.
  - Walking and cycling can partly be separated using standard deviations.
  - The standard deviation, as well as the total mean, total rms and PP values can be used to separate the different walking speeds.
  - Lying can be recognised using the up-down orientation.
  - Sitting can be separated from the difference between the X and Y value, which is dependent on the angle of the device.
  - No feature for the separation of the different cycling speeds has yet been found.
- For LOO the difference between the participants is too large to draw conclusions, but the same features seem to be applicable for LOO as were for WS for the same classifications.

Not all features have been tested separate, but one of the steps in the future is to check the stability of the data. The following quote from Khusainov, R. states the importance of checking the noise of the data received: “Standard Deviation represents the probability distribution of data and is less useful for noisy data, because spurious values can distort results”. In Matlab no filtering is applied at all with the current script. For the mean, rms, etc. some noise will not give a problem, but for the peak-peak values and the standard deviation (as mentioned in the quote), the stability and low noise of the signal is important. In the future it can be analysed whether a filter improves the results.

All the features have been checked manually. In the future it will be wise to use methods to automatize this. There are enough feature selection tools available, but the “Sequential Forward Floating Search (SFFS) method” is said to perform best. [11]

### Statistical analysis

Different statistical analysis methods have been analysed in order to find the best features for the separation of the classes. A design choice must be made before clear statements can be made. From the ANOVA clear graphs state the differences between the different classes for the features, they are however not scientific supported using a certain value to compare with. The post-hoc analysis gives clear figures, showing the separating power of the different features, but the confidence interval plotted often suggests that classes can be separated with that feature, while this is not the case when looking at the ANOVA plot. An analysis script has been written to find all the  $\alpha$  values of the post-hoc analysis and use these to find the features that satisfy the originally set value of  $\alpha$  and as well finds the amount of features satisfying this  $\alpha$  per class division. The results of this script were stated in Table 18. From this data, an analysis can be made which classifications need more features. These results correspond quite well to the confusion matrices from for instance Table 11. There are therefore several options to analyse the features wanted, they are once more summarised below. From this list a choice should be made in another research and worked out in detail for detailed feature analysis in the future.

- ANOVA: clear graph, but limited scientific foundation
- Post-hoc analysis: confusing graph, but clear scientific foundation
  - Script processing the scientific foundation, creating tables for checking overall performance
- Confusion matrices: clear values, shows classifications that need to be improved, but no scientific foundation, such as an ANOVA or post-hoc analysis.

Below an example of the script, processing the post-hoc analysis is shown with again the reasoning why and conclusions that could be drawn from it.

### Post-hoc graph substituted with other script

When looking at the right side of Figure 22, the confidence interval of the plotting of the post-hoc suggests that classes are separated clearly, but in practice this is not the case, when looking at the left side of the figure. A script has been written to use the data from the analysis to create other comparisons, without using the plot. The detailed results are visible in Appendix D: Detailed results. It is visible that using the numbers of the post hoc analysis, four classes cannot be separated at all and 33 class pairs have at least one

person that is not able to classify that activity pair within that person. The table also shows the class pairs that still need improvement (marked yellow), assuming an average of at least 3.0 features per one-to-one classification needed. The marked areas are again the cycling part and the transitions and walking at 2.4 km/h, which were also found in the confusion matrix. Assuming this 3.0 features per classification (with  $\alpha < 0.001$ ), the walking speeds from 4.0 km/h and up are classified correctly. The numbers from this script are more accurate than the post-hoc plots and are therefore preferred, but as mentioned before, further analysis is needed to make an educated choice whether the easier ANOVA plot might already be enough, or other less intuitive methods must be used.

## SVM

The SVM method has shown that depending on the settings, interesting accuracies can be reached, at least for WS. When looking at the confusion matrix of SVM as in Table 11, different areas could be recognised.

- Around the diagonal (1) the values being close to it represent the fact that the classification is already good enough that only similar activities are misclassified (and some other areas). Also the walking section is seen to follow the diagonal and not creating a square area, therefore also stressing the separating ability within this section: low speeds and high speeds are not mixed up. In order to improve the classification further, features must be found stressing the small differences between the similar activities, for instance the frequency analysis. It can be seen that the cycling speeds are most off and need most correction.
- Since a combination of walking and cycling is found on both sides of the diagonal (2), there are apparently not enough features to separate these activities correctly. New features should be found to stress this difference between walking and cycling.
- The fact that for almost all activities standing is predicted (3), might be explained by the fact that standing is a basic posture, such that it is recognised in all activities and therefore assigned to it. The other activities on the other hand are not recognised as standing and assigned to standing, stressing the fact that standing is a basis for the other activities. If later instead of SVM for instance decision tree is used, this analysis of standing could be one of the features to be looked at first: is the participant standing; then the activities (3) are possible.
- Cycling at 19.0 km/h is often predicted for all activities, in this case however all activities are seen as cycling, instead of cycling being seen as all other activities. It is interesting to see that the most active form of cycling is getting the most static activities predicted. A possible reasoning for this misclassification thought of is the inclusion of some curves in the measurements. If curves in the cycling part are included the speed is reduced and the participant stops peddling. If this is the case the data consisted of cycling data and this transition behaviour. It is possible that this transitional behaviour is recognised in a few of the activities and assigned to cycling. The other way around would then not be possible since this data would also contain cycling data and therefore not be fitting to the other activities. This last part is all a hypothesis. In order to check this, the first eight participants must be left out for a new analysis, since it is not annotated correctly for these participants whether curves were included in the data or not.

For an idea of the accuracy: when looking at an already smartphone implemented activism classification, ambulation accuracies of 87.2% and posture accuracies of average 97.6% are reached (when taking the accuracy of one column).[22] Ambulation is defined as walking on a flat surface and on stairs, postures are also defined as standing, sitting and lying. This research achieved an ambulation accuracy of 92.2% which is comparable and a posture accuracy of 90.4%, which is slightly lower. A side note must be made that they seem to have used LOO, but had a more standardised position of the accelerometer and the numbers from this study are WS, therefore leaving the different positions of the accelerometer out.

The difference between Within Subject and Leave One Out analysis can be seen from the higher amount of wrong predictions for the Leave One Out analysis, but these seem to be divided in the same manner; the errors occur on the same places, but the amount of errors is higher. For LOO all the different characteristics of the different participants are also included as extra variables, such as the position of the accelerometer,

the step size, the tread of walking: putting the foot down flat or rolling it off, etc.. The addition of these extra variables might cause a shift in the features between different participants, therefore misclassifying the data.

In order to improve the Leave One Out analysis, a participant identifying feature could be used. When looking at the different participants, corresponding behaviour can be found between certain participants, as visible in Figure 14. There are indeed differences due to the choice of position of the accelerometer (left or right hip, slightly more forward or back and on the belt or on the elastic band), probably the length of the legs, the type of shoes, etc. It is however expected that there are a limited amount of these variables and a limited amount of variations the accelerometer data can assume. If a feature is made of these settings, this enables the classification system to, even though it is LOO, classify the behaviour using a type of WS analysis. This feature however must first be determined. There are various methods thought of to achieve this. First analysis is needed to check everything, but a few methods thought of are:

- Calculate the participant determining feature(s) from manually inserted settings, such as body length and position of the accelerometer (left or right).
- Calibrate the data at the start of the usage of the accelerometer: perform a certain routine and let the algorithm determine the type of accelerometer data found and generate features corresponding.
- Real time determination of the type of accelerometer data found: analyse the data over a longer time and find the participant determining features.

This method enables the possibility to separate the data further, allowing a more accurate classification of the data. One of the first steps towards this conversion would be to analyse whether corresponding datasets (with the same relation between the x, y and z- axis for different activities) can be found between different participants or measurement sessions. The measurements could for instance be done multiple times with the accelerometer each time on a different position to find the relation between the position and the type of data received. With this knowledge a protocol can be designed to check different participants of different length, different types of shoes, etc. This new dataset could be used to assign the new features, determining the type of accelerometer data received.

## **IMA**

The IMA has been implemented differently than previous expected, since the values of the thresholding are determined using SVM instead of ROC curves. It is however expected that the other method would not have given any different results. As visible in Figure 23, the postures cannot be distinguished, which is logical, since the IMA only looks at the intensity and independent of the posture, the intensity is low, therefore putting the postures on one group. The IMA can be used to get a general idea of the intensity of the different activities, but as expected not in detail and cannot give information on the sedentary behaviour, since the posture cannot be determined, this last part is also confirmed by the LOO analysis of the IMA in Table 14, since all the data of the postures are combined in sitting (except one datapoint).

## **Summary of accuracies**

Analysis of different settings, such as IMA vs SVM, six seconds vs ten seconds and WS vs LOO has been worked out in Table 15 and Table 16.

- From the summary of the accuracies it can clearly be seen that the Leave One Out analysis is not yet very effective: on all different aspects it performs significantly worse compared to the within subject analysis. The before mentioned option of adding a feature that can recognise the type of data received could, with a certain efficiency, convert the LOO to WS, therefore increasing the accuracy.
- The six second windows have performed better than the ten second windows and are therefore suggested to be used.
- The IMA values are easy to calculate and currently consisting datasets, only containing the IMA values can be used to separate the postures from different walking speeds, but further classification is not possible, especially not for classifying sedentary behaviour.

A side note must be placed to the total accuracies: all measurements had the same weight; WS or LOO was calculated and that value was used in the calculation of the total accuracy. There were datasets that had been split in two separate datasets. These had therefore only half of the data, but were weighted evenly, therefore had in total more influence on the accuracy. A weight compensation function should be built in to compensate for potential errors caused by this method.

The standard deviations of the accuracies show that there is still quite some difference in quality between different calculated support vectors. On average the results are quite good, there are even systems performing with a 100%, but it is dependent on the activity how stable it is in practice. The most interesting settings are the only merge cycle and the sitting accuracy with respectively standard deviations of 5.3 and 19.6. The 5.3 seems fine for the current system with the simple features used, but the 19.6 shows that the system is still quite unstable in its performance. In later research, attention must be paid to the calculation of features and the corresponding SD, this influences the SD of the accuracies.

From the confusion matrices it was concluded that an overall increase in accuracy could be achieved if more and more detailed features were added.

The found results are not yet as high as the current literature accuracies in practice, this mostly because the features have not been completely implemented. In literature ambulation is often 93% accurate and posture 99.63% accurate[11], it is sometimes unclear whether this is WS or LOO analysis. In this research, the ambulation was 92.2% accurate and the posture 94.0% accurate for WS. In case of WS the results are already quite good, but in case of LOO, the system is clearly underperforming with an accuracy of 40% less.

### 5.3. General

When the plot constructed from the activity compendium (Figure 4) is reviewed again, it is visible that walking has speeds at light, moderate and vigorous PA intensity, while cycling only has speeds at moderate and vigorous PA intensity. The recommendations state that a minimum of moderate PA intensity is needed for the PA, further differentiation of the speeds is interesting, but not needed. If cycling is merged to one class, it can still be used to classify the different PA intensities, but the total accuracy is increased.

The Within Subject results can be used for the feedback system if the cycling is merged, with an accuracy of  $94.1 \pm 5.3\%$ .

The results are however limited to Within Subject analysis for accurate results. This also means that people have to calibrate the system with all different activities first before it can be used in the current state with this accuracy.

Furthermore the reason for the difference between the Leave One Out and Within Subject analysis has not been researched, so it is possible that each time a participant starts the system, the results will be different. Before, a few methods have been mentioned to improve the results of the LOO more towards the WS. Looking at the literature and seeing the high accuracies reached there, it is estimated that if the position of the accelerometer is standardised or a dataset of all possible orientations is made, the LOO might be increased to an accuracy corresponding to the WS.

#### Implementation in feedback system

The results of this study show that it is possible to implement the PA activity monitoring system with an accuracy of 94% and the sedentary behaviour with 90.7% (with a SD of 19.6). Further research however is first preferred. As mentioned before, the following options should be researched first before actual implementation:

- An analysis of the stability of the dataset
  - The influence of the orientation of the accelerometer on the dataset
  - The influence of the body characteristics, such as length and shoes on the dataset

- The ability to overcome the stability problems, for instance:
  - Standardisation of the position of the accelerometer
  - Implementation of these datasets as features to add the type of dataset to the classification
- The features
  - New features should be calculated and tried.
    - Frequency behaviour
    - Energy/entropy
    - Averaged mean of cross covariance between each two axes
  - An automatic feature selection should be used to select better features and removing accuracy lowering features.

When looking at literature, the expectation rises, that with the previously mentioned research, the LOO can be improved to reach an overall accuracy of 90% for activity classification after merging the cycle speeds to one class. All these percentages are classification per 6 seconds. In the feedback system this data must be processed, for instance using an averaging function, to be able to connect the recommendations to timing, for instance the recommendation of the bouts of 10 minutes for the PAs.

When looking at the feedback for the PA intensity: since the postures can be separated quite efficiently from the dynamic activities, the feedback is estimated to achieve an accuracy of 85%, at the borders of the different PA intensity levels and increasing to 100% for more intensive activities.

For the feedback of the sedentary behaviour: the variation in quality of classification ( $SD=19.6$ ), suggests that the system must be improved quite a bit, but since literature shows that often an accuracy of 95% till 100% has been achieved, this should be fine after research to the above suggestions. Also in this case an accuracy of 90% is estimated to classify the sedentary behaviour.

At this moment the system is not accurate enough to implement it and expect large differences compared to the old system. With the suggested changes a system could be developed that would work for the calibrated activities, but also for other dynamic activities, since for instance the SD and RMS are taken into account, classifying activities, even if not directly very intensive. The accuracy for the system is expected to behave the same including new activities: 85% around the borders of the PA intensity and increasing for more intensive activities. Activities such as lifting heavy objects and vertical arm work will still be underestimated, but they will be registered as more active compared with the previous system, since posture changes (when bending to pick up, etc.) can be used in the PA intensity classification.



## 6. Conclusion

This study showed that it is possible to operationalise the “Global recommendations on physical activity for health” using an tri-axial accelerometer on the waist and applying support vector machine (SVM) as a data analysis method. For an accurate implementation, the system must be able to distinguish different physical activity intensities, which can be determined from the activities cycling (as one class for all different speeds from 10.1 km/h and above) and walking (with a specification of the speeds in at least 4.0 and 5.3 km/h). The accuracy of SVM analysis, analysing within subject is  $94\pm5\%$  for activity classification and therefore physical activity (PA) intensity classification and  $90\pm20\%$  accuracy of sedentary behaviour. The analysis is also possible using the integral of the modulus of the accelerometer (IMA) method, but then only an accuracy of  $66\pm15\%$  is achieved for the activity classification in general and sedentary behaviour cannot be classified. For a universal feedback system a between subject system is needed, making it universal applicable: IMA is then no longer possible with an accuracy of  $45\pm13\%$  for the PA intensity and SVM is less preferable with an accuracy of  $71\pm13\%$ .

The basic setup has been made and offers a structure for future analysis of new settings. In literature, accuracies up to 90% can be achieved, even for between subject analysis. It is expected that these higher results can also be achieved for the between subject analysis, if the algorithms are improved as discussed. This improved monitoring system should be extensive enough that it can be used for a feedback system with the PA intensity and sedentary behaviour as new features, besides the old PA classification system that worked per day, based on IMA values.

Sitting and the intensity of walking and cycling can be classified with an accuracy of 90% for within subject analysis and is expected to be able to perform at 85% for between subject analysis when the algorithms are improved as discussed. This would result in a monitoring system that corresponds to the accuracy found in literature (when looking at the classification method), which is a prerequisite for a good feedback system, which might then be designed with accurate predictions.

This study has started the analysis of the possibilities to implement a classification system to improve the Roessingh Research and Development (RRD) feedback system and has created a basis for this. This basis could already be implemented for within subject analysis, but if it is extended using the information found, it could provide the basic monitoring needed for an universal feedback system.

## 7. References

1. World Health Organization.
2. Sedentary Behaviour Research Network. [cited 2015 01 July]; Available from: <http://www.sedentarybehaviour.org/what-is-sedentary-behaviour/>.
3. World Health Organization. *WHO Physical Inactivity: A Global Public Health Problem*. 2015 [cited 2015 20th May]; Available from: [http://www.who.int/dietphysicalactivity/factsheet\\_inactivity/en/](http://www.who.int/dietphysicalactivity/factsheet_inactivity/en/).
4. Nederlands Instituut voor Sport & Bewegen. *Beweegnormen/NISB*. Available from: <http://www.nisb.nl/weten/normen/beweegnormen.html>.
5. World Health Organization, *Global recommendations on physical activity for health*. 2010: p. 58.
6. Esliger, D.W. and M.S. Tremblay, *Physical activity and inactivity profiling: the next generation* This article is part of a supplement entitled *Advancing physical activity measurement and guidelines in Canada: a scientific review and evidence-based foundation for the future of Canadian physical activity guidelines* co-published by *Applied Physiology, Nutrition, and Metabolism* and the *Canadian Journal of Public Health*. It may be cited as *Appl. Physiol. Nutr. Metab.* 32(Suppl. 2E) or as *Can. J. Public Health* 98(Suppl. 2). *Applied Physiology, Nutrition, and Metabolism*, 2007. **32**(S2E): p. S195-S207.
7. Lowe, S.A. and G. ÓLaighin, *Monitoring human health behaviour in one's living environment: A technological review*. *Medical Engineering and Physics*, 2014. **36**(2): p. 147-168.
8. Ellis, K., et al., *A random forest classifier for the prediction of energy expenditure and type of physical activity from wrist and hip accelerometers*. *Physiological Measurement*, 2014. **35**(11): p. 2191-2203.
9. Rosenberger, M.E., et al., *Estimating activity and sedentary behavior from an accelerometer on the hip or wrist*. *Medicine and Science in Sports and Exercise*, 2013. **45**(5): p. 964-975.
10. Kim, D.Y., et al., *Different location of triaxial accelerometer and different energy expenditures*. *Yonsei Medical Journal*, 2014. **55**(4): p. 1145-1151.
11. Khusainov, R., et al., *Real-time human ambulation, activity, and physiological monitoring: taxonomy of issues, techniques, applications, challenges and limitations*. *Sensors (Basel, Switzerland)*, 2013. **13**(10): p. 12852-12902.
12. Selderhuis, A., *Het activiteitenpatroon van kantoorwerkers*, 2015, Universiteit Twente: Enschede. p. 43.
13. Chen, K.Y. and D.R. Bassett, *The Technology of Accelerometry-Based Activity Monitors: Current and Future*. *Medicine and Science in Sports and Exercise*, 2005. **37**(Supplement).
14. AINSWORTH, B.E., et al., *Compendium of Physical Activities: an update of activity codes and MET intensities*. *Medicine & Science in Sports & Exercise*, 2000. **32**(9): p. S498-S516.
15. Arizona State University. *Compendium of Physical Activities*. 2011 [cited 2015 05 May]; Available from: <https://sites.google.com/site/compendiumofphysicalactivities/Activity-Categories>.
16. Preece, S.J., et al., *Activity identification using body-mounted sensors - A review of classification techniques*. *Physiological Measurement*, 2009. **30**(4): p. R1-R33.
17. Matlab, *support vector machine example Figure*, in *Matlab*.
18. Stephen, J.P., et al., *Activity identification using body-mounted sensors—a review of classification techniques*. *Physiological Measurement*, 2009. **30**(4): p. R1.
19. Pliagkas, C., *Activity classification using a hip worn inertial sensor*. 2011.
20. Bastian, T., et al., *Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: Laboratory-based calibrations are not enough*. *Journal of Applied Physiology*, 2015. **118**(6): p. 716-722.
21. Khan, A.M., et al., *Activity Recognition on Smartphones via Sensor-Fusion and KDA-Based SVMs*. *International Journal of Distributed Sensor Networks*, 2014. **2014**: p. 14.
22. Anguita, D., et al., *Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine*, in *Ambient Assisted Living and Home Care*, J. Bravo, R. Hervás, and M. Rodríguez, Editors. 2012, Springer Berlin Heidelberg. p. 216-223.

## A: Informed consent

## B: Analogue annotation tool

Postures				Opmerkingen
			4x	
Stand				
Sit				
Lie				
Stand				
Lie				
Sit				
Walk tread		3x	Opmerkingen	
Step on				
2.4				
4				
6.9				
7.5				
6.9				
4				
2.4				
stand				
Walk outside		3x	Min km/h	Max
Stand				
4				
5.3				
6.9				
5.3				
4				
Cycle outside		3x	Min km/h	max km/h
Sit on				
10.1				
16				
19				
16				
10.1				

## C: Matlab code

Can be requested at RRD.

## D: Detailed results

### Analysis feature script

Below a list is visible, generated using a script, processing the outcomes of the post-hoc analysis: All class pairs are compared (stand with sit, stand with lie, etc.). The post-hoc analysis gives a measure of the separating power per class pair, this is  $\alpha$  or the p-coefficient. The script checks whether the  $\alpha$  is higher or lower than 0.001, if it is lower the separation is chosen to be significant and that corresponding feature is marked. After all features have been calculated for all pairs, the amount of marked features per class pair is summed. The average is taken of these sums over the amount of participants. This average and the minimum amount of features of all participants per class pair are noted. If the average nr of features is high, there are enough features that can be used to separate those pairs. A preferred average number is 3 (random chosen after looking at the numbers). If the minimum number of features is 0, there are people who cannot separate those class pairs and more or more efficient features are needed. Yellow marked are the class pairs that have less than 3.0 features on average and a minimum number of features of 0. These marked pairs need better feature analysis.

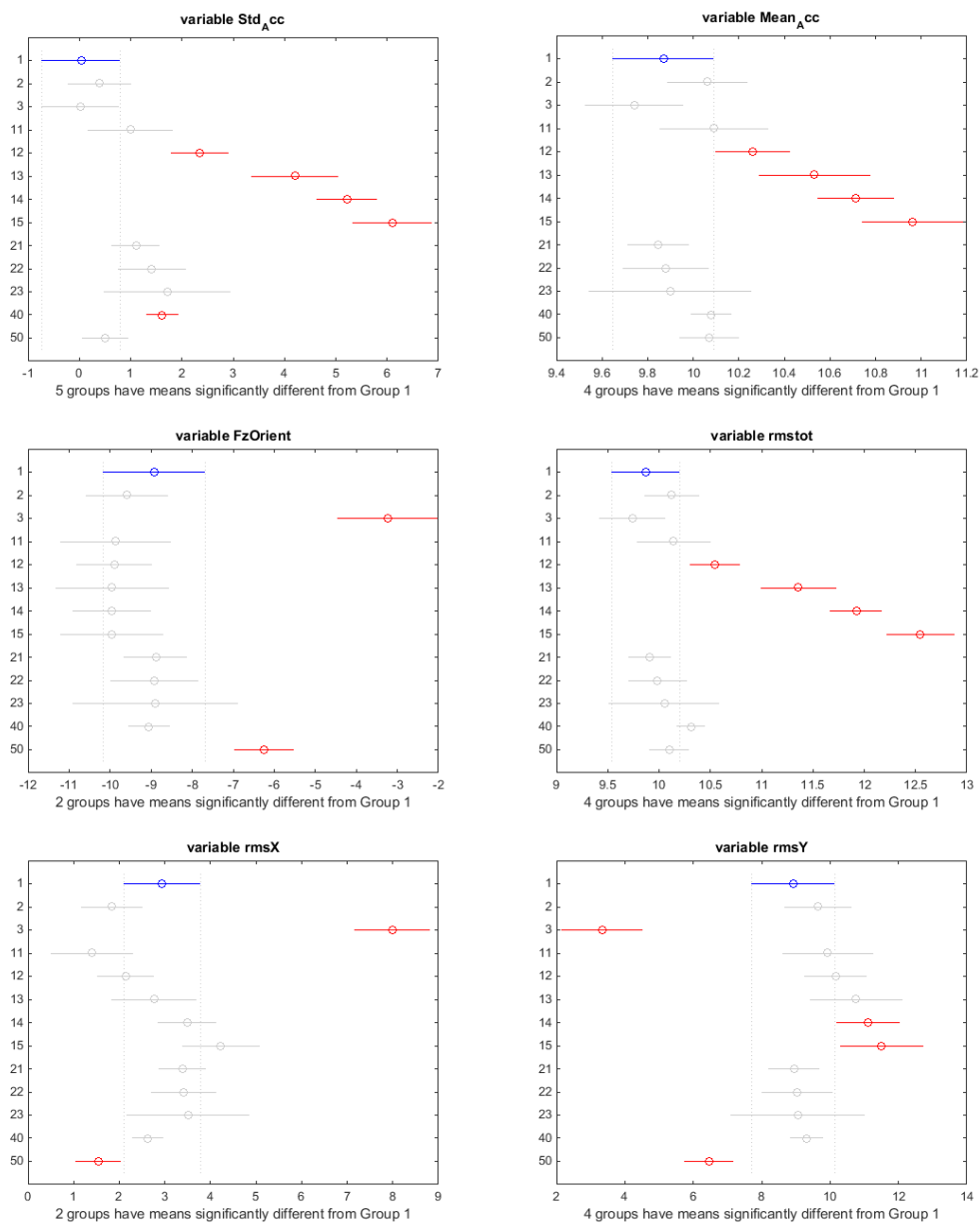
clas s 1	clas s 2	averag e nr of featur es	min nr of featur es		clas s 1	clas s 2	averag e nr of featur es	min nr of featur es		clas s 1	clas s 2	averag e nr of featur es	min nr of featur es
1	2	1,08	0		3	14	8,67	8		13	21	5,33	3
1	3	4,00	3		3	15	8,50	8		13	22	4,67	3
1	11	0,50	0		3	21	3,92	3		13	23	2,42	0
1	12	3,50	3		3	22	4,75	3		13	40	4,00	1
1	13	5,33	4		3	23	5,08	3		13	50	7,25	4
1	14	6,58	6		3	40	8,17	7		14	15	6,67	0
1	15	6,58	6		3	50	4,08	2		14	21	6,83	6
1	21	1,25	0		11	12	0,00	0		14	22	6,67	5
1	22	1,83	0		11	13	3,67	1		14	23	5,25	3
1	23	1,75	0		11	14	5,83	4		14	40	6,42	6
1	40	4,08	3		11	15	6,08	6		14	50	8,50	8
1	50	3,25	1		11	21	5,83	0		15	21	6,83	6
2	3	3,75	3		11	22	5,83	0		15	22	6,58	6
2	11	1,00	0		11	23	2,50	0		15	23	5,83	4
2	12	4,17	3		11	40	2,50	0		15	40	6,67	6
2	13	5,58	4		11	50	2,25	0		15	50	8,42	7
2	14	6,67	6		12	13	1,42	0		21	22	0,00	0
2	15	7,00	6		12	14	5,83	3		21	23	0,00	0
2	21	7,50	0		12	15	6,00	5		21	40	2,58	0
2	22	1,50	0		12	21	3,42	0		21	50	2,50	0

2	23	1,50	0		12	22	1,42	0		22	23	0,00	0
2	40	4,17	3		12	23	2,50	0		22	40	3,33	0
2	50	2,50	0		12	40	8,33	0		22	50	2,25	0
3	11	4,17	3		12	50	5,67	3		23	40	8,33	0
3	12	8,00	7		13	14	2,58	0					
3	13	8,42	7		13	15	4,08	2					

Table 18 Feature compare. Per class pair the number of significant separating features is given. Determined using all participants, Within Subject analysis and a p-coefficient of 0.001. Max features is 7, calculated from a total of 12 participants. In yellow marked those with an average lower than 3 numbers of features

## Features

Below 8 features are given, showing the resulting separating abilities.



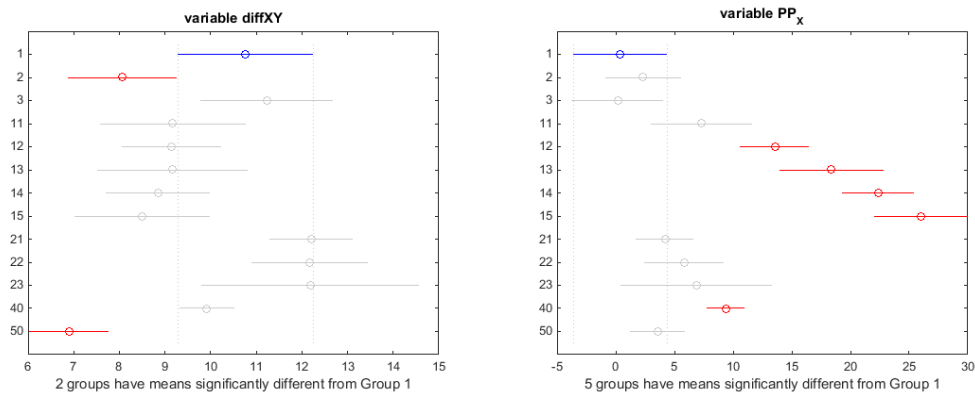


Figure 26 Overview of features with post hoc analysis. On the x-axis the value of the features are plotted in  $m/s^2$  and on the y-axis the different classes with their own numbering.

### Total accuracy WS vs LOO & kernel: RBF vs linear

Below the comparison can be seen between the WS and LOO analysis and the difference between a RBF and a linear kernel. The WS performs better than the LOO, since it has a higher accuracy, except for sitting. The RBF performs better than the linear kernel, but the difference is so small that it might not be significant.

SETTINGS \ METHOD	IMA WITHIN SUBJECT ( % $\pm$ SD ) RBF KERNEL	IMA LEAVE ONE OUT ( % $\pm$ SD ) RBF KERNEL	IMA WITHIN SUBJECT ( % $\pm$ SD ) LINEAR KERNEL	IMA LEAVE ONE OUT ( % $\pm$ SD ) LINEAR KERNEL
NORMAL	62,8 $\pm$ 14,2	35,6 $\pm$ 9,8	60,7 $\pm$ 14.1	36,0 $\pm$ 11.2
INCLUDE TRANS	27,2 $\pm$ 7,4	15,2 $\pm$ 3,8	26,3 $\pm$ 7.5	15,4 $\pm$ 4.2
MERGE WALK AND CYCLE	68,9 $\pm$ 12,5	61,4 $\pm$ 11,2	60,8 $\pm$ 12.4	60,6 $\pm$ 10.7
JUST POSTURE	41.0 $\pm$ 14,4	24,6 $\pm$ 11,8	39,5 $\pm$ 13.6	24,6 $\pm$ 11.8
JUST WALK	91.0 $\pm$ 11,4	54,7 $\pm$ 18,6	90,7 $\pm$ 12.2	55,2 $\pm$ 20.2
JUST CYCLE	65,5 $\pm$ 19,4	59,2 $\pm$ 14,8	65,4 $\pm$ 19.6	54,7 $\pm$ 16.9
ONLY MERGE CYCLE	66,1 $\pm$ 14,6	44,6 $\pm$ 13,2	64,8 $\pm$ 15.0	45,5 $\pm$ 13.8
SITTING ACCURACY	14,6 $\pm$ 17,7	23,2 $\pm$ 11,0	11,9 $\pm$ 18.4	23,1 $\pm$ 10.9

Table 19 Comparing overall performance for different settings. Only IMA compared for 6 seconds with as kernel linear and rbf