Personalization of energy expenditure and cardiorespiratory fitness estimation using wearable sensors in supervised and unsupervised free-living conditions

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus, prof.dr.ir. F.P.T. Baaijens, voor een commissie aangewezen door het College voor Promoties in het openbaar te verdedigen op dinsdag 15 december 2015 om 16.00 uur

door

Marco Altini

geboren te Ravenna, Italie

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr.ir. A.C.P.M. Backx
1^e promotor:	prof.dr.sc. O. Amft
copromotor:	prof.dr.ir. J.W.M. Bergmans
leden:	prof.dr. G.-Z. Yang FREng (Imperial College London)
	prof.dr.ir. C. van Hoof (Katholieke Universiteit Leuven)
	dr.ir. P.J.M. Cluitmans
	dr. A. Di Bucchianico

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

This thesis was prepared with the LaTeX documentation system.
Cover design: Mari Tosmin

A catalogue record is available from the Eindhoven University of Technology Library.
ISBN: 978-90-386-3971-0

Personalization of energy expenditure and cardiorespiratory fitness estimation using wearable sensors in supervised and unsupervised free-living conditions, by Marco Altini
Eindhoven: Technische Universiteit Eindhoven, 2015 – Proefschrift

Abbreviations

ACC	Accelerometer
ACT	Active
ADL	Activity of daily living
ANOVA	Analysis of variance
AS	Activity-specific
BMI	Body mass index
BMR	Basal metabolic rate
CO_2	Carbon dioxide
CRF	Cardiorespiratory fitness
DIT	Diet induced thermogenesis
DLW	Doubly labeled water
ECG	Electrocardiogram
EE	Energy expenditure
FFT	Fast Fourier transform
GPS	Global positioning system
GSR	Galvanic skin response
HMM	Hidden Markov model
HR	Heart rate
HRaR	Heart rate above rest
HRV	Heart rate variability
HP	High pass
HWBM	High whole body motion
IRB	Institutional review board
LDA	Latent Dirichlet allocation
LP	Low pass
LWBM	Low whole body motion
MAPE	Mean absolute percentage error
MET	Metabolic equivalent
MI	Motion intensity
MCU	Microcontroller unit

O_2	Oxygen
PA	Physical activity
PAEE	Physical activity energy expenditure
PAL	Physical activity level
REE	Resting energy expenditure
RMR	Resting metabolic rate
RMSE	Root mean square error
SED	Sedentary
SVM	Support vector machines
TEE	Total energy expenditure
TM	Topic models
VO_2max	Maximal oxygen uptake

Summary

Personalization of energy expenditure and cardiorespiratory fitness estimation using wearable sensors in supervised and unsupervised free-living conditions

At present, two thirds of the world population is overweight and fails to achieve the minimum physical activity recommendations, making lack of physical activity one of the major health problems worldwide. Physical activity has been defined as any bodily movement produced by skeletal muscles which results in energy expenditure (EE). Thus EE, together with cardiorespiratory fitness (CRF), i.e. the ability of the circulatory and respiratory systems to supply oxygen during sustained physical activity, are among the most important determinants of health and wellbeing. In the past few years, ubiquitous sensing technologies showed unprecedented insights into the relation between physical activity and health and have been driving behavioral change. Wearable sensors are getting more and more widespread due to improvements in miniaturization, battery capacity and user experience design, reaching ubiquitousness in the quantified-self community and being rapidly adopted by the general population. As a result, a multitude of EE estimation systems were developed in the recent past. However, currently such systems rely on population-based estimation approaches that often do not provide accurate estimates at the individual level.

Physiological data such as heart rate (HR) is key in providing accurate, personalized EE estimates. For example, HR at the individual level is highly correlated with EE due to the strong relation between oxygen consumption, HR and EE. However, the individual-specific relation between HR and EE differs between persons, challenging the generalization of standard population-based approaches for EE estimation. As a result, individual calibration and laboratory tests are needed to normalize HR. The rationale behind the need for normalization is that individuals with similar body size expend a similar amount of energy during a certain activity, however their HR differs depending on other factors, for example, CRF.

When performing the same activity, fitter individuals will have lower HR compared to less fit ones. Thus, EE estimation models relying on HR to predict EE will result in underestimations and overestimations of EE, unless HR is normalized for physical fitness level.

Another limitation of current physical activity monitoring devices and algorithms is the focus on EE only. EE reflects the individuals' behavior and not the individuals' actual health status. On the other hand, CRF can be considered a proxy to cardiovascular and cardiorespiratory health, and therefore a marker of health tightly coupled with physical activity. Current practice for CRF measurement is affected by multiple limitations. Direct measurement of oxygen volume during maximal exercise (i.e.VO_2max) is the gold standard. However, VO_2max tests require medical supervision and can be risky for individuals in non-optimal health conditions. Submaximal methods to estimate VO_2max are limited by the necessity to perform laboratory tests or strict exercise protocols. Thus, novel methods that can provide personalized physical activity monitoring and estimate markers of health such as VO_2max in free-living are needed.

The aim of this thesis is to introduce new methods and models to provide accurate EE estimation at the individual level without requiring individual calibration and to estimate VO_2max in free-living conditions. We rely on wearable technology to acquire combined inertial and physiological data, such as accelerometry and HR. Novel EE estimation techniques are proposed in this thesis to account for variability in physiological data between individuals and determine normalization parameters without the need for laboratory tests or individual calibration. Furthermore, we propose methods to bring normalization techniques to free-living conditions, avoiding laboratory protocols that are often required by current solutions. To this aim, we propose to contextualize HR in free-living conditions using a context-recognition architecture to determine different levels of context, from low level activity primitives (e.g. walking) to high level activity composites (e.g. commuting or working). Thus, by determining HR in specific contexts in free-living conditions, we obtain submaximal HR at predefined intensities while avoiding the need for strict exercise protocols. We use contextualized HR to personalize EE estimation models and to estimate VO_2max.

This thesis includes nine scientific publications addressing four objectives:

1. Selection of methods, sensor number and positioning: to determine which combination of methods, sensors number and positioning is best for EE estimation according to state of the art solutions.

2. Physiological data normalization: to develop methods to normalize physiological data and therefore personalize EE estimation regression models, increasing the estimate accuracy at the individual level, without the need for individual calibration.

3. VO_2max estimation using wearable sensor data: to develop methods and models to estimate VO_2max using wearable sensor data, without the need for laboratory tests or strict exercise protocols.

4. Personalized EE estimation and VO_2max estimation in free-living: to develop methods to contextualize HR in free-living conditions, therefore avoiding the need for strict exercise protocols to be performed under laboratory conditions. Then, to use contextualized HR to both personalize EE estimation regression models and estimate VO_2max in free-living.

By addressing the four objectives above, the research included in this thesis shows that machine learning techniques can be used to normalize and contextualize physiological data in either laboratory or unsupervised free-living conditions. Thus, wearable sensors can be used to contextualize physiological data and provide personalized EE estimation and VO_2max estimation without the need for laboratory equipment or specific protocols. The results included in this thesis advance state of the art in terms of providing EE estimates more accurate at the individual level, as well as moving towards quantification of aspects related to health status, such as CRF, and not only behavior (e.g. EE). Finally, we envision different opportunities for future work relying on the proposed methods. Such future prospects could be new applications guiding behavioral change by closing the loop between objective monitoring of physical activity behavior (e.g. EE) and changes in health status as quantified by CRF estimation. Additionally, some of the proposed methodologies could be applied to other applications, beyond physical activity monitoring. For example, hierarchical modeling for implicit signal normalization could be employed in the context of monitoring more accurately at the individual level psychological stressors.

Contents

Summary **vii**

Contents **xi**

1 Introduction **1**

1.1 Physical activity and health . . . 1

1.1.1 Energy expenditure . . . 2

1.1.2 Cardiorespiratory fitness . . . 4

1.2 Current methods to estimate EE and CRF using wearable sensors . 6

1.2.1 Wearable sensors for EE and CRF estimation . . . 6

1.2.2 Methods for EE estimation . . . 7

1.2.3 Activity-specific estimation methods . . . 8

1.2.4 Methods for personalization of physiological data . . . 10

1.2.5 Methods for CRF estimation . . . 11

1.3 Goals of this thesis . . . 13

1.3.1 Selection of methods, sensor number and positioning . . . 13

1.3.2 Physiological data normalization . . . 13

1.3.3 VO_2max estimation using wearable sensor data . . . 14

1.3.4 Personalized EE estimation and VO_2max estimation in free-living . . . 14

1.4 Thesis outline . . . 16

1.4.1 Selection of methods, sensor number and positioning . . . 16

1.4.2 Physiological data normalization . . . 16

1.4.3 VO_2max estimation using wearable sensor data . . . 16

1.4.4 Personalized EE estimation and VO_2max estimation in free-living . . . 17

1.5 Chapter overview of the thesis . . . 18

Part I: Selection of methods, sensor number and positioning **19**

2 Energy expenditure estimation using wearable sensors: a new methodology for activity-specific models **21**
2.1 Introduction 22
2.2 Related work 23
Epidemiological studies 23
Methods based on machine learning 23
Towards activity-specific EE estimation 24
2.3 A new methodology 25
2.4 Use Case - A necklace monitor 27
Method Implementation 27
2.5 Data collection and analysis 29
Participants 29
Instruments 29
Experimental design 30
Study design choices 30
Pre-processing 30
Feature extraction 31
Feature selection 31
Statistics and performance measures 32
2.6 Results 33
Activity clusters classification 33
Activity clusters models 33
EE estimation performance 33
2.7 Conclusions and future work 35

3 Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning **39**
3.1 Introduction 40
3.2 Related work 41
Counts-based estimation methods 41
Activity-specific estimation methods 41
Comparisons 42
3.3 Analysis approach 42
Statistics and performance measure 45
3.4 Implementation 46
Activity type clusters 46
Features extraction and selection 46
Activity recognition 46
Energy expenditure 47
3.5 Evaluation study 47
Participants 47
Instruments 47
Experimental design 48
3.6 Results 48
Estimation methods 48

Sensors number and positioning . . . 49
3.7 Discussion . . . 51
Estimation methods . . . 51
Sensors number and positioning . . . 52
Limitations . . . 54
3.8 Conclusions . . . 54

Part II: Physiological data normalization **55**

4 Personalizing energy expenditure estimation using a cardiorespiratory fitness predicate **57**
4.1 Introduction . . . 57
4.2 Related work . . . 59
4.2.1 Epidemiological research . . . 59
4.2.2 Activity-specific EE estimation . . . 59
4.2.3 CRF estimation . . . 60
4.3 Fitness and heart responses . . . 61
4.4 Approach . . . 62
4.5 Methods . . . 63
4.5.1 Activity recognition . . . 63
4.5.2 Automatic HR normalization system . . . 64
4.5.3 HR normalized . . . 65
4.5.4 Personalized activity-specific EE estimation . . . 65
4.5.5 Feature extraction . . . 65
4.5.6 Feature selection . . . 67
4.5.7 Statistics and performance measure . . . 67
4.6 Measurement setup and data collection . . . 68
4.6.1 Participants . . . 68
4.6.2 Instruments . . . 68
4.6.3 Experimental design . . . 69
4.6.4 Pre-processing . . . 70
4.7 Results . . . 70
4.7.1 Activity recognition . . . 70
4.7.2 Walking speed estimator . . . 70
4.7.3 Heart rate normalization factor estimator . . . 70
4.7.4 Energy expenditure estimation . . . 72
4.8 Discussion and conclusions . . . 72

5 Automatic heart rate normalization for accurate energy expenditure estimation: an analysis of activities of daily living and heart rate features **75**
5.1 Introduction . . . 76
5.1.1 Scientific background . . . 76
5.1.2 Rationale for the study . . . 77
5.1.3 Objectives of the study . . . 77
5.2 Methods . . . 77

5.2.1 Participants . . . 77
5.2.2 Study design . . . 78
5.2.3 Outcome measures . . . 79
5.2.4 Methods for data acquisition and measurement . . . 79
5.2.5 Methods for data analysis . . . 79
5.3 Results . . . 81
5.3.1 Automatic HR Normalization Factor Estimation from ADLs 81
5.3.2 EE Estimation . . . 82
5.4 Discussion . . . 83
5.4.1 Answers to study questions . . . 83
5.4.2 Strength and weaknesses of the study . . . 84
5.4.3 Results in relation to other studies . . . 84
5.5 Conclusions . . . 85

6 Personalizing energy expenditure estimation using physiological signals normalization during activities of daily living 87
6.1 Introduction . . . 87
6.2 Related work . . . 88
6.2.1 EE estimation in epidemiological research . . . 88
6.2.2 Machine learning methods for EE estimation . . . 88
6.2.3 Normalization of physiological signals . . . 89
6.3 Relation between EE, accelerometer and physiological data . . . 90
6.4 Methodology overview . . . 90
6.5 Measurement setup and data collection . . . 92
6.5.1 Participants . . . 92
6.5.2 Instruments . . . 92
6.5.3 Experimental design . . . 94
6.5.4 Statistics and performance measures . . . 94
6.6 Methods implementation . . . 95
6.6.1 Pre-processing . . . 95
6.6.2 Activity recognition . . . 96
6.6.3 Automatic physiological signals normalization using ADLs 96
6.6.4 Personalized activity-specific EE estimation . . . 97
6.7 Results . . . 98
6.7.1 Automatic physiological signals normalization using ADLs 98
6.7.2 Personalized activity-specific EE estimation . . . 98
6.8 Discussion . . . 99
6.8.1 Automatic physiological signals normalization using ADLs 100
6.8.2 Personalized activity-specific EE estimation . . . 102
6.8.3 Conclusion and further work . . . 102

Part III: VO_2max estimation using wearable sensor data 105

7 Personalized cardiorespiratory fitness and energy expenditure estimation using hierarchical Bayesian models 107
7.1 Introduction 108
7.2 Related work 109
7.2.1 Maximal oxygen uptake 109
7.2.2 Submaximal CRF estimation 109
7.2.3 CRF estimation in free living 110
7.2.4 EE estimation 110
7.3 Methods 112
7.4 Evaluation study 116
7.4.1 Participants and data acquisition 116
7.4.2 Experiment protocol 116
7.4.3 Statistics and performance measures 117
7.5 Implementation 117
7.5.1 Pre-processing 117
7.5.2 Features extraction and selection 118
7.5.3 Activity recognition 118
7.5.4 Hierarchical Bayesian regression models 118
7.6 Results 119
7.7 Discussion 122

Part IV: Personalized EE estimation and VO_2max estimation in free-living 127

8 Personalization of energy expenditure estimation in free living using topic models 129
8.1 Introduction 130
8.2 Related work 131
8.2.1 Personalized EE estimation 131
8.2.2 Context recognition 132
8.3 Methods 133
8.3.1 HR normalization parameter estimation 134
8.3.2 Relevant activities composites 136
8.4 Evaluation study 137
8.4.1 Participants and data acquisition 137
8.4.2 Experimental design and validation procedure 139
8.5 Implementation 140
8.5.1 Features extraction and selection 140
8.5.2 Activity primitives 140
8.5.3 Walking speed 141
8.5.4 Stay regions 142
8.5.5 Relevant activity composites discovery 142
8.5.6 HR normalization parameter estimation 142
8.5.7 Personalized EE estimation 143
8.6 Results 144
8.6.1 Activity primitives and speed 144

8.6.2 HR normalization parameter 144
8.6.3 EE estimation 145
8.7 Discussion 145

9 Cardiorespiratory fitness estimation in free living using wearable sensors 151
9.1 Introduction 151
9.2 Related work 153
9.2.1 Maximal and Submaximal Tests 153
9.2.2 CRF estimation in free living 154
9.2.3 Context recognition 155
9.3 Approach 155
9.3.1 CRF estimation 156
9.3.2 Context recognition 156
9.3.3 Relevant activity composites 158
9.4 Evaluation study 159
9.4.1 Participants and data acquisition 159
9.4.2 Experimental design and validation procedure 161
9.5 Implementation 163
9.5.1 Context recognition 163
9.5.2 CRF estimation 164
9.6 Results 165
9.6.1 Activity primitives and walking speed 165
9.6.2 Relevant activity composites 166
9.6.3 CRF estimation 166
9.7 Discussion 167
9.8 Conclusion 170

10 Cardiorespiratory fitness estimation using wearable sensors data: analysis of context-specific submaximal heart rates 173
10.1 Introduction 174
10.2 Methods 175
10.2.1 Participants 175
10.2.2 ECG and accelerometer device 175
10.2.3 Study design 176
10.2.4 Data processing 177
10.3 Results 181
10.3.1 Descriptive statistics 181
10.3.2 CRF estimation from context-specific submaximal HR during simulated activities of daily living 181
10.3.3 Context recognition; activity type and walking speed 181
10.3.4 Relation between context-specific submaximal HR during activities of daily living simulated in laboratory settings and in free-living 182

10.3.5 CRF estimation from context-specific submaximal HR in free-living . . . 183
10.3.6 Cross-validation of VO_2max estimates . . . 185
10.4 Discussion . . . 186
10.4.1 Context-specific HR during activities of daily living simulated in laboratory settings . . . 189
10.4.2 Context recognition in free-living . . . 190
10.4.3 Context-specific HR in free-living . . . 191
10.4.4 Fat free mass . . . 191
10.4.5 Cross-validation of VO_2max estimates . . . 192
10.4.6 Comparison with prior work . . . 192
10.4.7 Limitations and future work . . . 194
10.4.8 Conclusions . . . 194

11 Conclusions and future directions **195**
11.1 Selection of methods, sensor number and positioning . . . 196
11.2 Physiological data normalization . . . 197
11.3 VO_2max estimation using wearable sensor data . . . 199
11.4 Personalized EE estimation and VO_2max estimation in free-living . 200
11.5 Limitations . . . 202
11.6 Future directions . . . 203

Acknowledgments **205**

Resume **207**

Publications by the Author **209**
Journal Papers . . . 209
Conference Papers . . . 210
Patents . . . 212

1 Introduction

1.1 Physical activity and health

Lack of physical activity is one of the major health problems worldwide. At present, two thirds of the world population is overweight and obesity affects a third of the population in the United States [33, 79, 92]. Other diseases, such as type II diabetes [77, 70, 71] and cardiovascular disease [24, 57] are rapidly becoming widespread epidemics as well [133]. Physical activity has been defined as any bodily movement produced by skeletal muscles which results in energy expenditure (EE) beyond resting energy expenditure (REE) [42]. The role of physical activity is therefore key in maintaining a healthy lifestyle for several reasons. First, for the role of EE in regulating energy balance and body weight in the context of obesity prevention and management. Secondly, in a broader sense in the context of many other diseases resulting by lack of physical activity, such as cardiovascular disease or type II diabetes.

Historically, the first epidemiological studies on physical activity date back to the 50's. The seminal work of Morris et al. [90] showed lower mortality rates in bus conductors and postmen compared to bus drivers and telephone operators, initiating physical activity and health research. Since then, a multitude of studies showed that physical activity improves quality of life (e.g. lowering stress) and functional capacity (e.g. fitness) in both healthy and non-healthy individuals [28, 46]. Positive outcome of exercise was shown in children, adults as well as elderly, even regardless of total amount of physical activity or cardiorespiratory fitness (CRF) [86, 64]. However, while the health benefits associated with the effect of a physically active lifestyle are today recognized worldwide in our society [68, 97], more than 60% of the world population currently fails to achieve the minimum physical activity recommendations of 30 minutes of moderate-intensity physical activity daily [65].

The current physical activity level of the western world is considerably different from our Paleolithic ancestors, who spent their day hunting and gathering food [49]. Our ancestors used to expend much more energy compared to the energy we expend nowadays, and showed levels of CRF about 50% greater than ours [50]. Initially the agricultural revolution, subsequently the industrial revolution, and finally the shift towards computer-based work of the past two decades, completely disrupted human lifestyles. However, our genome has not changed much over the past thousand years and more [49]. As a result, the world population is now facing a multitude of epidemics due to lack of or reduced physical activity.

1.1.1 Energy expenditure

Being able to quantify physical activity is important for epidemiological research so that relations between physical activity, health status, environmental factors, and so on, can be determined. Similarly, quantification of physical activity can be key in deploying just in time interventions and promote behavioral changes by providing individuals with an objective assessment of their physical activity behavior. The definition of physical activity, i.e. any bodily movement produced by skeletal muscles which results in EE beyond rest energy expenditure (REE), introduces the concepts of EE and REE as key parameters to identify in order to quantify physical activity. Total energy expenditure (TEE or simply EE) is mainly a sum of internal heat produced and external work [58]. The internal heat produced is, in turn, mainly a sum of basal metabolic rate (BMR), simply put the EE an individual consumes while at rest, to sustain basic body functions such as breathing and the diet induced thermogenesis (DIT), which is allocated for processes such as digestion. External work may be estimated by measuring physical activity energy expenditure (PAEE). Therefore EE is typically defined as composed of three factors, BMR (or REE), DIT, about 10% of TEE, and finally PAEE, the EE due to physical exercise (see Fig. 1.1). PAEE is the most variable component of EE in humans, and the most relevant to quantify, since it is the component we can act upon with interventions targeting increased EE. EE is therefore the most commonly used single metric to quantify physical activity. Nevertheless, the effects of physical activity go beyond EE. Type, duration, frequency and intensity of physical activity have all impact on health outcome [82].

EE measurement gold standards

EE can be measured in three different ways; direct calorimetry, indirect calorimetry and doubly labeled water (DLW). Direct calorimetry is the gold standard, it measures the actual heat lost by the body during activity or rest. Until now, the only way to measure body heat has been to place an individual in a special, sealed room, i.e. the room direct calorimeter. Room calorimeters have been used since the early 1800's to measure metabolism but are the least practical way of measuring EE, since expensive and cumbersome equipment is required [74]. Alternatively, indirect calorimeters have been developed. Indirect calorimetry analyzes inhale and exhale gases concentration to measure EE [85]. Measuring the con-

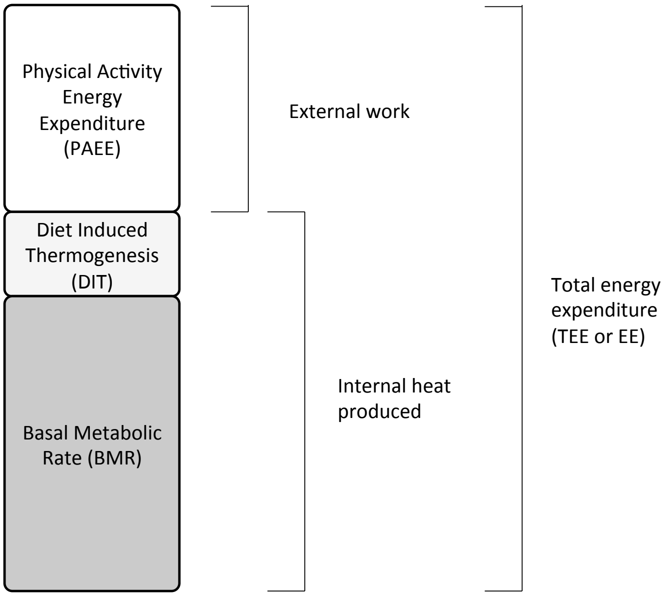

Figure 1.1: Breakdown of EE components as a sum of internal heat produced, comprising BMR and DIT, and external work, i.e. PAEE.

sumption of oxygen (O_2) rate and the production of carbon dioxide (CO_2) rate can be converted to EE following Weir's equation [128]. This method is a bit more practical, since it can be used both in room calorimeters similar to the direct ones, as well as using portable devices able to analyze O_2 and CO_2 gases and determine EE. Portable indirect calorimeters typically consist of a mouth piece and a central unit where the gas analyzers reside (see Fig. 1.2), and have a battery life of $2-3$ hours. Finally, one of the most recent methods developed to measure EE is DLW. DLW is water in which both the hydrogen and oxygen have been partly or completely replaced for tracing purposes with an uncommon isotope of these elements. Since one of the two isotopes (deuterium) is eliminated only via water loss (urine and sweat mainly), and the ratio between the two isotopes is known, the amount of oxygen-18, the second isotope, can be determined easily once we know how much deuterium was expelled. Oxygen-18 exits the body via both water loss and CO_2 and therefore can be used to determine how much CO_2 was used by metabolism. Finally, assuming a known relation between CO_2 loss and oxygen (respiratory ratio), EE can be estimated. DLW is an accurate method ($3-10\%$ error) to estimate the mean total carbon dioxide production (and therefore EE) however it provides only information on total EE, not minute by minute. Therefore DLW is a valid gold standard for measurements of TEE over a period of 7 to 14 days, but does not allow researchers to determine the EE of specific activities.

Given the confined settings of room calorimeters, practical limitations of portable indirect calorimeters (mouth piece, short battery life) and the inability to provide minute by minute EE estimates for DLW, significant limitations beyond cost affect even such reference systems.

Figure 1.2: Example of indirect calorimeter, Cosmed $K4b^2$.

1.1.2 Cardiorespiratory fitness

When analyzing the importance of physical activity in health, another important aspect to consider is how performed physical activity reflects into changes in physical fitness, and health status. Is health status improving because of additional physical activity and EE? Changes in physical fitness and heath status are typically measured in terms of CRF. CRF is defined as the ability of the circulatory and respiratory systems to supply oxygen during sustained physical activity, and is among the most important determinants of health and wellbeing [113]. With regular exercise the circulatory and respiratory systems become more efficient by enlarging the heart muscle, enabling more blood to be pumped, and increasing the number of small arteries in trained skeletal muscles. As a result, more blood is supplied to working muscles, increasing the amount of oxygen carried to the muscles. CRF is not only an objective measure of habitual physical activity, but also a useful diagnostic and prognostic health indicator for patients in clinical settings, as well as healthy individuals [81]. Epidemiological research has shown that in both individuals affected by disease [119] and healthy individuals [26, 127] higher level of CRF resulted in better outcomes in term of slower disease progression, lower risk of cardiovascular disease as well as lower risk of all cause mortality [81] (see Fig. 1.3).

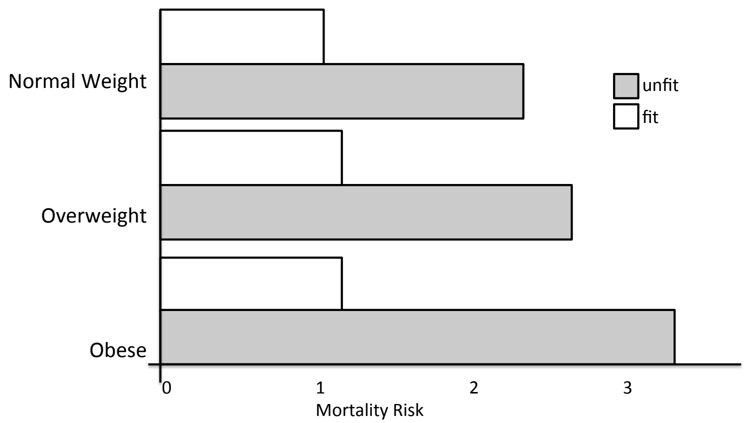

Figure 1.3: Relative risk of all cause mortality stratified by fitness and BMI. Individuals with higher BMI and higher level of fitness show reduced relative risk for all cause mortality with respect to individuals with lower BMI and lower level of fitness, highlighting the importance of fitness level for health. BMI categories (normal weight, overweight and obese) were assigned using criteria from guidelines for the evaluation and treatment of obesity: normal weight (BMI, $18.5 - 24.9$ kg/m^2), overweight (BMI, $25.0 - 29.9$ kg/m^2), or obese (BMI > 30.0 kg/m^2). Participants were labeled as unfit using MET thresholds based on age: 10.5 METs for 20-39 years, 9.9 METs for 40-49 years, 8.8 METs for 50-59 years and 7.5 METs for > 60 years or otherwise labeled as fit. Adapted from [127].

CRF measurement gold standards

Current practice for CRF measurement is direct measurement of oxygen volume (VO_2 in ml/min) during maximal exercise (i.e. VO_2max), the gold standard. Typically, VO_2max tests consist in measuring VO_2 using an indirect calorimeter during an incremental exercise test, either on a bike or treadmill [78]. However, VO_2max tests are affected by multiple limitations. Medical supervision is required and the test can be risky for individuals in non-optimal healthy conditions.

The ability to properly measure and quantify both physical activity in terms of EE and physical fitness in terms of CRF could be key in developing new applications around behavioral change and personalized coaching. By providing tailored feedback between physical activity behavior (or EE) and health markers (i.e. estimated CRF level), individuals could be helped in maintaining a healthy lifestyle. To this aim, technological solutions able to unobtrusively measure both physical activity and fitness in free-living conditions are necessary. Ubiquitous technology could avoid the impracticalities of gold standards currently employed for both EE and CRF measurements.

1.2 Current methods to estimate EE and CRF using wearable sensors

In the past few years, ubiquitous sensing technologies that could objectively monitor human behavior, started providing unprecedented insights into the relation between physical activity and health as well as driving behavioral change [52, 29, 34, 17]. Wearable sensors are getting more and more widespread due to improvements in miniaturization, battery power and user experience design, reaching ubiquitousness in the quantified-self community and being rapidly adopted by the general population. Before moving to the specifics of different methods proposed in literature to estimate EE and CRF, we provide a brief overview of the current technological solutions as well as of the wearable sensors used in this thesis.

1.2.1 Wearable sensors for EE and CRF estimation

Given the shortcomings of reference systems for EE measurement (cost, invasiveness, non-usability in free-living conditions or unreliability), the scientific community started developing early prototype of wearable sensors (electronic monitors) in the early 2000s. Over the last 10 years many electronic monitors have been introduced either on the market as commercial devices or in literature, with the aim of taking the place of the previous methods in order to accurately estimate EE in free-living conditions in a non-invasive manner.

Motion sensors were among the first to be employed. Using single axis accelerometers at the beginning, three axial accelerometers afterwards [52, 29, 118, 19, 31, 38], researchers tried to quantify physical activity and EE in different modalities. First, by exploiting the relation between whole body motion as measured by accelerometers, and EE [38, 52]. Then, by using accelerometers to distinguish activity types and then develop activity-specific EE models (see next sections for details) [29, 118, 19, 31]. Usually, accelerometers were mounted at the hip, since motion close to the body's center of mass is more representative of EE. Other body positions that were similarly linked to overall body motion, instead of e.g. limb motion [118], showed to be effective in distinguish activity types and estimate EE. Among the best-performing locations were; hip or back mounted [52, 118], chest mounted [34, 35] and ear-worn accelerometers [19, 31]. Solutions composed of multiple accelerometers were also proposed [118, 3].

Subsequently, physiological signals acquisition has been included more consistently in electronic monitors [118, 34, 96], in order to capture more accurately information about resistance work or work load effort. Signals explored in literature are GSR, skin temperature, heat flux and heart rate [130, 35, 43]. Physiological data showed consistently high correlation at the individual level during moderate to vigorous exercise, however was also very susceptible to artifacts due to factors such as psychological stress, emotions and differences among individuals [118].

ECG Necklace

In all chapters included in this thesis, we rely on imec's ECG Necklace as wearable sensor used to acquire data for our machine learning models. Thus, in this section we provide a more detailed description of the ECG Necklace hardware [98].

The low-power ECG necklace (see Fig. 1.4) monitors 1-lead ECG (bipolar) and 3D acceleration. The core of the hardware architecture component of the system is an ultra-low-power application specific integrated circuit for bio-potential readout, used to acquire ECG [134, 98]. The necklace also integrates a low-power microprocessor from Texas Instruments (MSP430) and a low-power radio from Nordic Semiconductor (nRF24L01). A 3D accelerometer from Analog Devices (ADXL330) is used to measure acceleration in three axes. An on-board SD card provides 2GB of data storage, which can last up to 2 weeks depending on the sensor's configuration. The power management unit consists of dedicated circuitry and a rechargeable Li-ion battery, giving the ECG Necklace a lifetime of 2 days to a week between charges, depending on the configuration.

The ECG Necklace provides also on board processing for beat detection. RR intervals can be extracted from the ECG signal in real-time using a beat detection algorithm based on Continuous wavelet transform. The algorithm has been optimized for robustness to motion artifact and achieves best-in-class performances, with 99.8% sensitivity and 99.77% positive predictivity on both the MIT-BIH database and a proprietary database of ambulatory ECG recordings [102]. Data gathered by the ECG necklace can be transmitted wirelessly to a PC via a USB receiver equipped with a compatible radio. Alternatively, the data can also be stored locally and later downloaded. The ECG is captured using standard snap-on electrodes connected to the necklace lead wires, typically placed in the lead II configuration. Reliability in ambulatory conditions is achieved by minimizing the effect of motion artifact using a light-weight design, and a beat detection algorithm optimized for noise robustness.

All electronics and battery are packaged in a pendent that can be worn around the neck, or attached to the waist or limbs using an elastic band. The size of the packaged necklace is $60\text{mm} \times 40\text{mm} \times 10\text{mm}$, and total weight is about 20 grams. Due to the size and attachment modality of the ECG Necklace, we explored different sensor locations on the body, however we were limited to positions where the ECG Necklace could be practically attached, and had to exclude others (e.g. the ear).

1.2.2 Methods for EE estimation

A multitude of EE estimation systems were developed in the recent past [130, 52, 34, 29, 118]. Accelerometers and heart rate (HR) monitors are the most com-

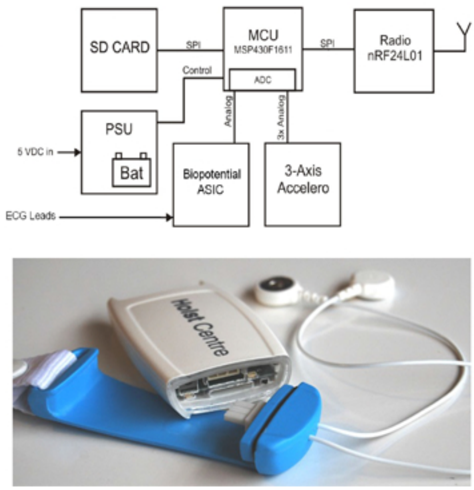

Figure 1.4: ECG Necklace system architecture and hardware prototype. The ECG Necklace could provide accelerometer and ECG data in a single sensor, plus memory storage, and was therefore used in both laboratory and free-living studies included in this thesis. Figure adapted from [98].

monly used single sensor devices in epidemiological studies. Accelerometers typically use *activity counts*, a unit-less measure representative of whole body motion intensity, as the independent variable in the linear regression model developed to predict EE [52]. The rationale being the relation between movement and EE, for weight bearing activities (see Fig. 1.5). Shortcomings of simple linear regression models are that a single model does not fit all the activities, and non-weight bearing activities cannot be correctly estimated in terms of EE. Therefore, estimates based on single accelerometer based regression models show larger error [123, 118]. Alternatively, HR monitors have been widely employed for EE estimation. HR monitors suffer from different problems, the most common being the low accuracy during sedentary behavior [43], given that HR is affected by many other factors (e.g. stress and emotions), and the need for individual calibration due to differences in fitness between individuals [34, 35].

1.2.3 Activity-specific estimation methods

The latest algorithms extended estimation methods based on single models by performing activity recognition over a predefined set of activities - or clusters of activities -, and then applying different methods to predict EE [9, 3, 118, 29, 123], based on the activity detected (activity-specific EE approaches). Other machine learning based methods were developed, trying to directly estimate EE from accelerometer features, using for example neural networks [103, 60]. However the latter approaches suffer from the same limitations of the counts-based estimation

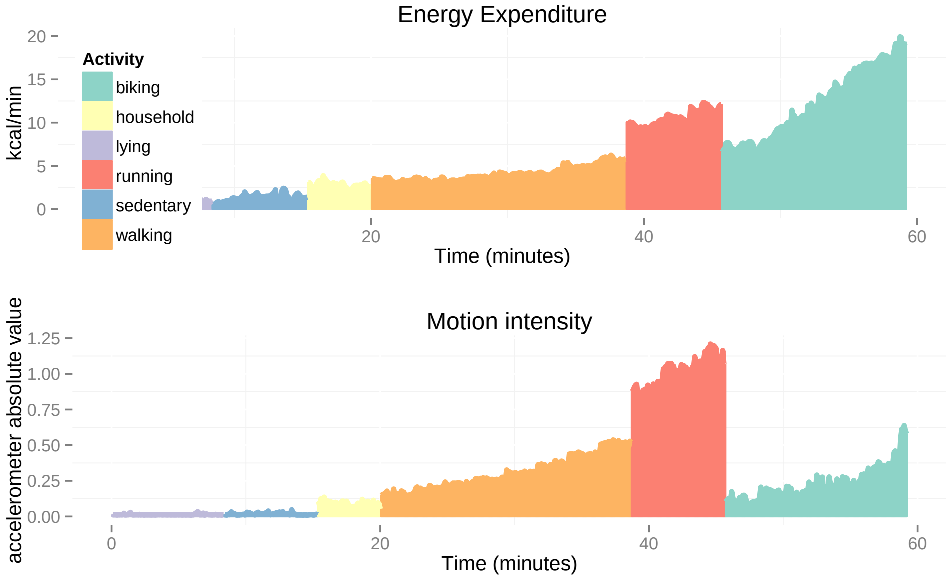

Figure 1.5: Relation between accelerometer and EE data for a series of activities performed in laboratory conditions. Walking, running and biking activities are performed at increasing intensities over time. Accelerometer data shows a strong relation with EE for weight bearing activities such as walking at different speeds. However, the relation between accelerometer and EE gets very weak for non-weight bearing activities, such as biking.

methods, being unable to capture the relation between accelerometer features and EE during different activities [30]. The most common activity-specific approaches are the following:

Activity-specific using METs lookup

One approach is to assign static MET values from the compendium on physical activities [2] to each one of the clusters of activities [29, 3], and use anthropometric features or other static features (e.g. HR at rest) to personalize the activity-specific models for different individuals or groups of individuals.

Activity-specific using wearable sensors features

Another approach is to apply a regression equation for each activity classified [118, 123], extending counts-based approaches to multiple clusters of activities. The regression models typically use accelerometer features, HR and anthropometric characteristics as independent variables.

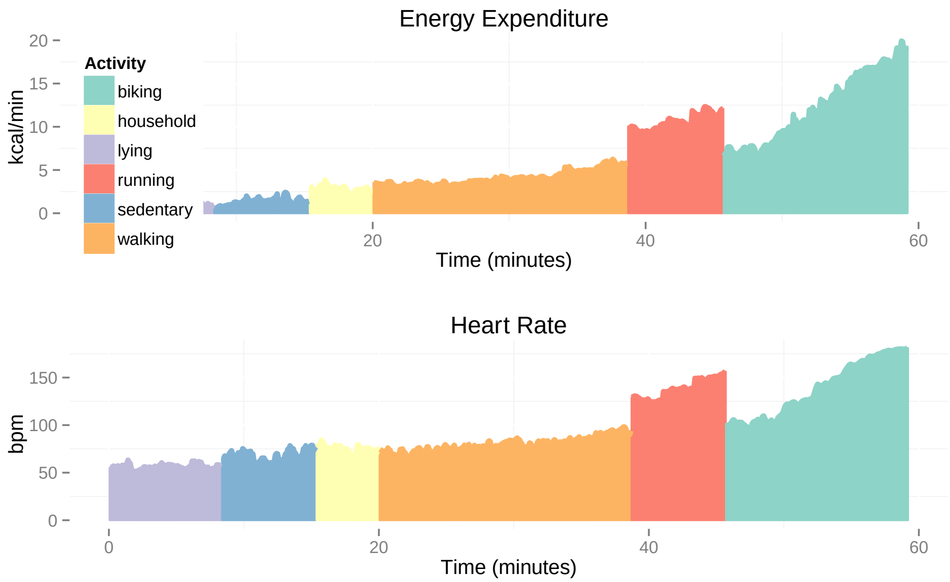

Figure 1.6: Relation between HR and EE data for a series of activities performed in laboratory conditions. Walking, running and biking activities are performed at increasing intensities. HR is highly correlated with EE, especially for moderate and vigorous activities such as walking, running and biking. We can see a weaker link between HR and EE during lying, sedentary behavior and household.

1.2.4 Methods for personalization of physiological data

Using HR in activity-specific regression equations showed consistent improvements in EE estimation compared to using acceleration only [3, 118, 34]. Additionally, when using activity-specific EE estimation models HR can be excluded from sedentary activities, therefore mitigating some of the issues due to the low correlation between HR and EE during sedentary behavior. However, the main limitation of HR based EE estimation models, i.e. the need for individual calibration, does not get resolved when using activity-specific models. HR during an activity is specific to a person since it depends on the individual's CRF level [104]. For two individuals with similar body size, EE during similar activities is comparable, as shown in Fig. 1.7, first two plots on the left end side. However, differences in HR between a sedentary and trained individual can easily reach values up to 30 bpm (or 20%) when doing the same activities, as shown in Fig. 1.7, third on fourth plot on the right end side of the figure. As a result, since HR depends on CRF, differences in HR between individuals with different CRF level can cause underestimations and overestimations of EE when developing HR based EE estimation models.

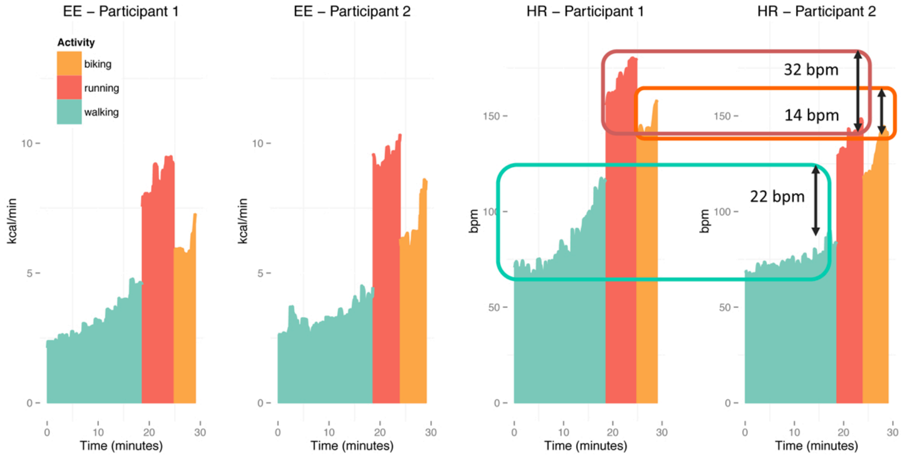

Figure 1.7: EE and HR for two individuals with similar body weight and height but different fitness level during a series of moderate to vigorous activities performed at increasing intensities. EE during similar activities is comparable, however, HR differs consistently during moderate to vigorous physical activities. Therefore, the need for HR normalization when predicting EE based on HR data.

To derive a reliable EE estimate, it is therefore necessary to normalize HR according to an individual's fitness. In turn, the normalized HR could serve as independent variable in EE regression models. Individual calibration currently limits practical applicability of HR monitors for EE estimation, since the individual relation between HR and EE needs to be determined for the algorithm to be accurate.

1.2.5 Methods for CRF estimation

CRF is typically estimated in terms of VO_2max. Given the limitations of VO_2max tests, less risky submaximal tests have also been developed [18]. Some are *non-exercise* CRF models, others are specific lab protocols performed while monitoring HR at predefined running speeds (e.g. treadmill tests) or output powers (e.g. bike tests) [18, 48, 54], without requiring maximal effort. Several *non-exercise* models of CRF have been developed using easily accessible measures such as age, sex, self reported physical activity level, body composition [69, 73]. Results typically provide decent accuracy at the group level [93]. However significant limitations apply at the individual level, since each individual is assumed to be equal to group averaged characteristics. Limited accuracy at the individual level is a common problem when physiological variables are not measured. As a result, for individuals with similar anthropometric characteristics, CRF levels cannot be discriminated, as shown in Fig. 1.8.

Submaximal tests have been developed to estimate VO_2max during specific

Figure 1.8: Relation between body weight, HR and CRF for participants with similar body size (weight and height) characteristics. a) Positive relation between VO_2max and body weight disappears when participants with similar body size characteristics are considered. b) Negative relation between VO_2max and HR while walking holds on a subset of participants with similar body size, and can potentially be used to discriminate CRF levels. Figure adapted from [4].

protocols while monitoring HR at predefined workloads, typically while running at a certain speed or biking at a certain intensity. The inverse relation between HR at a certain exercise intensity, defined by the strict exercise protocol that has to be sustained, and CRF, is the rationale behind this approach. The need for laboratory equipment and the necessity to re-perform the test to detect changes in CRF limit practical applicability of submaximal tests, since submaximal exercise tests should be re-performed every time CRF needs to be assessed.

Both maximal and submaximal tests to estimate CRF are affected by important limitations. To maximise applicability across a wide population, VO_2max should be estimated during activities of daily living, without the need for a predefined exercise protocol.

1.3 Goals of this thesis

The aim of this thesis is to introduce new methods and models to provide accurate EE estimation at the individual level without requiring individual calibration and to estimate VO_2max, in free-living conditions. More specifically, the following goals were investigated:

1.3.1 Selection of methods, sensor number and positioning

Recent work [20, 95, 19] showed that activity type can be reliably detected with wearable sensors, opening new opportunities for EE modeling beyond simple linear regression models. Over the last years, a few activity-specific algorithms have been reported [118, 29, 3]. What is not clear at this stage, is the methodology to follow when developing such an algorithm. There is currently little agreement in literature, on which activities to detect. Some [3, 118] used multi-accelerometer systems and extensive protocols to detect a large amount of activities (26 and 52 respectively). Others [29], developed single accelerometer systems able to recognize a smaller set of activities with higher accuracy. Once the activity set has been selected, even less agreement is found on how to predict EE given an activity. Some works assign static Metabolic equivalents (METs), the ratio of metabolic rate during a specific activity to a reference metabolic rate), combined with the subjects' anthropometric parameters. Others applied a linear regression equation for each model. Moreover, little agreement is found in literature regarding number of accelerometers, location on the body, and the role of accelerometer features (e.g. used for activity recognition only (activity-specific models using METs lookup), or for both activity recognition and activity-specific EE models (activity-specific using accelerometer features)). Determining the optimal method, number of sensors and on-body positioning of accelerometers to accurately estimate EE requires addressing the following issues related to the influence of activity type misclassification on EE estimation error, activity-specific models performance during different activities, activity recognition accuracy and EE estimation error performance depending on sensors number and positioning.

1.3.2 Physiological data normalization

Normally, activity-specific equations use accelerometer features and anthropometric characteristics to predict EE. Some activity-specific algorithms use equations where HR or other physiological parameters (e.g. galvanic skin response, GSR, skin temperature, etc.) are included as well [34, 35, 3, 118, 130], showing consistent improvement compared to accelerometers alone. However, inter-individual differences in physiology, as well as the consequent need for individual calibration, limit accuracy and practical applicability of such systems. For example, during moderate to vigorous activities, differences in HR between individuals performing the same activity are mainly due to CRF. Combined with activity-specific algorithms, information on CRF could provide more accurate EE estimation. Nevertheless, algorithms in the past tackled CRF-related variance only by means of individual calibration [35], and no algorithm includes information on CRF in the EE estimation equations. While CRF is the main factor driving changes in HR during physical exercise [121], differences in respiration, skin temperature or GSR might be caused by different underlying processes or characteristics of the person [109]. Breaking down the EE estimation process into activity-specific sub-problems is not sufficient to take into account the different relation between physiological signals and EE in different individuals. Therefore, new methods are needed to automatically normalize physiological signals without requiring individual calibration and fully exploit the relation between physiological signals and EE, reducing EE estimation error.

1.3.3 VO_2max estimation using wearable sensor data

While EE is the most commonly used single metric to quantify physical activity, with many algorithms proposed in the recent past [118, 29, 9, 67], CRF is not only an objective measure of habitual physical activity, but also a useful diagnostic and prognostic health indicator for patients in clinical settings, as well as healthy individuals [81]. EE and CRF are tightly coupled when EE estimation is performed based on HR data acquired using wearable sensors. The inverse relation between HR and CRF is one of the main causes behind the need for individual calibration of HR monitors, since differences in CRF cause differences in HR but not in metabolic responses (EE) [104]. Thus, CRF estimation could both provide a relevant health marker and be used to personalize EE estimation models, improving estimation accuracy. Instead of performing VO_2max tests or submaximal tests involving specific exercise intensity at which HR is measured and then use the collected HR during predefined exercise to predict CRF, non-invasive and unsupervised methods could be developed. For example, using wearable sensor data and machine learning techniques specific contexts could be automatically determined during activities of daily living. Thus, new methods using context-specific HR during activities of daily living as predictor for CRF could be developed.

1.3.4 Personalized EE estimation and VO_2max estimation in free-living

In free-living conditions, contextualizing and interpreting physiological data is challenging, due to the effect of both low-level activity primitives (e.g. lying down, walking, etc.) and high-level activity composites (e.g. commuting, working, socializing, etc.) on physiological data. Almost no previous work addressed the need for normalization of physiological data in EE estimation as well as VO_2max estimation in free-living conditions [100]. The person-specific nature of high level activity composites brings additional challenges, such as the inability of reliably using supervised methods for activity classification. Thus, the need for the development of new methods able to bring algorithms developed under controlled laboratory conditions to free-living settings and practical use. To this aim, ways to incorporate high level activity-composite information together with low-level activity and reliably interpret physiological data such as HR in free-living conditions, should be developed.

1.4 Thesis outline

This section provides an overview of how the thesis contributions are organized in the different chapters.

1.4.1 Selection of methods, sensor number and positioning

State of the art methods for EE estimation are reviewed, implemented and compared in Chapter 2. We analyze the accuracy of simple linear regression models as well as different types of activity-specific models. Additionally, we also analyze the impact of different sensor modalities, such as accelerometer and HR data in reducing EE estimation error. As a result, we developed a methodology combining different activity-specific approaches to reduce the estimation error. Our method uses as predictors static EE values for sedentary activities and a combination of accelerometer and HR features for dynamic activities. Additionally, in Chapter 3, we evaluated performance of the proposed methodology and other comparison methods on a dataset comprising five on-body sensor locations and nearly fifty activities of varying intensities, analyzing trade-offs between estimation methods, sensor number and positioning.

1.4.2 Physiological data normalization

Chapters 4, 5 and 6 introduce methods to normalize physiological data without the need for individual calibration. In Chapter 4 we introduce a method to personalize EE estimations relying on HR data. We first contextualize HR during low intensity activities, i.e. determine HR while walking at a specific speed. We contextualized HR by combining an activity recognition classifier and a regression model for walking speed estimation. Then, contextualized HR is used to predict a normalization parameter, thus avoiding the need for individual calibration. The proposed method using contextualized HR data as predictor for normalization parameters is then extended to determine the importance of HRV features and different activity intensities in Chapter 5. Finally, the proposed normalization methodology is extended to other physiological signals, such as galvanic skin response and respiration rate, as a generic methodology proposed in Chapter 6.

1.4.3 VO_2max estimation using wearable sensor data

In Chapter 7, we develop methods and models to estimate health markers, such as VO_2max, without the need for specific protocols. A method is developed to estimate VO_2max from contextualized HR data collected during activities of daily living, simulated in laboratory settings. A hierarchical Bayesian regression approach is used with model coefficients that vary depending on the performed activity. In this way participants are not constrained to specific activities. Given the tight relation between HR, VO_2max and EE, an EE estimation model is also developed to account for between-individual variability in HR without the need for explicit HR normalization. We propose another hierarchical Bayesian model which uses a hierarchical structure with nested and non-nested groupings to estimate EE using coefficients varying by both activity type (as in activity-specific methods) and participant-level characteristics, such as the predicted VO_2max.

1.4.4 Personalized EE estimation and VO_2max estimation in free-living

Finally, we introduce methods to personalize EE estimation regression models and estimate VO_2max in unsupervised free-living settings, outside of the laboratory environment. We cover this objective in Chapters 8, 9 and 10. In Chapter 8, a method is developed to combine low-level activity primitives and high-level activity composites using topic models. Optimal contexts (i.e. combinations of activity primitives and composites) for analyzing HR were determined without supervision. We show that the proposed method can be used to estimate HR normalization parameters in free-living conditions, and therefore personalize laboratory derived EE models. In Chapter 9, we also apply the proposed method to determine contextualized HR in free-living conditions to the application of VO_2max estimation, showing reduced estimation error compared to existing methods. Finally, in Chapter 10, we further analyze context-specific HR in both laboratory and free-living settings, showing how VO_2max estimation can be obtained with the same accuracy in both conditions, and therefore laboratory protocols are not required.

1.5 Chapter overview of the thesis

Ch.	Thesis goal	Related publication
2	1: Selection of methods, sensor number and positioning	M. Altini, J. Penders, and O. Amft. Energy expenditure estimation using wearable sensors: A new methodology for activity-specific models. In Proceedings of the Conference on Wireless Health, WH '12, pages 1:1-1:8, New York, NY, USA, 2012. ACM.
3	1: Selection of methods, sensor number and positioning	M. Altini, J. Penders, R. Vullers, and O. Amft. Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning. IEEE Journal of Biomedical and Health Informatics, 19(1):219-226, 2015.
4	2: Physiological data normalization	M. Altini, J. Penders, and O. Amft. Personalizing energy expenditure estimation using a cardiorespiratory fitness predicate. In Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on, pages 65-72. IEEE, 2013.
5	2: Physiological data normalization	M. Altini, J. Penders, R. Vullers, and O. Amft. Automatic heart rate normalization for accurate energy expenditure estimation. Methods Inf Med, 53(5):382-388, 2014.
6	2: Physiological data normalization	M.Altini, J.Penders, R.Vullers, and O.Amft. Personalizing energy expenditure estimation using physiological signals normalization during activities of daily living. Physiol Meas, 35(9):1797, September 2014.
7	3: VO_2max estimation using wearable sensor data	M. Altini, P. Casale, J. Penders, and O. Amft. Personalized cardiorespiratory fitness and energy expenditure estimation using hierarchical Bayesian models. Journal of Biomedical Informatics, 56:195-204, 2015.
8	4: Personalized EE estimation and VO_2max estimation in free-living	M. Altini, P. Casale, J. Penders, and O. Amft. Personalization of energy expenditure estimation in free living using topic models. IEEE Journal of Biomedical and Health Informatics, 19(5):1577-1586, 2015.
9	4: Personalized EE estimation and VO_2max estimation in free-living	M. Altini, P. Casale, J. Penders, and O. Amft. Cardiorespiratory fitness estimation in free living using wearable sensors. Submitted to Artificial Intelligence in Medicine.
10	4: Personalized EE estimation and VO_2max estimation in free-living	M. Altini, P. Casale, J. Penders, G. ten Velde, G. Plasqui, and O. Amft. Cardiorespiratory fitness estimation using wearable sensors data: analysis of context-specific submaximal heart rates. Submitted to the Journal of Applied Physiology.

Part I: Selection of methods, sensor number and positioning

2

Energy expenditure estimation using wearable sensors: a new methodology for activity-specific models

M. Altini, J. Penders, O. Amft
Adapted from: proceedings of the Conference on Wireless Health, ser. WH '12. New York, NY, USA: ACM, 2012, pp. 1:1-1:8.
Runner-up best paper award.

Abstract

Accurate estimation of Energy Expenditure (EE) in ambulatory settings is a key element in determining the causal relation between aspects of human behavior related to physical activity and health. We present a new methodology for activity-specific EE algorithms. The proposed methodology models activity clusters using specific parameters that capture differences in EE within a cluster, and combines these models with Metabolic Equivalents (METs) derived from the compendium of physical activities. We designed a protocol consisting of a wide set of sedentary, household, lifestyle and gym activities, and developed a new activity-specific EE algorithm applying the proposed methodology. The algorithm uses accelerometer (ACC) and heart rate (HR) data acquired by a single monitoring device, together with anthropometric variables, to predict EE. Our model recognizes six clusters of activities independent of the subject in 52.6 *hours of recordings from* 19 *participants. Increases in EE estimation accuracy ranged from* 18 *to* 31% *compared to state of the art single and multi-sensor activity-specific methods.*

2.1 Introduction

Lack of physical activity is one of the major health problems in most of the western world. Even though our genome has not changed much over the last ten thousand years and more [49], activity patterns of our hunter-gatherers ancestors have been first modified by the agricultural and industrial revolution, and then completely disrupted by the shift towards computer-based work which took place over the past twenty years. As a result, two thirds of the world population is overweight and obesity affects a third of the population in the US at present. Other diseases, such as diabetes, are rapidly becoming widespread epidemics as well [133]. Accurate quantification and assessment of habitual physical activity in ambulatory settings is essential in order to find subtle but important links between not only sedentary time, but all the aspects of habitual physical activity, and health [84]. New technologies, seamlessly integrated in everyone's life, able to monitor objectively and non-invasively our behavior, can provide unprecedented insights on these links.

Currently, epidemiologists use accelerometers [52] and HR monitors [43] to objectively gather information about physical activity. Traditionally, they make use of regression equations developed using data acquired over a certain protocol [21, 43, 52] to predict EE. For accelerometers, the rationale behind this approach is that body motion measured close to the body center of mass, is linearly related to EE. On the other hand, HR monitors exploit the linear relation between HR and oxygen uptake. Limitations of these approaches are the inability of single accelerometers worn close to the body center of mass to detect low and upper body motion [43, 118], the low accuracy of HR monitors during sedentary behavior and the need for individual calibration [34].

Recent work [20] showed that activity type can be reliably detected with wearable sensors, opening new opportunities for EE monitors. Over the last years, a few activity-specific algorithms have been reported [3, 29, 107, 118]. They first recognize the activity performed, and then apply a model developed for the specific activity, showing consistent improvements compared to previous methods. What is not clear at this stage, is the methodology to follow when developing such an algorithm. There is currently little agreement in literature, on which activities to detect. Some [3, 118] used multi-accelerometer systems and extensive protocols to detect a large amount of activities (26 and 52 respectively), exploiting the fact that frequent misclassification of the activities will most likely result in small EE errors, due to the similarity in the movement involved. Others, [29, 107] developed either multi-sensor or single accelerometer systems able to recognize a smaller set of activities with higher accuracy. Once the activity set has been selected, even less agreement is found on how to predict EE given an activity. Some works assign static *Metabolic equivalents* (METs, the ratio of metabolic rate during a specific activity to a reference metabolic rate), combined with the subjects' anthropometric parameters or fitness indicators such as the HR at rest [3]. Others applied a linear regression equation for each model [123, 118, 107]. At this stage, it is not clear whether combining METs values and regression models could provide better estimates [122], and whether each activity-specific regression model requires the same parameters. As a matter of fact, differences in protocols and evaluation measures make it impossible to compare the different approaches.

In this paper we present a methodology which aims at clarifying the relation between physical activity patterns detectable with wearable sensors, and EE. Our paper includes four main contributions. *a)* We propose a new methodology, which combines METs values from the compendium of physical activities [2] with regression equations, depending on the type of activity. *b)* We show that by carefully selecting activity-specific features able to explain differences in EE within the activity, EE estimations can be improved. *c)* We develop a new algorithm applying our methodology to the case of a single monitoring device able to measure acceleration and HR. *d)* We compare our algorithm to models used in epidemiological studies, as well as to state of the art activity-specific EE methods.

2.2 Related work

Epidemiological studies

Accelerometers and HR monitors are the most commonly used single sensor devices in epidemiologic studies. Accelerometers use *activity counts*, a unit-less measure representative of whole body motion, as the independent variable in the linear regression model developed to predict EE [52]. Shortcomings of single regression models are; *a)* the accuracy of the monitor is highly dependent on the activities used to develop the model, *b)* a single model does not fit all the activities, since the slope and intercept of the regression model change based on the activity performed while data is collected [123, 118]. As a result, even when activity counts are representative of EE, the output can be misleading. Additionally, activity counts are defined differently by each sensor's manufacturer (i.e. *Actigraph counts*, and the equations derived from them, are not directly comparable to *Actical* or *Actiheart counts* [129]). HR monitors suffer from different problems, the most common being the low accuracy during sedentary behavior [43], given that HR is affected by many other factors (e.g. stress and emotions), and the need for individual calibration [34]. Some of the issues have been tackled developing models that use more than one linear regression equation, such as Crouter's 2-regression model [51] or Brage's [35] branched equations. Even though these methods are promising, especially the ones combining HR and ACC data, they have shown limited improvements compared to ACC based simple linear regression models [21, 115].

Methods based on machine learning

The latest monitors go towards two directions. Both strategies make use of pattern recognition and machine learning techniques. Some authors applied these methods to directly estimate EE from ACC features [60, 103]. Others, extended

Crouter's and Brage's approach, performing activity recognition over a pre-defined set of activities, and then applying different methods to predict EE [3, 29, 123, 107, 118]. Given the significant amount of work adopting activity recognition as a first step to estimate EE, and the consistent improvements obtained [30], we believe this is the best methodology to follow when developing such algorithms. The principle behind activity recognition as a first step in EE estimation is that the slope and intercept of the regression models change based on the activity performed [123]. Tapia [118] developed a system composed of three accelerometers and applied a different regression equation for each activity classified. The regression models use ACC features as independent variables. The system can recognize about fifty activities with 50% accuracy in a subject-independent manner. *Root Mean Square Error* (RMSE) was reduced from 2.7 to 1.0 $METs$ compared to Crouter's approach. Bonomi [29] proposed a similar approach, but with a single sensor device, mounted on the lower back. His system recognizes six clusters of activities and assigns a MET value to each one of them. *Total Energy Expenditure* (TEE) was validated against *Doubly Labeled Water* and showed accuracy up to $1MJoules/day$ when simple anthropometric parameters are used as independent variables together with the assigned METs. van Hees [123] also used a single sensor able to distinguish four activities, and then applied linear regression using a measure of motion intensity as the independent variable (similarly to Tapia). Albinali [3] developed a multi-sensor system, composed of three accelerometers able to distinguish twenty-two activities with 26% accuracy (subject-independent). He extended Bonomi's approach, developing a custom MET table, which takes into account anthropometric variables, as well as the HR at rest, to predict EE more accurately at the individual level. This method showed improvements up to 15% compared to non-activity-based models. Rumo [107] also combined HR and ACC. His system consists of three sensors, two accelerometers and a HR belt, and can classify seven types of activities. Manual selection and the bootstrapping method were used to determine which independent variables to adopt for the activity-specific models. RMSE for the individual models ranged from 2.2 to 9.7 $KJoules$.

Towards activity-specific EE estimation

In this section we analyze shortcomings of state of the art activity-specific EE algorithms. More specifically, we believe the following limitations should be tackled; *a)* Activity-specific models that assign METs values to each activity classified assume that EE is constant within a cluster of activities [3, 29]. Nevertheless, most activities can be performed at different intensities, and including information about whole body motion or other features (ACC or HR) representative of variations in EE within an activity, would improve the estimate. *b)* Activity-specific regression models apply linear regression (e.g. using activity counts) even though there is no whole body motion involved [118], and therefore the motivation for applying linear regression, which is the linear relation between intensity of motion and EE, does not hold anymore. *c)* Activity-specific models that assign custom METs values [3] should carefully select the independent variables used. For example HR at rest, which is often used as an index of *cardiorespiratory fitness*, should not be used to predict EE at rest, since cardiorespiratory fitness is not related to *Basal Metabolic Rate* (BMR). Care should be taken when considering anthropometric variables as well. The energy cost of activities such as walking depends on body weight, while the assumption does not apply to biking or sitting.

2.3 A new methodology

In this section we present our new approach to activity-specific EE modeling, and apply this methodology to the case of imec's ECG necklace. Our approach to combine static METs with activity-specific regression equations takes four steps to derive an EE model. First, we categorize activities into clusters meaningful for EE estimates. Secondly, we separate sedentary and non-sedentary activities, and assign a static EE value to sedentary ones. Then, we examine the motion patterns of non-sedentary clusters to select the best independent variables for the prediction models. Finally, we include anthropometric characteristics to take into account differences in body size (see Figure 2.1).

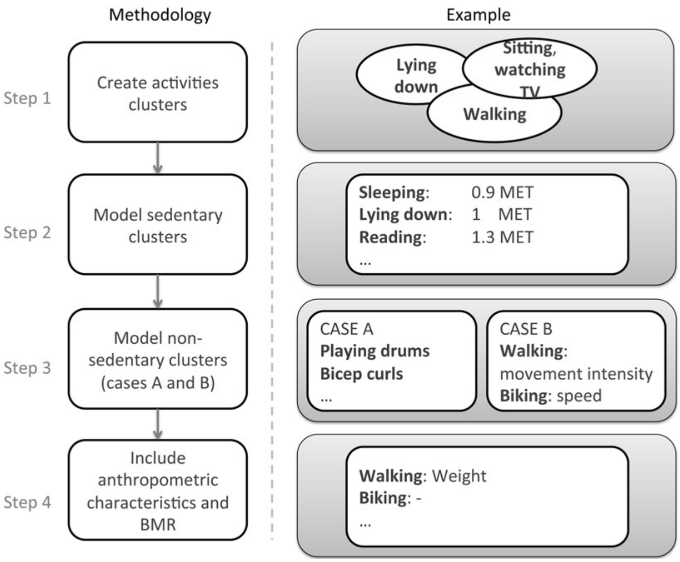

Figure 2.1: Block diagram of the proposed methodology, and example of an application.

Step 1 - Create activities clusters

Different postures provoke different levels of EE, due to the energy cost of holding a specific posture [1] and should be individually distinguished. *Activities clusters* (i.e. groups of activities) should contain at least the basic human postures (e.g. lying, sitting and standing), in order to avoid ambiguous results when using the monitor to detect activities that were not part of the original training set. The detection of more activities, which are often very specific (e.g. brushing teeth, stretching, fidgeting hands, etc.) requires validation on a different population and in real-life settings, to evaluate the *sensitivity* and *specificity* of such activities. If more sensors are used, the distinction between sedentary and non-sedentary activities while holding a specific posture (e.g. sitting resting or sitting lifting weights) will most likely produce better EE estimates, since different models could be applied. When designing an activity-specific EE algorithm, we recommend to group activities into clusters containing at least the following postures: lying, sitting and standing. For a specific posture, we recommend to distinguish ambulation and transportation, as well as to separate between resting and non-resting activities whenever possible.

Step 2 - Model sedentary clusters

Sedentary activities (e.g. lying down, sitting resting, watching TV, etc.) are such that between-individual differences in EE cannot be further explained by ACC features or HR data, but only by anthropometric variables. Therefore we assign to sedentary-only clusters (i.e. clusters which contain only resting activities) a static EE value, derived from the compendium of physical activities.

Step 3 - Model non-sedentary clusters

Our methodology splits modeling of non-sedentary clusters into two parts. Non-sedentary activities may or may not involve specific patterns of motion representative of changes in EE that an accelerometer can detect. For example, non-sedentary activities such as some gym exercises or playing instruments might involve no whole body motion or other specific motion pattern representative of changes in EE (case A). Other non-sedentary activities (e.g. household, walking, etc.) involve a different amount of whole body motion depending on the intensity of the action (case B).

Case A: We predict EE for clusters of activities that do not involve whole body motion or specific ACC patterns using only physiological signals as independent variables.

Case B: For the remaining clusters of activities, the independent variables used to predict EE will include ACC features as well. Which features to introduce will depend on the activity performed. The question to answer in order to develop a good model is the following: what ACC features are representative of changes in EE within the cluster? Whole body motion is a good candidate for most of the activities that involve significant movement (e.g. walking). Other activities, such as biking, do not involve whole body motion, but show specific patterns that ACC features can capture. When a cluster contains activities with similar motion patterns but different EE, physiological signals should be used in the model to capture the remaining variance.

Step 4 - Include anthropometric characteristics and resting metabolic rate

Once activities have been clustered, split between sedentary and non-sedentary, and the independent variables for each cluster have been selected, anthropometric characteristics and RMR should be taken into account as well. The type of anthropometric characteristics used for each cluster model depends on the activity performed. For example, the EE of walking and running is related to body weight, while non-weight bearing activities, such as biking, are not. If the dependent variable of the regression models is TEE, RMR should be included in the models.

2.4 Use Case - A necklace monitor

In this section we apply our proposed methodology to the use case of necklace which combines ACC and HR data in a single monitoring device. Details on the necklace and the experimental protocol can be found in Sec. 2.5 and 6.5.3.

Method Implementation

Step 1 - Create activities clusters

Different clusters of activities have been evaluated based on their impact on EE and the ability of a single device placed on the chest to detect them. By using a single monitoring device located at the chest it was not feasible to differentiate between sitting and standing. Therefore we grouped activities in the following clusters: *lying, low whole body motion (LWBM), high whole body motion (HWBM), walking*, *biking* and *running* - see Table 2.1. LWBM and HWBM are clusters similar to the *sitting-standing* and *active standing* introduced in [29] to distinguish between sedentary and household activities with a single accelerometer, and are useful in isolating sedentary behavior even when sitting and standing cannot be distinguished.

Activities clusters classification: Four pattern recognition methods were tested on the six clusters of activities: Classification Trees (C4.5), Artificial Neural Networks, Support Vector Machines, and Naive Bayes. The best performance was obtained by the C4.5 classification tree, which was used for the EE model.

Speed estimation: We estimated speed using multiple linear regression. Independent variables for walking and running included both ACC and anthropometric features. Biking speed was predicted by ACC features only.

Table 2.1: Distribution of the activities into the six clusters used for activity recognition.

Cluster name	Original activities
Lying	Lying down resting
LWBM	Sitting resting, sitting stretching, standing stretching, desk work, reading, writing, working on a PC, watching TV, sitting fidgeting legs, standing still, bicep curls, shoulder press
HWBM	Stacking groceries, washing dishes, preparing a salad, folding clothes, cleaning and scrubbing, washing windows, sweeping, vacuuming
Walking	self-paced, self-paced carrying books, treadmill (flat: $3, 4, 5, 6$ km/h, 4 km/h carrying weights, incline: $3, 5km/h, 5, 10\%$))
Biking	Cycle ergometer, low, medium and high resistance level at 60 and 80 rpm
Running	$7, 8, 9, 10$ km/h on a treadmill

Step 2 - Model sedentary clusters

The only sedentary-only cluster of our model is *lying*, which was assigned a value of 1 MET. The LWBM cluster contains resting activities mixed with non-sedentary activities, therefore it was modeled as a non-sedentary cluster.

Step 3 - Model non-sedentary clusters

Using a single monitoring device sets limits regarding the number of distinguishable activities. Thus, the other activity clusters contain some variability in EE and have been modeled as non-sedentary clusters (Case B in Sec 2.3). What ACC features are representative of variations in EE within these cluster? LWBM and HWBM are clusters involving diverse and irregular motion patterns, that we captured using features representative of intensity and variability of motion over the three axes. Walking, running and biking involve repetitive patterns, that can be easily captured using measures of motion intensity or motion speed (see Table 2.4 for details). All models include the Heart Rate above Rest (HRaR) to complement acceleration features in capturing differences in EE, e.g. walking vs. walking carrying weights.

Step 4 - Include anthropometric characteristics and resting metabolic rate

Body weight was included for all the clusters involving ambulation (HWBM, walking and running). No anthropometric variables were included for lying, LWBM and biking. RMR was in included in models not involving ambulation (lying, LWBM and biking). We computed RMR using simple anthropometric variables only (gender, age, weight and height). The Harris-Benedict formula estimates BMR, which is between 10 and 20% lower than RMR [21]. Therefore, we chose

Table 2.2: Study participants' characteristics

characteristic	mean ± std	range
Age ($years$)	29.5 ± 4.6	$24 - 39$
Height (m)	1.76 ± 0.11	$1.59 - 1.97$
Weight (kg)	72.7 ± 14.7	$50.2 - 102.1$
BMI (kg/m^2)	23.3 ± 3.1	$18.6 - 28.7$

to increase BMR by 15%, to estimate RMR.

2.5 Data collection and analysis

Participants

Participants were 19 (14 male, 5 female) healthy imec-nl employees from diverse ethnic background - see Table 2.2. Imec's internal Ethics Committee approved the study, and each participant signed an informed consent form.

Instruments

ECG Necklace

The ECG Necklace [98] is a low power wireless ECG platform. The system relies on an ultra-low-power ASIC for ECG read-out, and it is integrated in a necklace, providing ease-of-use and comfort while allowing flexibility in lead positioning and system functionality. It achieves up to 6 days autonomy on a 175 mAh Li-ion battery. For the current study, the ECG Necklace was configured to acquire one lead ECG data at 256 Hz, and ACC data from a three-axial accelerometer (ADXL330) at 64 Hz. The sensor was placed on the chest with an elastic belt. The *x*, *y*, and *z* axes of the accelerometer were oriented along the vertical, medio-lateral, and antero-posterior directions of the body, respectively. Two gel electrodes were placed on the participant's chest, in the lead II configuration. Data were recorded on the on-board SD card to ensure no data loss. Data were also streamed in real-time to provide visual feedback of the system functionality to the experimenter.

Indirect calorimeter

Breath-by-breath data were collected using the *Cosmed* $K4b^2$ indirect calorimeter. The Cosmed $K4b^2$ weights 1.5 kg, battery included, and showed to be a reliable measure of EE [85]. The system was manually calibrated before each experiment according to the manufacturer instructions. This process consists of allowing the system to warm-up, following a double calibration, first with ambient air and then with calibration gas values. A delay calibration was performed weekly to adjust for the lag time that occurs between the expiratory flow measurement and the gas analyzers.

Experimental design

Participants were invited for recordings on two separate days. They reported at the lab at $8.00\ a.m.$, after refraining from drinking (except for water), eating and smoking in the two hours before the experiment. The protocol included a wide range of lifestyle and sport activities, including sedentary and household activities. More specifically, day one consisted of activities selected as representative of common daily leaving of many people in industrialized countries [21]. The activities were: *lying down, resting, sitting stretching, standing stretching, desk work, reading, writing, working on a PC, watching TV, fidgeting legs, standing still, standing preparing a salad, washing dishes, stacking groceries, folding clothes, cleaning the table, washing windows, sweeping, vacuuming, walking self-paced, walking self-paced carrying books ($4.5\ kg$), climbing stairs up, climbing stairs down*. Each sedentary and household activity was carried out for a period ranging from 4 to 12 minutes, with a 1 or 2 minutes break between the activities. Day two was carried out at the gym, where subjects performed a series of more vigorous activities, including: *step-test, biceps curls, shoulder press, walking at 3,4,5 and 6 km/h on a treadmill, walking at 4 km/h carrying a weight (5% of the subject's weight), walking at 3 km/h, 5 and 10% inclination, walking at 5 km/h, 5 and 10% inclination, cycle ergometer at 60 and 80 rpm, low, medium and high resistance levels, running at 7,8,9 and 10 km/h*. Activities carried out at the gym were 4 minutes duration, except for free weights and running, which lasted for 1 to 2 minutes.

Study design choices

We included a wide set of activities, ranging from sedentary to vigorous, recorded in laboratory settings. Even though performance for activities that were not part of the dataset should be assessed outside of the lab, there is currently no reference system able to measure breath-by-breath EE in unconstrained settings. For example, DLW – which is the standard reference system for EE in daily life – provides only TEE after one or two weeks, averaging under and over-estimations. Thus, it provides limited information about the algorithm performance under different conditions, which is key in understanding advantages of activity-specific models.

Pre-processing

The dataset acquired in this work contains 52.6 hours of annotated data collected from nineteen subjects, consisting of reference VO_2, VCO_2, three axial acceleration and ECG.

ECG Necklace data

Raw ECG and ACC data were downloaded from the SD card of the ECG Necklace using proprietary software developed by imec-nl. Raw data were exported into *csv* files containing time-stamped ECG and acceleration samples. A *Continuous*

Wavelet Transform based beat detection algorithm was used to extract R-R intervals from ECG data, which output was examined to correct for missed beats [102].

Indirect calorimeter data

Breath-by-breath data acquired from the Comsed $K4b^2$ was resampled at $0.5\ Hz$. EE was calculated from O_2 consumption and CO_2 production using Weir's equation [128]. The first 1 or 2 minutes of each activity were discarded to remove non-steady-state data.

Feature extraction

Features extracted from the ECG necklace raw data were used to derive activity recognition and EE models. Activity recognition was performed on the six activity clusters introduced in Sec. 6.6.1.1. An activity-specific EE model was derived for each cluster. ACC data over the three axes were segmented in 4 second windows, band-pass (BP) filtered between 0.1 and $10\ Hz$, to isolate the dynamic component caused by body motion, and low-pass (LP) filtered at $1\ Hz$, to isolate the static component, due to gravity. Time and frequency features were extracted from each window over the three axes of the LP and BP signal. Time features included *mean, mean of the absolute signal, magnitude, mean distance between axes, skewness, kurtosis, variance, standard deviation, coefficient of variation, range, min, max, correlation, inter-quartiles range, median* and *zero crossing rate*. Frequency features included: *spectral energy, entropy, low frequency band signal power* $(0.1 - 0.75\ Hz)$*, high frequency band signal power* $(0.75 - 10\ Hz)$*, frequency and amplitude of the FFT coefficients*. These features were selected due to high accuracy showed in past research [20, 29, 118].

Three features were extracted from R-R intervals, computed over 15 seconds windows; *mean, variance and standard deviation*. Additionally, sleep HR was derived from the HR while lying down [34], and used to extract the HRaR. R-R intervals features were not included in the activity recognition model. Feature extraction was performed in MATLAB (MathWorks, Natick, MA).

Feature selection

Feature selection for the activity recognition model was performed according to different criteria. First of all, we removed features that depend on the range and sensitivity of the accelerometer used to ease implementation of the algorithm on different hardware. Secondly, we evaluated features based on the individual predictive ability of the feature alone, along with the degree of redundancy between them. This step was implemented in Java using libraries provided by the WEKA machine learning toolkit (University of Waikato, Hamilton, New Zeland).

The final feature set was manually selected, taking into account the output from the automatic feature selection scheme when features showed high correlation. It includes: *mean of the absolute band-passed signal and inter-quartile range* — which capture the intensity of whole body motion, *mean distance between axes* *and median* — which capture posture information, *variance and standard deviation* — measures of the spread of the distribution, *zero crossing rate and main frequency peak* — which provide useful information on the repetitive pattern of certain activities, *low and high frequency band signal power*. We manually selected features for the EE models, according to the criteria illustrated in Sec. 2.4.

Statistics and performance measures

Activity recognition

Performance of the activity recognition model was evaluated independent of the subject, using *leave-one-subject-out-cross-validation*. Metrics used are the *sensitivity* and *specificity* of the recognition of each activity, as well as the *percentage of the correctly classified instances* over the entire set used for validation. Walking, biking and running speeds were evaluated according to the *Root Mean Square Error*.

Energy expenditure

Performance of the EE models were evaluated in a subject independent fashion, developing regression models on all the subjects but one, and validating them on the remaining one. The procedure was carried out N times (N = number of subjects), and results were averaged. The performance measures used is the RMSE, averaged within an activity and between subjects. Results are reported only in terms of RMSE because of the great between-subject variability typical of EE estimates, which makes averages predictions between subjects less informative than the average error. Normalization procedures do exist (e.g. estimating in $kcal/kg$), but do not take into account that EE during different activities is affected differently by body weight.

Comparisons

Reported performance of EE models are highly dependent on the protocol used to validate the algorithms, which makes it impossible to compare different models from published results. We re-implemented six methods; two simple methods used in epidemiological studies, using ACC (*method ACC* [52]) or HR (*method HR* [43]) as independent variable of the regression model, and four activity-specific (AS) EE algorithms. The four models derive EE assigning static values to the detected activity (*method AS-static* [29, 3]), using a single linear regression model per activity and a measure of Motion Intensity (MI) as the independent variable (*method AS-MI* [118, 123]), combining ACC and HR features following automatic variables selection (*method AS-mixed* [107] - where HR is always used, and accelerometer features are part of one model only) or following the proposed methodology (*method AS-new*). To the best of our knowledge, this is the first comparison of state of the art activity specific models on the same dataset.

2.6 Results

Activity clusters classification

Subject independent classification accuracy of the classification tree used to select the cluster model to apply to estimate EE was 92.9%. Table 2.3 shows the performance of the classifier in terms of sensitivity and specificity for the six clusters. RMSE for walking, running and biking speed were $0.31\ km/h$, $0.77\ km/h$ and $8.43\ rpm$. Biking speed errors can be reduced increasing the frequency resolution (i.e. using windows ¿ 4 seconds). Utilizing 4 seconds window our system cannot detect speeds other than multiples of 0.25 Hz.

Table 2.3: Classification performance of the C4.5 classifier used to select the cluster model to predict EE.

Activities Cluster	**Sensitivity**	**Specificity**
Lying	1	0.99
LWBM	0.91	0.97
HWBM	0.87	0.95
Walking	0.98	0.99
Running	0.99	0.99
Biking	0.91	0.99

Activity clusters models

We derived six models (see Table 2.4), applying the proposed methodology. The total RMSE over the whole protocol, assuming a perfect classification of the activities, was $0.86\ kcal/min$. RMSE for lying, LWBM, HWBM, walking, running and biking were $0.24, 0.42, 0.63, 1, 27, 1.06$ and $1.29\,kcal/mim$ respectively. Misclassification lowers performance to $RMSE = 0.87\ kcal/min$. RMSE for the single clusters after classification were $0.24, 0.42, 0.61, 1, 27, 1.07$ and $1.44\ kcal/mim$. These results confirm that the classifier can be used to select activity cluster models.

EE estimation performance

Table 2.5 shows results in terms of RMSE averaged over all of the activities and per cluster. Simple methods used in epidemiological studies (methods ACC and HR) show the lowest performance and will not be further discussed.

Results of the *AS-static* method showed improvements compared to non-activity-specific models, but higher error compared to other activity-specific models, in all of the clusters. Recognizing an activity and assigning a static EE value works well on average but cannot capture the variability in EE within the cluster. Measures of motion intensity (*AS-MI*) seem to outperform HR for low to medium intensity activities (LWBM and HWBM), while activities where whole body motion is not

Table 2.4: Predictors and models used to estimate EE for each activities cluster. BW is body weight, MI is motion intensity, VAR is variance, STD is standard deviation, IQR is inter-quartile range.

Cluster	**Model**
Lying	$RMR \times 1MET$
LWBM	$-0.43 + 0.00068\ RMR + 0.015\ HRaR + 18.23\ MIx + 15.35\ MIy + 2.31\ MIz - 11.83\ VARx - 25.71\ VARy - 5.03\ VARz$
HWBM	$-2.42 + 0.029\ HRaR + 5.23\ MIx + 1.76\ MIy + 1.25\ MIz - 33.10\ VARx - 39.92\ VARy - 9.28\ VARz + 14.96\ STDx + 12.11\ STDy + 1.76\ STDz + 0.04\ BW$
Walking	$-5.31 + 0.068\ HRaR + 6.00\ MIx + 0.087\ BW$
Biking	$-6.78 + 0.0035\ RMR + 0.073\ HRaR + 0.026\ speed$
Running	$-10.62 + 0.027\ HRaR + 5.47\ IQRx + 0.16\ BW$

representative of EE, such as biking (see Fig. 2.6), were better modeled by methods using HR as well (*AS-mixed* and *AS-new*).Walking patters were predicted accurately by methods using ACC only features (*AS-MI*) when differences in EE could be explained by motion patterns alone. The inability of these methods to detect the higher energy cost of carrying weights or walking uphill results in decrease of performance during these activities (see Fig. 2.5).

Overall, combining manually selected ACC and HR features, representative of variations in EE within a cluster, shows significant improvements compared to other methods. Estimates of compendium-based models (*AS-static*) were improved by 31 %. Regression based models that use a measure of motion intensity (*AS-MI*) or automatically selected variables (*AS-mixed*) as predictors, were improved by 18 and 19 % respectively. Figures 2.2 to 2.6 show how combining features specifically selected for a cluster, based on motion patterns involved in the cluster, as well as physiological signals able to capture variations in EE when motion is constant, provides better estimates compared to other activity-specific methods, on almost all of the activities included in our protocol.

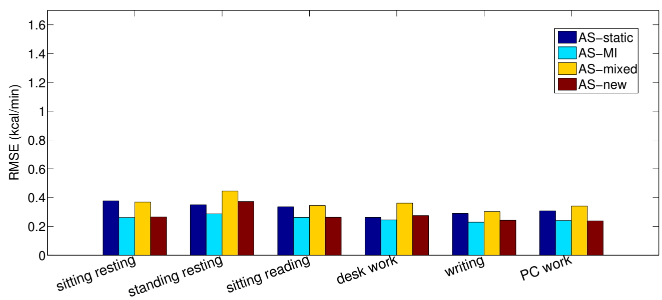

Table 2.5: Overall and per cluster performance (RMSE) of the methods implemented. Results are in kcal/min. AS is Activity Specific, MI is Motion Intensity. Refer to Sec. 2.5 for details on the methods.

Cluster	**ACC**	**HR**	**AS-static**	**AS-MI**	**AS-mixed**	**AS-new**
Lying	0.65	1.21	0.29	0.26	0.24	0.24
LWBM	0.68	1.45	0.66	0.48	0.59	0.42
HWBM	0.75	1.32	1.19	0.80	0.89	0.63
Walking	1.55	1.65	1.66	1.49	1.43	1.27
Running	2.00	2.72	1.54	1.20	1.50	1.06
Biking	4.38	1.67	1.88	1.84	1.52	1.29
Overall	1.51	1.57	1.25	1.05	1.06	0.86

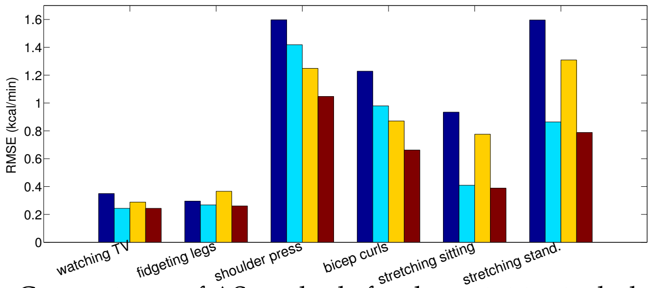

Figure 2.2: Comparisons of AS methods for the activities included in the *LWBM* cluster.

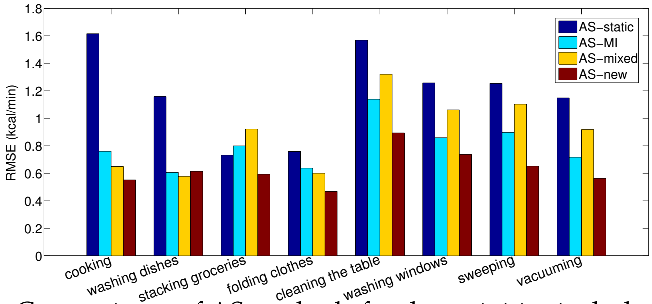

Figure 2.3: Comparisons of AS methods for the activities included in the *HWBM* cluster.

2.7 Conclusions and future work

We introduced a new methodology, which aims at clarifying the relation between type of physical activity and EE. Our approach consists of four steps. First, we separated activities into clusters meaningful for EE estimates. Secondly, we split

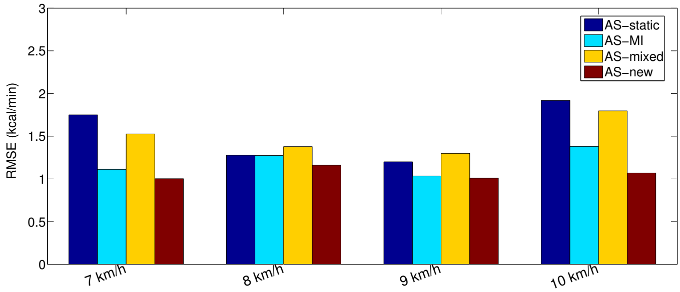

Figure 2.4: Comparisons of AS methods for the activities included in the *running* cluster.

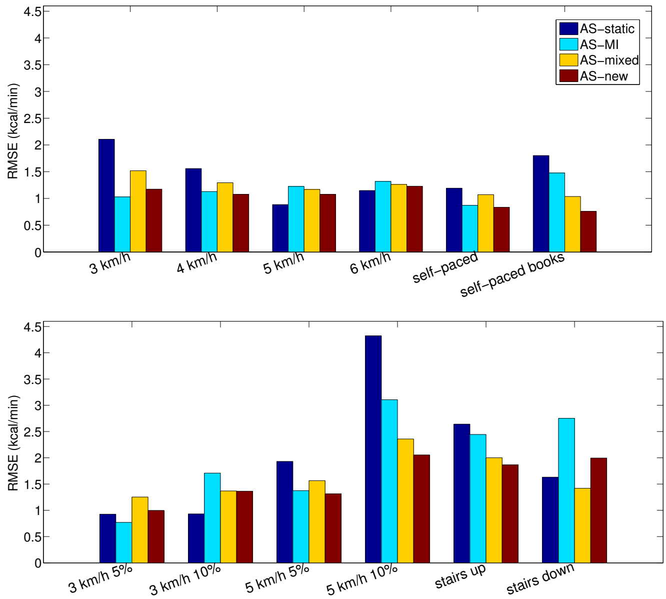

Figure 2.5: Comparisons of AS methods for the activities included in the *walking* cluster.

sedentary and non-sedentary activities, and assigned a static MET value to sedentary activities. The motion patterns of non-sedentary clusters were examined, to select ACC features representative of intra-individual differences in EE within the cluster. When no differences in motion were distinguishable within one cluster, physiological signals were used to discriminate between different levels of EE. Finally, we included anthropometric characteristics to take into account differences in body size. By applying this methodology to the development of a new algorithm for a single monitoring device, we showed improvements in EE estimates, ranging from 18 to 31% compared to state of the art activity-specific methods.

An aspect of interest that was not further investigated during this study is the

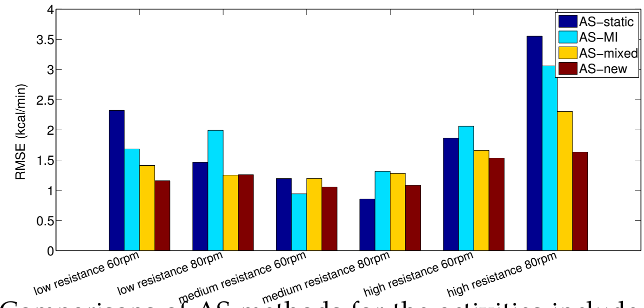

Figure 2.6: Comparisons of AS methods for the activities included in the *biking* cluster.

personalization of EE models that use physiological signals. Physiological signals (e.g. HR) differ greatly at the individual level, and require either individual calibration or normalization. We used the heart rate above rest as the only heart rate feature, to reduce between-subject differences in HR during different activities. We are currently investigating the possibility to include other factors able to explain between-subject differences in HR during different activities (e.g. cardiorespiratory fitness level), in order to further improve the activity-specific models.

3

Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning

M. Altini, J. Penders, R. Vullers, O. Amft
Adapted from: IEEE Journal of Biomedical and Health Informatics, no. 99, p. 1, 2014.
Featured as research highlight on JBHI's homepage.

Abstract

Several methods to estimate Energy Expenditure (EE) using body-worn sensors exist, however quantifications of the differences in estimation error are missing. In this paper, we compare three prevalent EE estimation methods and five body locations to provide a basis for selecting among methods, sensors number and positioning. We considered (a) *counts-based estimation methods,* (b) *activity-specific estimation methods using metabolic equivalents (METs) lookup tables, and* (c) *activity-specific estimation methods using accelerometer features. The latter two estimation methods utilize subsequent activity classification and EE estimation steps. Furthermore, we analyzed accelerometer sensors number and on-body positioning to derive optimal EE estimation results during various daily activities. To evaluate our approach, we implemented a study with* 15 *participants that wore five accelerometer sensors while performing a wide range of sedentary, household, lifestyle, and gym activities at different intensities. Indirect calorimetry was used in parallel to obtain EE reference data. Results show that activity-specific estimation methods using accelerometer features can outperform counts-based methods by* 88% *and activity-specific methods using METs lookup for active clusters by* 23%. *No differences were found between activity-specific methods using METs lookup and using accelerometer features for sedentary clusters. For activity-specific estimation methods using accelerometer fea-*

tures, differences in EE estimation error between the best combinations of each number of sensors (1 to 5), analyzed with repeated measures ANOVA, were not significant. Thus, we conclude that choosing the best performing single sensor does not reduce EE estimation accuracy compared to a five sensors system and can reliably be used. However, EE estimation errors can increase up to 80% if a non-optimal sensor location is chosen.

3.1 Introduction

Energy expenditure (EE) is the most commonly used single metric to quantify PA. Different methods to estimate EE have been developed in the past, from counts-based estimation methods to activity-specific EE equations, developed using one or more accelerometers. Counts-based estimation methods are developed by fitting a single regression line to all the data, independently of the activity performed. On the other hand, in activity-specific estimation methods, the estimation process is split into two steps. First, activities are classified into clusters that group them according to a certain criteria (e.g. EE level [19], motion patterns [29], etc.). Secondly, an activity-specific model is applied to estimate EE. Activity-specific EE models [9, 29, 118] showed higher performance compared to single models [52, 60]. However, little agreement is found in literature regarding number of accelerometers, location on the body, and the role of accelerometer features (e.g. used for activity recognition only (activity-specific models using METs lookup), or for both activity recognition and activity-specific EE models (activity-specific using accelerometer features)) [3, 118]. Even though the use of a single sensor is more practical, recent advances in sensor technology and the ease of integrating small accelerometers into shoes [91], watches or mobile phones, reduced obtrusiveness of wearable sensors, allowing researchers to deploy multi-sensor systems.

Determining the optimal number and on-body positioning of accelerometers to accurately estimate EE requires addressing the following issues, that have not been studied: *1) On activity recognition:* what is the influence of activity type misclassification on EE estimation error when using activity-specific approaches?, *2) On differences in EE within an activity cluster:* which activity-specific approach performs best during different activities? and *3) On EE estimation:* how do activity recognition accuracy and EE estimation error change based on sensors number and positioning?

In this paper we analyze three prevalent EE estimation methods as well as on-body sensors number and positioning to estimate EE. In particular, this paper provides the following contributions:

1. We analyze EE estimation error for three common EE estimation approaches (counts-based, activity-specific using METs lookup and activity-specific using accelerometer features). We show that activity-specific using accelerometer features approaches outperform counts-based approaches and activity-specific using METs lookup approaches for active clusters.

2. We analyze all combinations of five accelerometers on-body positions and evaluate their impact on activity recognition and EE estimation error. We show that a single accelerometer is sufficient to maintain the lowest EE estimation error when suitably placed.

3.2 Related work

Counts-based estimation methods

Counts-based methods were the first EE estimation algorithms developed. They are simple linear regression models using as predictor activity-counts, i.e. a dimensionless unit representative of whole body motion. Counts-based models rely on the relation between motion intensity close to the body's center of mass and EE [52]. However, single regression models are unable to fit all the activities, since the slope and intercept of the regression model change based on the activity performed while data is collected [123]. As a result, even when motion intensity (*activity counts*) is representative of EE, the output can be inaccurate. Additionally, the inability of these systems to recognize high or low body movement (e.g. biking or arm exercises) caused high estimation error for activities not involving whole body motion. In [51] the authors had to remove biking activities from their evaluation, due to the inability of their system to capture EE changes when there is limited motion close to the body's center of mass.

Activity-specific estimation methods

The latest algorithms extended estimation methods based on single models by performing activity recognition over a predefined set of activities - or clusters of activities -, and then applying different methods to predict EE [9, 3, 118, 29, 123], based on the activity detected (activity-specific EE approaches). Other machine learning based methods were developed [60], trying to directly estimate EE from accelerometer features, using for example neural networks [103, 60]. However these approaches suffer from the same limitations of the counts-based estimation methods, being unable to capture the peculiarities of the relation between accelerometer features and EE during different activities [30]. The most common activity-specific approaches are the following:

Activity-specific using METs lookup

One approach is to assign static MET values from the compendium on physical activities [2] to each one of the clusters of activities [29, 3], and use anthropometric features or other static features (e.g. heart rate at rest) to personalize the activity-specific models for different individuals.

Activity-Specific Using Accelerometer Features

Another approach is to apply a regression equation for each activity classified [118, 123], extending counts-based approaches to multiple clusters of activities. The regression models typically use accelerometer features and anthropometric characteristics as independent variables.

Comparisons

Comparisons of estimation methods

[29] showed that activity-specific estimation methods using METs lookup outperform counts-based approaches when a single sensor is used. In [3], the authors extended the static approach of [29], developing a custom MET table, which takes into account the heart rate at rest, to predict EE, showing 15% improvement in performance compared to the best counts-based estimation methods. In both our previous work [9] and in [3], activity-specific estimation methods using METs lookup and accelerometer features were implemented and compared. However, while we proposed a combined approach using METs lookup for sedentary clusters of activities and using accelerometer features for active clusters, the authors of [3] opted for using METs lookup only. The two systems used a different sensor setup. One single sensor on the chest was used in [9], while three sensors placed on the upper arm, thigh and waist were used in [3]. The different activity types, sensors number and positioning might have motivated the different choices made by the authors. Thus, it is unclear what estimation method works best as well as if different estimation methods require different sensors number.

Comparisons of sensors number and positioning

When it comes to sensors number and positioning, comparisons are lacking. Some works investigated the accuracy of sensors placed on different parts of the body to detect a specific set of activities [19, 95, 118, 20, 47]. However, none of these works considered how sensors number and positioning affects EE. Some researchers showed high accuracy in EE estimates adopting one sensor placed on the lower back [29] or chest [9]. Others used two or three accelerometers [118, 3]. Small differences between protocols used to collect data, algorithms evaluation metrics, as well as the inclusion of extra sensors in only some of the systems (e.g. heart rate), limit our understanding of what is the best solution in terms of sensors number and positioning.

3.3 Analysis approach

This section covers the approach we used to analyze the role of different estimation methods, sensors number and positioning for EE estimation.

1. *Estimation methods:* we compared three common methods to estimate EE: (a) counts-based, activity-specific using (b) METs lookup and using (c) accelerometer features (see Fig. 9.1 and Fig. 3.2).

   a) Counts-based estimation methods: These methods consist in a linear regression model. The model can be formalized in vector form as follows: $y = X\beta + \epsilon$. In the context of EE estimation, y is the vector of target EE values, β is the vector of regression coefficients, and X is the vector of input features. The vector X contains p features, features that can be grouped into two categories: accelerometer features (X_{acc}) and anthropometric characteristics (X_{ant}).

   b) Activity-specific estimation methods using METs lookup: They are composed of two parts: activity recognition and activity-specific models. Assuming n clusters of activities $C = \{c_1, \ldots, c_n\}, \forall c_i \in C, \quad \exists \quad y_{act_i} = X_{act_i}\beta_{act_i} + \epsilon$. Each y_{act_i} maps an activity cluster to EE. y_{act_i} is the vector of target EE values for a specific cluster of activities, β is the vector of regression coefficients, and X_{act_i} is the vector of input features. The vector X_{act_i} contains r features, a MET value depending on the activity type, taken from the compendium of physical activities (X_{met_i}), and anthropometric characteristics (X_{ant}), used to personalize models between individuals.

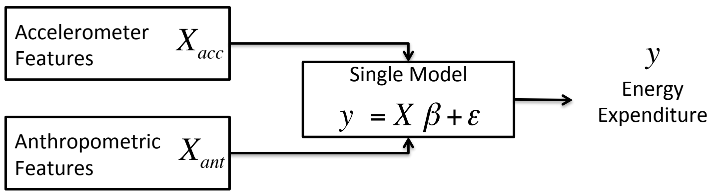

Figure 3.1: Block diagram of counts-based estimation methods. Accelerometer and anthropometric features are used independently of the activity type.

   c) Activity-specific estimation methods using accelerometer features: Similarly to b), we assume n clusters of activities $C = \{c_1, \ldots, c_n\}, \forall c_i \in C, \quad \exists \quad y_{act_i} = X_{act_i}\beta_{act_i} + \epsilon$. Where each y_{act_i} maps an activity cluster to EE. As in b), y_{act_i} is the vector of target EE values for a specific cluster of activities, β is the vector of regression coefficients, and X_{act_i} is the vector of m input features. Features can be grouped into accelerometer features (X_{acc_i}) and anthropometric characteristics (X_{ant}). X_{acc_i} differ from X_{met_i} introduced in b), since they are not constant and change within a cluster.

2. *Sensors number and positioning:* we evaluated all possible combinations of 5 sensors (see Fig. 3.3 and Sec. 3.5 for details). Our analysis is structured as follows:

   a) Activity recognition: $\forall$ sensors number $j \in \{1, \ldots, 5\}$, and $\forall$ combinations k of j sensors, $k = \binom{5}{j}$, we implemented an activity recognition model to classify clusters of activities $c_i \in C = \{c_1, \ldots, c_n\}$. Additionally, activity recognition accuracy was evaluated in the ability to discriminate between *sedentary* and *active* clusters of activities. This analysis was performed to understand to which extent misclassification of the activity class can affect EE estimation accuracy for activity-specific EE models.

b) Differences in EE within an activity cluster: We assumed perfect activity recognition (i.e. $\forall$ instance d, we assume $c_{d_p} = c_{d_a}$ where c_{d_p} is the predicted cluster, while c_{d_a} is the actual cluster). Assuming n clusters of activities $C = \{c_1, \ldots, c_n\}, \forall c_i \in C$, $\forall$ sensors number $j \in \{1, \ldots, 5\}$, and $\forall$ combination k of j sensors, $k = \binom{5}{j}$, we implemented an activity-specific model using accelerometer features; $y_{i,j,k} = X_{i,j,k}\beta_{i,j,k} + \epsilon$. Where $X_{i,j,k}$ is the vector of the input features (as in III.1.c, features include $X_{acc_{i,j,k}}$ and X_{ant}). $X_{acc_{i,j,k}}$ includes features from one of the k combinations of j sensors for the activity i. On the other hand, as shown in section III.1.b, activity-specific estimation methods using METs lookup do not include accelerometer features in the activity-specific models. Thus, once perfect activity recognition is assumed there is no difference in EE estimation due to sensor number and positioning. Assuming n clusters of activities $C = \{c_1, \ldots, c_n\}, \forall c_i \in C$, we implemented one activity-specific regression model using METs lookup per activity, as in Sec. III.1.b; $y_{act_i} = X_{act_i}\beta_{act_i} + \epsilon$. This analysis was performed

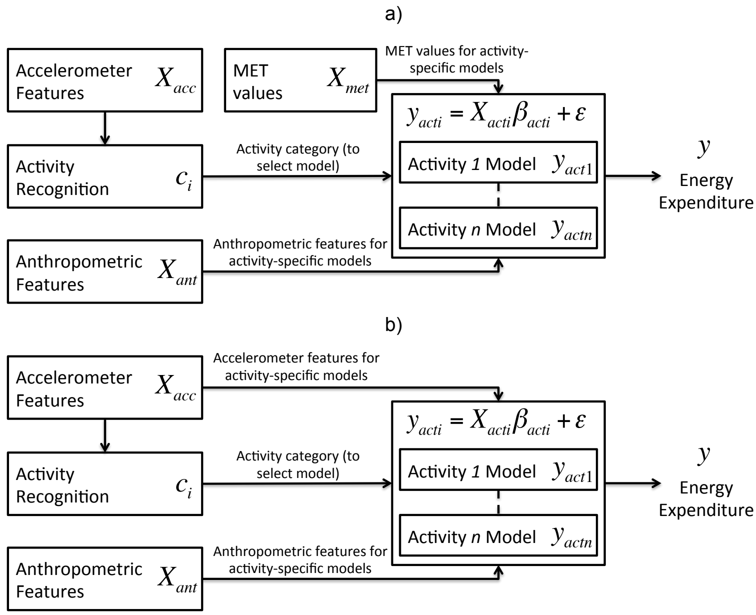

Figure 3.2: Block diagram of the activity-specific estimation methods considered for comparison in this work. $a)$ shows approaches using METs lookup while $b)$ shows approaches using accelerometer features as predictors.

to understand in which activities accelerometer features can improve EE estimation accuracy, compared to activity-specific estimation methods using METs lookup, and if higher sensors number can reduce EE estimation error. *c) EE estimation:* Combining activity-recognition and activity-specific EE models, we analyzed the impact of multiple accelerometers in EE estimation. Misclassification rates were taken into account by applying the wrong activity-specific EE model in the estimation process. As in all activity-specific models, $\forall c_i \in C = \{c_1, \ldots, c_n\}, \quad \exists \quad y_{act_i} = X_{act_i}\beta_{act_i} + \epsilon$. Given an instance d, we can apply n EE models, one $\forall c_i$. if $c_{d_p} \neq c_{d_a}$, the wrong activity-specific EE model will be applied (e.g. $y_{act_p} = X_{act_p}\beta_{act_p} + \epsilon$ instead of $y_{act_a} = X_{act_a}\beta_{act_a} + \epsilon$). This analysis was performed to understand if more sensors improve not only activity-recognition, as known from literature, but also EE estimation accuracy, due to reduced misclassification rates.

Statistics and performance measure

Models were derived using data from all but one participants, and validated on the remaining one (leave-one-participant-out cross validation). Performance of the activity recognition models was evaluated using the average of the percentage of correctly classified instances (i.e. accuracy). Results for EE estimates were reported using Root mean square error (RMSE), where the outcome variable was gross EE expressed in kcal/min. A one-way repeated-measures within-subjects ANOVA with six levels was used to compare EE models. The Tukey test was used to perform pairwise comparisons. Paired t-tests were used to compare RMSE between the best and worst sensor for each number of sensors (1 to 5). Significance was assessed at $\alpha < 0.05$.

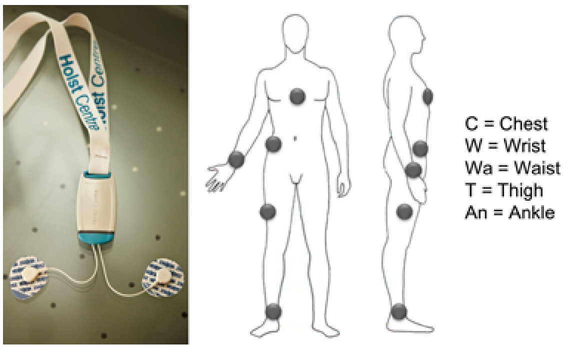

Figure 3.3: ECG Necklace and on-body accelerometer positioning.

3.4 Implementation

Activity type clusters

We grouped all recorded activities into two categories to separate sedentary and active behavior. We included *lying (lying down resting)*, *sitting (sitting resting, desk work, reading, writing, working on a PC, watching TV)* and *standing (standing resting, standing cooking)* postures in our sedentary clusters. Active clusters were four, one representative of household activities, namely the Òhigh whole body motion (HWBM)Ó cluster (*stacking groceries, washing dishes, folding clothes, cleaning and scrubbing, washing windows, sweeping, vacuuming*) and three representative of locomotion and active transportation, such as *walking (self-paced, self-paced carrying books, treadmill flat:* $3, 4, 5, 6\ km/h$*, incline:* $3, 5km/h, 5, 10\%$*)*, *biking (cycle ergometer, low, medium and high resistance level at* 60 *and* $80\ rpm$*)* and *running* ($7, 8, 9, 10\ km/h$ *on a treadmill)*.

Features extraction and selection

Features extracted from the sensors' raw data were used to derive activity recognition and EE models. Accelerometer data from the three axes of all five sensors were segmented in 4 seconds windows, band-pass (BP) filtered between 0.1 and 10 Hz, to isolate the dynamic component caused by body motion, and low-pass (LP) filtered at 1 Hz, to isolate the static component, due to gravity. Feature selection for activity recognition was based on correlation, due to the hypothesis that a good feature set includes features correlated with the class, but uncorrelated to each other. The final feature set included: *mean of the absolute BP signal*, *inter-quartile range*, *mean distance between axes*, *median*, *variance*, *standard deviation*, *zero crossing rate*, *main frequency peak*, *low and high frequency band signal power*. Feature selection for EE was based on how much variation in EE each feature could explain within one cluster. The process was automated using linear forward selection. Features to be selected depended on the combination of sensors considered for a model. Additionally, anthropometrics features (*body weight* and *resting metabolic rate (RMR)*, estimated with the Harris-Benedict formula [59]) were added depending on the cluster, following the methodology for activity-specific EE models presented in [9].

Activity recognition

We adopted a constant set of parameters for sliding window and classifier type of the activity recognition. We selected a time window of 4 seconds, which is short enough to detect short breaks in sedentary time, and long enough to capture the repetitive patterns of some activities (e.g. walking). Given the positive results in past research on activity recognition, we selected Support Vector Machines (SVMs) as classifiers. For the SVMs, we used a polynomial kernel with degree 5 ($\lambda = 10$, $C = 1$), fixing these parameters for all models.

Energy expenditure

Counts-based Methods

We implemented single regression models using data from all activities and motion intensity (i.e. *mean of the absolute BP signal summed over the three axis*) as the only accelerometer feature, together with anthropometric characteristics (*body weight* and *RMR*), as typically done in epidemiological studies (see Fig. 9.1).

Activity-specific estimation methods using METs lookup

Activity-specific estimation methods using METs lookup relied on the activity recognition system of Sec. 3.4. METs values were used together with anthropometric features (*body weight* and *RMR*), for the activity-specific linear regression models (see Fig. 3.2.a). METs values were chosen based on compendium values for the activities included in each cluster, resulting in 1 for *lying*, 1.3 for *sitting* and *standing*, 3.5 for *HWBM*, 3 for *walking*, 6.7 for *biking* and 11 for *running*.

Activity-specific estimation methods using accelerometer features

Within one activity cluster, EE can be estimated using other features, representative of EE changes within the activity cluster [118, 123, 9]. Depending on sensors selected, we created different EE activity-specific linear models, using the selected set of features for those sensors (see Fig. 3.2.b).

3.5 Evaluation study

Participants

Participants were 15 (11 male, 4 female) healthy individuals, mean age $29.8 \pm 5.2\ years$, mean weight $71.8 \pm 15.9\ kg$, mean height $1.75 \pm 0.10\ cm$, mean BMI $23.2 \pm 3.0\ kg/m^2$. Imec's IRB approved the study. Each participant signed an informed consent form.

Instruments

Body area network

The sensor platform used was the ECG Necklace. Five ECG Necklaces were synchronized in a wireless network [7] (see Fig. 3.3). One ECG necklace was placed on the chest (C) and configured to acquire one lead ECG data at 256 Hz, and accelerometer data at 64 Hz (ADXL330). Sampling frequency was chosen as 64 Hz since it is considered to be much higher than typical human motion. The other four ECG Necklaces were configured to acquire only accelerometer data at 64 Hz and placed on the dominant ankle (An), dominant thigh (T), dominant wrist (W) and waist (Wa) - at the right hip. All sensors were attached to the body using elastic bands. ECG data was not used for this study. Activity type was annotated manually by experimenter.

Indirect calorimeter

Breath-by-breath data were collected using the Cosmed $K4b^2$ indirect calorimeter. The Cosmed $K4b^2$ weights $1.5\ kg$ and showed to be a reliable measure of EE [85]. The system was manually calibrated before each experiment according to the manufacturer instructions.

Experimental design

Participants were invited for recordings on two separate days. They reported to the lab at 8:00 am, after refraining from drinking (except for water), eating and smoking in the two hours before the experiment. The protocol included a wide range of sedentary, lifestyle and sport activities. Each activity was carried out for a period from 4 to 12 minutes, except for running (1 to 4 minutes). The first minute of each recording was removed to discard non-steady-state data.

3.6 Results

Given the high number of models implemented, we report only results for the best combinations of 1 to 5 sensors (Fig. 6.3.a-b, Fig. 8.2.a-b and Fig. 9.2.a), as well as information on exactly which sensors provide these optimal performance, together with the worst performance obtained with the same number of sensors, for comparison.

Estimation methods

Fig. 6.3 shows the effect of different feature sets on EE estimation performance for activity-specific EE models, assuming perfect activity recognition.Only one activity-specific model using METs lookup is needed for comparison, since these approaches don't use accelerometer features. The RMSE obtained for activity-specific estimation methods using METs lookup was 1 kcal/min, while for activity-specific estimation methods using accelerometer features it ranged between 0.84 and 0.86 kcal/min (18% error reduction, $p < 0.05$, Fig. 6.3.a). 23% error reduction was shown for active clusters using accelerometer features (Fig. 6.3.b). Fig. 8.2 shows performance of the EE estimation models in combination with activity recognition, as well as counts-based estimation methods. For clarity, results for the activity-specific estimation methods using METs lookup were omitted in Fig. 8.2. Activity-specific estimation methods using METs lookup rely on the same activity recognition algorithms used by the activity-specific method using accelerometer features, thus the METs-based method would still perform sub-optimally. The RMSE for activity-specific estimation methods using accelerometer features ranged from 0.85 to 0.89 kcal/min. RMSE for counts-based estimation methods was between 1.6 and 2.6 kcal/min depending on sensor position (88% error increase for the best-performing sensor, C, $p < 0.05$). The error obtained using the counts-based estimation was significantly higher compared to activity-specific models even when counts were considered separately for sedentary and active clusters (Fig. 9.2.b).

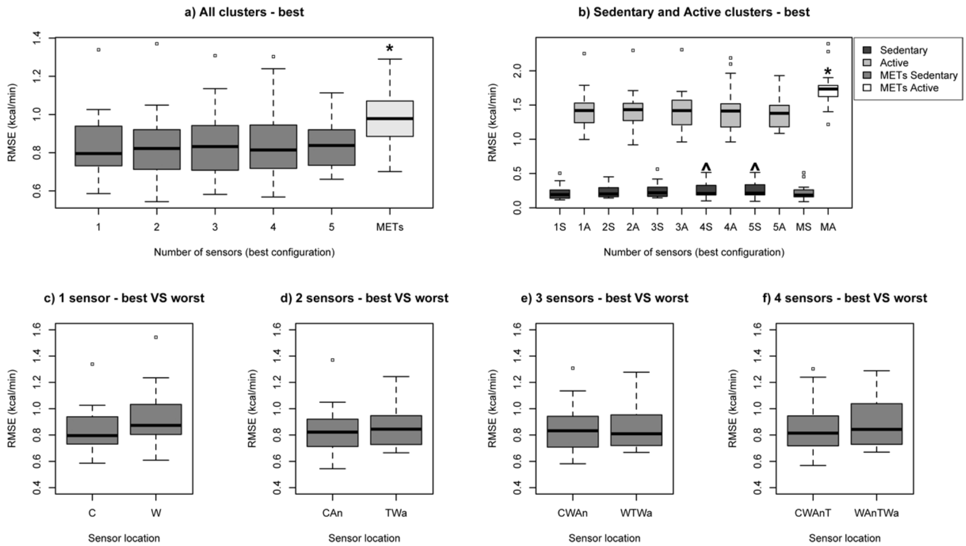

Figure 3.4: EE estimation RMSE for sedentary and active clusters when perfect activity recognition is assumed in activity-specific estimation methods. Boxplots 1 to 5 in plot a) as well as $1S-1A$ to $5S-5A$ on plot b) concern activity-specific estimation methods using accelerometer features, while an activity-specific model using METs lookup is shown as *METs* on plot a) and $MS-MA$ on plot b). $*$ indicates significant differences between the annotated model (activity-specific model using METs lookup) and all of the other models, i.e. the ones using accelerometer features ($p < 0.05$). Λ indicates significant differences between models the annotated models, i.e. 4S (4 sensors, sedentary clusters) and 5S (5 sensors, sedentary clusters) and model sedentary model when only one sensor is used, i.e. 1S ($p < 0.05$). RMSE for the best and worst activity-specific model using accelerometer features for each number of sensors is shown on the bottom row. C is Chest, T is Thigh, An is Ankle, W is Wrist and Wa is Waist.

Sensors number and positioning

Sensors number and positioning is evaluated according to the three criteria of Sec. 9.3: (*1) activity recognition*, *2) differences in EE within an activity cluster* and *3) EE estimation*).

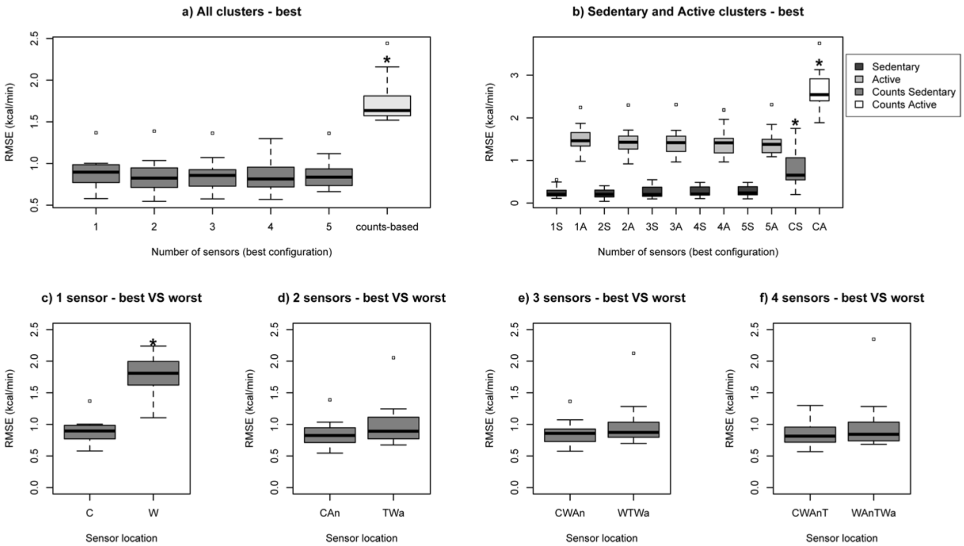

Figure 3.5: EE estimation RMSE for sedentary and active clusters in activity-specific estimation methods using accelerometer features, after activity classification. Activity-specific estimation methods using METs lookup are not shown due to sub-performing results. Comparison with a counts-based model is shown in $a)$ as *counts-based* and $b)$ as *CS* and *CA*. $*$ indicates significant differences between the annotated counts-based model and all of the other models, i.e. activity specific models using accelerometer features ($p < 0.05$). RMSE for the best and worst activity-specific models using accelerometer features for each number of sensors is shown on the bottom row. C is Chest, T is Thigh, An is Ankle, W is Wrist and Wa is Waist.

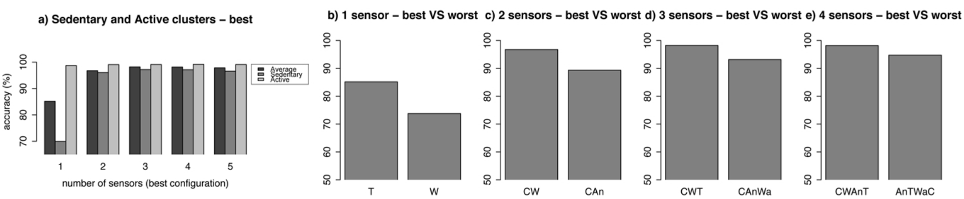

Figure 3.6: Activity recognition accuracy for sedentary (average accuracy of *lying, sitting* and *standing*) and active (average accuracy of *HWBM, walking, biking* and *running*) clusters, and their average. Classification accuracy for different sensors number and positioning are shown on the right. C is Chest, T is Thigh, An is Ankle, W is Wrist and Wa is Waist.

Activity Recognition

Fig. 9.2.a shows the performance of the activity recognition models. Additionally, the impact of sensor location (best VS worst for each number of sensors) is shown

(Fig. 9.2.b-e). Accuracy varied between 85% for 1 sensor - and 98% for 3 or more sensors (Fig. 6.a). Accuracy for active clusters was always above 98%, with differences of only 1% between the best single sensor system and a 5 sensors body area network (Fig. 9.2.a). Sedentary clusters accuracy ranged between 69.9 and 97%. Sensor location affected the accuracy by 12% for a single sensor, while the decrease in performance was reduced to 7%, 5% and 4% for two, three and four sensors respectively (see Fig 9.2.b-e).

Differences in EE within an activity cluster

Fig. 6.3 shows the effect of different feature sets on EE estimation performance for activity-specific estimation methods using accelerometer features, assuming perfect activity recognition. No significant differences were found when different locations on the body were considered to extract activity-specific features. However, differences are found when analyzing separately sedentary and active clusters, showing higher errors in sedentary clusters using accelerometer features from four or five sensors (see Fig. 6.3.b).

EE estimation

Fig. 8.2 shows performance of the EE estimation models in combination with the activity recognition. In this analysis, differences in performance are due to a) higher misclassification rates of models based on a smaller number of sensors and b) different feature sets used for activity-specific estimation methods using accelerometer features, depending on the sensors that are part of the system. Sensor location analysis shows the Chest sensor as the best single sensor for EE estimation, while the Wrist sensor seems to perform worse than any other combination (Fig. 8.2.c-f).

3.7 Discussion

To the best of our knowledge, this is the first time that state of the art activity-specific EE estimation methods are evaluated to determine benefits of using multiple accelerometers for EE estimation. For activity-specific estimation methods, evaluating the benefit of multiple sensors is important, since additional accelerometers can contribute differently. Firstly, additional sensors can improve the accuracy of the activity recognition model. Thus, reducing EE estimation error due to the selection of the wrong activity-specific EE model. Secondly, features from more than one sensor could better explain the EE variance within one cluster of activities.

Estimation methods

Our estimation results show that activity-specific estimation methods using accelerometer features outperform counts-based estimation methods by 88% and activity-specific estimation methods using METs lookup by 18%. Counts-based estimation methods were outperformed by the activity-specific estimation, regardless of sensor location, with RMSE between 1.6 kcal/min at the chest to 2.6 kcal/min at the wrist. The results reflect a similar behavior to what was observed for activity-specific models, where wrist-based models were poorly performing due to weak relation between movement and EE. The inability of counts-based estimation methods to fit all activities is further reflected by the estimation error when considering sedentary and active clusters separately. Activity-specific estimation methods using accelerometer features provide no advantage compared to activity-specific estimation methods using METs lookup for sedentary clusters, but only for active clusters (23% error reduction). This is due to the fact that active clusters can be performed at varying intensities (e.g. *walking* at different speeds), and assigning static METs values prevents the model from capturing these differences in intensity within one cluster of activities. However, sedentary clusters of activities cannot be performed at varying intensities (e.g. *sitting* or *lying down*), making it possible to estimate EE accurately using METs lookup approaches. We assume model development did not lead to overfitting given the similar level of error variability between simple and complex methods. We expect that overfitting was avoided as the data from one participant was eight used for training or evaluation.

Sensors number and positioning

Our results on the sensors number and positioning point out three main findings: *1) On activity recognition*: if properly chosen, two sensors are sufficient to provide accurate physical activity type assessment (see Fig. 9.2). *2) On differences in EE within an activity cluster:* Adding features from more than one sensor in the activity-specific models using accelerometer features does not improve the accuracy of the EE estimate (see Fig. 6.3). *3) On EE estimation:* Applying a wrong EE model due to misclassification of the activity type has a small (non-statistically significant) impact on the EE estimate accuracy Ð provided that an optimal sensor positioning is chosen (e.g. the Chest sensor, see Fig. 8.2). Thus, choosing the best performing single sensor does not reduce performance for EE estimation compared to a five sensors system.

Activity recognition

Our results on the sensor number for activity recognition confirm previous works that considered multiple accelerometers [20, 95, 118, 47]. Adding more sensors improves accuracy, until a plateau is reached, when two or more sensors are used. In our case 97/98% accuracy using Chest and Wrist or Chest and Thigh sensors. It is of interest for our analysis, how activity recognition influences EE estimates as discussed below.

Differences in EE within an activity cluster

Our second finding concerns the accelerometer features needed to explain differences in EE within one cluster. To determine such features, we developed EE models assuming perfect activity recognition (see Fig. 6.3). We showed that accelerometer features from one sensor are sufficient to explain differences in EE within one cluster of activities. This finding can be explained by the fact that within one cluster of activities (for example *walking*) the variation in EE is explained mainly by the level of motion intensity of the *whole body*. Other features, such as *motion intensity of the wrist sensor*, can lead to errors, since high level of motion (e.g. while writing), do not correspond to high EE. This reasoning might explain why in Fig. 6.3 the error is shown to increase when features from 4 or 5 sensors are used for sedentary clusters (Fig. 6.3.b).

Even though adding features from more sensors does not reduce EE estimate error, accelerometer features from at least one sensor should be used for active clusters (23% error reduction compared approaches using METs lookup). In a recent review on activity-specific EE estimation [122], the controversy between applying static values (i.e. MET values) and the need of including accelerometer features in linear models had been raised. Past research showed inconsistency in the approach used for activity-specific models even after implementing and comparing estimation methods using METs lookup or accelerometer features [3, 9]. With this analysis we show that accelerometer features are relevant only for active clusters, and most importantly this is true regardless of the number of sensors used (see Fig. 6.3.b). Our findings are consistent with our previous work using one sensor [9], indicating that the best approach to obtain high accuracy and limit model complexity, is to use a combined approach. Activity-specific models using METs lookup can be used for sedentary activities, where static METs values and anthropometric features are sufficient to accurately estimate EE.

EE estimation

Provided that the best performing sensor is chosen, no significant error reduction was found when more than one sensor was used for EE estimation. This is due to the fact that errors are mainly due to misclassification of posture (one single sensor is unable to recognize all of the three postures in the sedentary cluster), resulting in applying a very similar activity-specific EE model. Thus, the EE estimation RMSE for a single sensor placed on the Chest is similar when compared to a 5 sensors system (no statistically significant difference), even if activity classification accuracy is decreased by up to 13% on average, and 28% for sedentary clusters. This is an important finding since past work showed good accuracy using one single accelerometer and activity-specific approaches [29, 9], but no previous work could compare performance of EE estimation methods when different sensors number and positioning were used, preventing us from understanding if systems relying on multiple sensors for activity recognition [3, 118] could still provide better results.

Limitations

Performance for activities that were not part of the dataset should be assessed outside of the lab. However, there is currently no reference system able to measure breath-by-breath EE in unconstrained settings. Only by using indirect calorimetry and supervised settings we can record data which allows us to analyze how multiple sensors affect the EE estimate process in both activity recognition and intra-individual differences within one activity. Another limitation was to limit the number of MET values used for our analysis to the ones associated to the activity clusters, while more fine grained values could be used for certain activities (e.g. walking at different speeds). However we believe that using individual MET values for activities may not generalize, since some activities (e.g. related to household) show different EE but cannot be accurately sub-divided when using a limited number of sensors. Hence some activity clusters would still require a single MET value to be used, while actual EE varies widely.Finally, due to the size and attachment modality of the ECG Necklace we were limited to analyzing positions where the ECG Necklace could be practically attached. While we explored different on-body sensor locations, we had to exclude others that showed promising results in previous research (e.g. the ear) [19, 31], due to practical limitations with our system.

3.8 Conclusions

We suggest using one single sensor close to the body's center of mass (chest or waist), together with a combined activity-specific estimation method, for accurate and unobtrusive EE estimation. The combined estimation method should be composed of activity-specific models using METs lookup for the sedentary activity clusters, and activity-specific models using accelerometer features for the physically active clusters. This approach showed to be both practically feasible, since it limits the number of sensors to one, and accurate in terms of EE estimation accuracy.

Part II: Physiological data normalization

4

Personalizing energy expenditure estimation using a cardiorespiratory fitness predicate

M. Altini, J. Penders, O. Amft

Adapted from: Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on. IEEE, 2013, pp. 65-72.

Abstract

Accurate Energy Expenditure (EE) estimation is key in understanding how behavior and daily physical activity (PA) patterns affect health, especially in today's sedentary society. Wearable accelerometers (ACC) and heart rate (HR) sensors have been widely used to monitor physical activity and estimate EE. However, current EE estimation algorithms have not taken into account a person's cardiorespiratory fitness (CRF), even though CRF is the main cause of inter-individual variation in HR during exercise. In this paper we propose a new algorithm, which is able to significantly reduce EE estimate error and inter-individual variability, by automatically modeling CRF, without requiring users to perform specific fitness tests. Results show a decrease in Root Mean Square Error (RMSE) between 28 and 33% for walking, running and biking activities, compared to state of the art activity-specific EE algorithms combining ACC and HR.

4.1 Introduction

New technologies, seamlessly integrated in everyone's life, able to monitor objectively and non-invasively our behavior, can provide unprecedented insights on aspects of behavior related to physical activity and health status. Among the technologies used to objectively monitor PA, accelerometers (ACC) and heart rate (HR) monitors are the most widespread [34, 35, 43, 9, 3, 107, 29, 118]. For ACC, the rationale behind their adoption is the linear relation between motion close to the body's center of mass, and energy expenditure (EE). On the other hand, HR shows a strong correlation with EE, due to the relation between oxygen consumption, HR and EE. Main limitations of these technologies are the inability of single accelerometers close to the body's center of mass to detect low and upper body motion, and the low accuracy of HR monitors during sedentary behavior, as well as the need for individual calibration. Some of these issues have been tackled by developing activity-specific EE algorithms [9, 3, 107, 29, 118].

By adopting an activity-specific approach, some HR limitations can be easily overcome. Issues due to the weak relation between HR and EE during sedentary time, where HR can be affected by artifacts due to emotions and stress, can be avoided by including HR only in some activity-specific equations (e.g. when moderate to vigorous PA is performed). The need for individual calibration of HR-based algorithms is motivated by the substantial inter-individual differences in the relation between HR and EE. During moderate to vigorous PA, differences in HR between individuals performing the same activity are mainly due to cardiorespiratory fitness (CRF). CRF, is not only the main cause of inter-individual variability, but also inversely related with several health outcomes, such as cardiovascular disease and coronary artery disease, being one of the most important health markers [81]. Combined with activity-specific algorithms, information on CRF could provide more accurate EE estimation. Nevertheless, algorithms in the past tackled CRF-related variance only by means of individual calibration [35], and no algorithm includes information on CRF in the EE estimation equations. For many practical applications individual calibration is not feasible since it would require every user to perform a calibration test.

In this paper, we present a new activity-specific EE algorithm that incorporates CRF-related variance by normalizing HR. The HR normalization is performed by estimating walking speed and activities, and integrating anthropometric information. In particular, the following contributions are made:

1. We detail the HR normalization procedure that, based on activities carried out during daily life (rest, walking at different speeds), can automatically estimate CRF-related variance. Thus, our approach does not require users to do specific fitness tests to estimate CRF.

2. We compare EE estimation performance of a standard current state of the art activity-specific algorithm with our personalizing version considering individual CRF. For this purpose, we used a dataset including 44 activities recorded with 29 subjects.

This paper is structured as follows. Related work and the relation between CRF, HR and EE are discussed in Sections 8.2 and 4.3. Section 4.4 introduces our approach to CRF estimation and HR normalization. The implementation of our approach is described in Section 6.4, while the measurement setup and data collection process can be found in Section 4.6. Results and conclusions are presented in Sections 4.7 and 4.8.

4.2 Related work

4.2.1 Epidemiological research

Accelerometer and HR monitors are the most commonly used single sensor devices in epidemiological studies [34, 35, 43, 21]. ACC use *activity counts*, a unitless measure representative of whole body motion, as independent variable in the linear regression model developed to predict EE [34]. Shortcomings of single regression models are; *a)* the accuracy of the monitor is highly dependent on the activities used to develop the linear model, *b)* a single linear model does not fit all the activities, since the slope and intercept of the regression model change based on the activity performed while data is collected. As a result, even when activity counts are representative of EE, the output can be misleading.

HR monitors suffer from different problems. First, HR monitors are typically inaccurate during sedentary behavior, given the fact that HR is also affected by non-activity related factors, such as stress and emotions [43]. Artifacts at rest were tackled by means of the so called *HR-flex point*, a point above which EE is estimated using an activity-model, while below which EE is estimated using a rest value or a sedentary-model [43]. Basically, a first version of today's activity-specific algorithms [9, 3, 107, 29, 118]. Secondly, HR monitors need individual calibration to perform accurately [34]. The high correlation between HR and EE within one individual, which motivated researchers in using HR monitors to estimate EE since the 80s, is indeed peculiar of a specific individual, and changes substantially between subjects. Even the HR-flex point, is often determined specifically for one individual, by for example averaging the HR at rest and the HR while walking at a certain speed.

4.2.2 Activity-specific EE estimation

The latest monitors extended approaches based on simple linear regression models performing activity recognition over a predefined set of activities, and then applying different methods to predict EE [9, 3, 107, 29, 118], based on the activity. The principle behind activity recognition as a first step in EE estimation is that the slope and intercept of the regression models change based on the activity performed. One approach [118] is to apply a different regression equation for each activity classified. The regression models typically use ACC features and anthropometric characteristics as independent variables. Another approach is to assign static values (e.g. Metabolic Equivalents (METs) from the compendium on physical activities) to each one of the clusters of activities. Assigning static values showed limitations during moderate to vigorous activities in a recent comparison between activity-specific models, since static values cannot capture intra-individual differences in EE [9]. Intra-individual differences in EE for an activity are caused by the fact that moderate to vigorous activities can be carried out at different intensities (e.g. walking at different speeds), resulting in different levels of EE. Activity-specific linear regression models require ACC and HR features to capture these differences [9].

Some authors included HR features as well in the activity-specific linear models. In [3], a multi-sensor system composed of three accelerometers was developed. The authors extended the static approach of [29], developing a custom MET table, which takes into account anthropometric variables, as well as the HR at rest, to predict EE. In [107] HR and ACC were combined as well. The system consisted of three sensors, two accelerometers and a HR belt, and could classify seven types of activities. Inter-individual differences in HR were not taken into account.

Activity-specific multiple linear regression models combining ACC and HR features showed consistent improvements in EE estimation accuracy compared to algorithms using static or ACC-only features [9]. However, inter-individual differences in HR due to CRF are not tackled by any activity-specific algorithm. One approach used to reduce inter-individual differences in HR during daily life, was proposed in [35]. The authors use the Heart Rate above Rest (HRaR), instead of the HR, as a predictor for their linear models. Using the HRaR does bring each subject to the same baseline, but it introduces a simple offset, which is unable to capture how HR evolves during physical exercise, as a result of differences in CRF.

4.2.3 CRF estimation

Even though the effects of CRF on HR are widely recognized, no algorithm up to date includes or models CRF to estimate EE. On the other hand, different groups proposed methods and algorithms to measure and estimate CRF alone [62, 55]. CRF is typically measured by means of a maximal oxygen uptake test. Maximal oxygen uptake (VO_2 max) is widely accepted as the single best measure of cardiovascular fitness and maximal aerobic power. Tests measuring VO_2 max can be dangerous in individuals who are not considered normal healthy subjects, as any problems with the respiratory and cardiovascular systems will be greatly exacerbated. Thus, many protocols for estimating VO_2 max have been developed for those for whom a traditional VO_2 max test would be too risky [62, 55]. Sub-maximal VO_2 max tests generally are similar to a VO_2 max test, but do not reach the maximum of the respiratory and cardiovascular systems. On the other hand, non-exercise VO_2 max estimation uses information about the person's anthropometric characteristics, activity level (derived with questionnaires), and HR at rest features to estimate CRF. Often, the predicted maximal HR is used as well. The shortcoming of this approach is that maximal HR is typically predicted using age only, and HR at rest is weakly related to CRF. Higher accuracy was shown by sub-maximal tests involving actual exercise, for example biking or running at sub-maximal rates.

Even though sub-maximal tests are less dangerous and showed good accuracy in past research [62], they are still affected by some limitations; *a)* a specific test is required to determine CRF, *b)* the specific test should be re-performed every time CRF needs to be assessed, *c)* in the context of EE estimation, it is not clear how to include information about CRF.

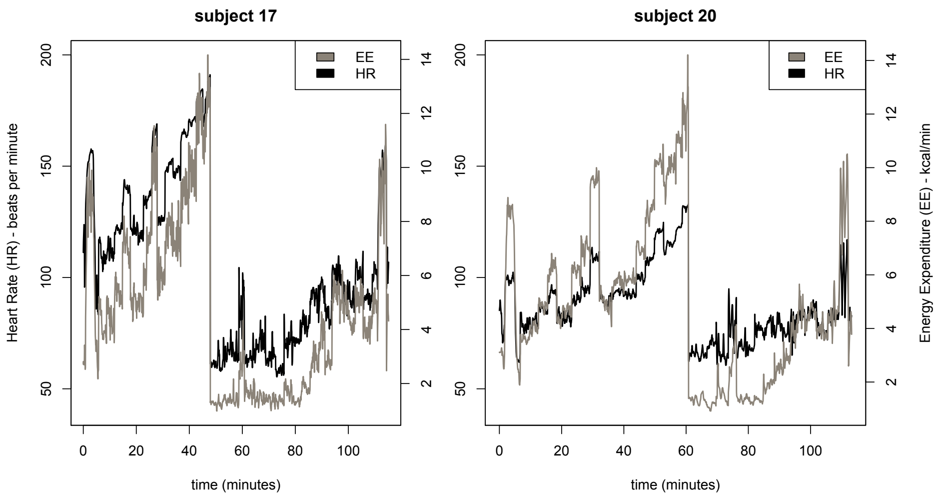

Figure 4.1: Relation between EE and HR in different subjects during a sequence of different PAs. Individual correlations between EE and HR are 0.97 and 0.96 respectively. The absolute EE levels are similar due to similar body weight. HR differs significantly between the subjects during moderate to vigorous activities.

4.3 Fitness and heart responses

This section covers more in detail the relation between CRF, HR and EE, which motivates our approach to personalized EE estimation. The main cause of differences in the HR-EE relation during activity is CRF. An individual with higher CRF (i.e. more fit), will have a lower HR during exercise, compared to an individual with low CRF. Fig. 9.1 shows the relation between HR and EE for two subjects during a series of intense and sedentary activities. Individual correlations between HR and EE are above 0.96 for both of them. The figure shows clearly that for two subjects with similar body weight (subject 17, body weight: 71.2 kg, and subject 20, 72 kg), EE is almost the same, while HR is very different, due to higher fitness level of subject 20 (subject 17 is inactive while subject 20 is a trained runner). Since EE is derived from HR, typically by means of a linear model, estimating EE from HR during exercise results into substantially high over and under-estimations. By individually calibrating the system, the relation between HR and EE becomes peculiar for an individual, since it is derived specifically for him/her, and not using data collected on a different sample of the population. Unfortunately, individual calibration is not practically feasible since it requires each new user to perform lab tests in supervised settings, using expensive devices such as an indirect calorimeter. Thus, alternative methods to tackle the problem are needed to objectively and accurately estimate EE at the individual level, and not only as group averages.

4.4 Approach

This section covers our approach to CRF estimation and its integration into EE multiple linear regression models, necessary to reduce EE estimation error due to inter-individual differences in the relation between HR and EE. To this aim, we developed the concept of automatic *Heart Rate Feature Normalization*. We propose the following steps to estimate EE using HR normalized by level of CRF:

1. Build a model to derive a normalization factor automatically during daily life, without requiring specific tests.
2. Use the normalization factor to normalize HR.
3. Use the normalized HR as predictor in activity-specific EE estimation equations.

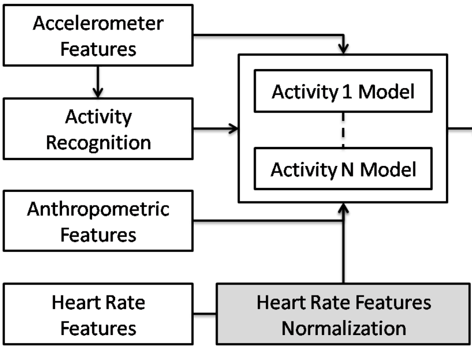

Figure 4.2: Architecture of an activity-specific EE estimation system. ACC features are used to recognize an activity, and to estimate EE for each model, together with anthropometric characteristics and the normalized HR. The *Heart Rate Features Normalization* block, shown in gray, is included in our model to remove confounding effects in the relation between HR and EE, due to CRF.

When normalizing HR, we are not only interested in the maximal HR an individual can reach, but in the HR the different individuals would reach when performing the same activity, at the same workload. We hypothesize that this HR at a constant workload is representative of CRF, and can be used to normalize HR. We selected running at $10km/h$ as the constant workload. Thus, our HR normalization factor is the HR while running at $10km/h$. Running at maximum 10 (for females) to 12 (for males) km/h, together with anthropometric characteristics, could explain 88% of the variance in VO_2 max in past research on VO_2 max estimation [62], showing that it is a normalization factor that well represents CRF. We implemented a system able to derive the CRF-related normalization factor automatically during daily life, without requiring specific tests (a). In this way the algorithm is able to self-adapt and learn from its user, without requiring any

Table 4.1: Distribution of the activities into the six clusters used for activity recognition.

Cluster name	Original activities
Lying	Lying down resting
Sedentary	Sitting resting, sitting stretching, standing stretching, desk work, reading, writing, working on a PC, watching TV, sitting fidgeting legs, standing still
Dynamic	Stacking groceries, washing dishes, cooking, folding clothes, cleaning and scrubbing, washing windows, sweeping, vacuuming
Walking	Self-paced, self-paced carrying books, stairs up and down, treadmill (flat: $3, 4, 5, 6\ km/h$, $4\ km/h$ carrying weights, incline: $3, 5km/h, 5, 10\%$))
Biking	Cycle ergometer, low, medium and high resistance level at 60 and $80\ rpm$
Running	$7, 8, 9, 10\ km/h$ on a treadmill

individual calibration to estimate CRF or EE. Once a normalization factor for an individual has been automatically determined, it is used to normalize HR (b). The resulting normalized HR is free of confounding effects due to CRF. As a last step, the normalized HR is used instead of the HR as a predictor for the activity-specific EE estimation equations (c), as shown in Fig. 9.2. The new predictor better represents the relation between HR and EE, since HR is not affected by CRF.

4.5 Methods

We considered ACC and HR data to implement all components of our approach, including activity recognition, HR normalization factor estimation, and EE estimation. This section details the components further. More details on participants, sensor device and experimental protocol can be found in Section 4.6.

4.5.1 Activity recognition

We implemented an activity recognition algorithm to classify the following clusters of activities (see Table 4.1): *lying, sedentary, dynamic, walking, running* and *biking*. We selected Support Vector Machines (SVMs) as classifier, and the following features: *mean absolute value of the band-passed signa, variance, standard deviation, main frequency peak, amplitude of the main frequency peak* and *high frequency band signal power*. See Sections 6.6.1.2 and 4.5.6 for details on the feature extraction and selection processes. For the SVM, a polynomial kernel with degree 5 was used ($\lambda = 10$, $C = 1$). Activity recognition is used for EE estimation (all six clusters), and as part of the automatic HR normalization system (*lying* and *walking* only).

Table 4.2: Walking Speed Model ($R^2 0.94$).

Variable	**Coefficient**
Intercept	-1.28
Height	0.015
MI	11.37
Var	-2.41
IQRX	1.79
IQRY	-2.96
HPowX	-0.00079
HPowZ	-0.00084
FFTpeakXf	-0.088

4.5.2 Automatic HR normalization system

We extended the architecture of activity-specific EE estimation algorithms (Fig. 9.2), by including the extra *Heart Rate Features Normalization* block. The block is detailed in Fig. 6.3, where all the components necessary to derive the normalization factor automatically, are listed. In order to provide automated and non-invasive CRF estimation, we estimate the normalization factor using activities of daily living only, and their associated HR. More specifically, the *Heart Rate Features Normalization* block uses the HR while resting and walking at different speeds as predictors for the HR normalization factor, together with anthropometric characteristics. Thus, in addition to the activity recognition algorithm, the automatic HR normalization system includes two more components; 1) a walking speed estimator, 2) and a normalization factor estimator. The next sections covers the components in detail.

4.5.2.1 Walking speed estimator

The walking speed estimator is a multiple linear regression model (see Table 4.2) which predicts walking speed using as features the individual's *height* and the following ACC features: *main frequency peak on the X axis (FFTpeakXf), mean absolute value of the band-passed signal (or Motion Intensity, MI), sum of the variance on the three axis (Var), inter-quartile range on the X and Y axis (IQRX and IQRY)* and *high frequency band signal power on the X and Z axis (HPowX and HPowZ).*

4.5.2.2 Heart rate normalization factor estimator

A multiple linear regression model (see Table 4.3) is built to predict the normalization factor (i.e. an individual's HR while running at 10 km/h) using activities of daily living only. The best model (see Section 4.7) relies on HR while *lying* down resting and while *walking at 4, 5 and 6 km/h*, together with the individual *height* and *age*, as independent variables.

Table 4.3: Heart Rate Normalization Factor Estimation Model ($R^2 0.87$).

Variable	Coefficient
Intercept	66.91
HR at rest	0.29
HR 4km/h	1.58
HR 5 km/h	−2.80
HR 6 km/h	2.18
Height	−0.17
Age	−0.23

4.5.3 HR normalized

Actual HR measurements are finally used after applying the HR normalization factor, derived with the normalization factor estimator, using the simple ratio: $NormalizedHR = \frac{CurrentHR}{Normalizationfactor}$

4.5.4 Personalized activity-specific EE estimation

Following the methodology applied in current state of the art EE estimation algorithm, EE is estimated by first classifying the activity performed, by means of ACC features, and then applying an activity-specific EE linear regression model. The activity-specific EE linear models use anthropometric characteristics, ACC and HR features. Thus, we developed six multiple linear regression models, one for each cluster of activities (see Table 4.4). Activity-specific ACC features for each model were selected using linear forward selection, in order to model intra-individual differences in EE. Normalized HR was used as a feature for the moderate to vigorous clusters (*dynamic*, *walking*, *running* and *biking*). The final feature set includes $Resting\ Metabolic\ Rate$ (RMR, computed with anthropometric variables only, according to the Harris-Benedict formula), $motion\ intensity\ (MI)$, $standard\ deviation\ (STD)$, $median\ (MED)$, $main\ frequency\ peak\ (FFTpeax_f)$ and its $amplitude\ (FFTpeak_a)$, $body\ weight\ (BW)$ and $Normalized\ Heart\ Rate$ $(HRNorm)$.

4.5.5 Feature extraction

ACC and HR features were used to derive activity recognition, walking speed, CRF (normalization factor) estimation and EE estimation linear models. ACC data from the three axes were segmented in 4 second windows, band-pass (BP) filtered between 0.1 and $10Hz$, to isolate the dynamic component caused by body motion, and low-pass (LP) filtered at $1\ Hz$, to isolate the static component, due to gravity. We selected a time window of $4s$, since it is short enough to detect changes in postures even for short breaks in sedentary time, and long enough to capture the repetitive patterns of activities, such as walking or running. Time and frequency features were extracted from each window over the three axes of the LP

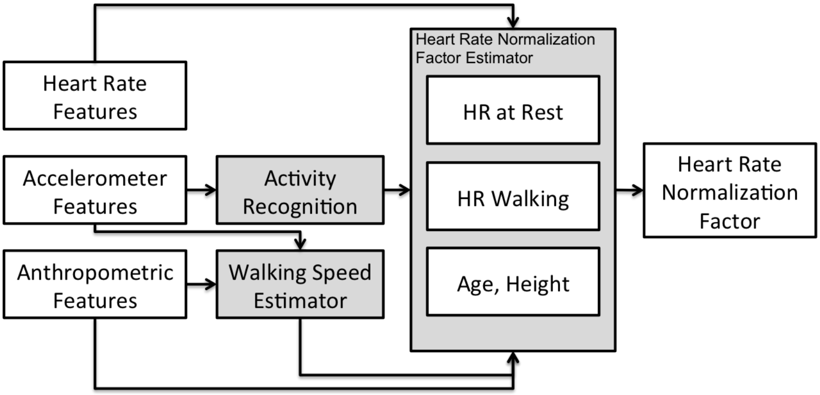

Figure 4.3: Components of the automatic HR normalization system used in this work to derive the normalization factor automatically.

Table 4.4: Activity-specific EE linear models.

Cluster	**Model**
Lying	$0.49 + 0.00068\ RMR - 29.66 MIx + 9.78\ STDx + 0.11\ MEDx + 0.68\ MEDy$
Sedentary	$0.31 + 0.00061\ RMR + 8.42\ MIx + 11.12\ MIy - 2.37\ MIz + 2.9\ STDx + 2.48\ STDy + 0.47\ MEDy - 0.14\ MEDz + 0.05\ FFTpeakYa$
Dynamic	$-3.43 + 5.95\ HRNorm + 0.035\ BW + 7.65\ MIy + 8.59\ MItot + 4.80\ STDx$
Walking	$-9.00 + 15.07\ HRNorm + 0.056\ BW + 3.91\ STDx$
Biking	$-10.58 + 0.0029\ RMR + 16.75\ HRNorm - 37.66\ MIx + 14.23\ MIy - 54.37\ VARy + 26.22\ STDx$
Running	$-8.73 + 11.50\ HRNorm + 0.12\ BW + 13.99\ MIy - 5.28\ STDy + 4.16\ MEDx - 3.70\ MEDz - 1.33\ FFTpeakXf$

and BP signal. Time features included *mean, mean of the absolute signal, magnitude, mean distance between axes, skewness, kurtosis, variance, standard deviation, coefficient of variation, range, min, max, correlation, inter-quartiles range, median* and *zero crossing rate*. Frequency features included: *spectral energy, entropy, low frequency band signal power (0.1 - 0.75 Hz), high frequency band signal power (0.75 - 10 Hz), frequency* and *amplitude of the FFT coefficients*. These features were selected due to high accuracy showed in past research [5-9]. The *mean HR* was extracted from R-R intervals, computed over 15 seconds windows. R-R intervals features were not included in the activity recognition and walking speed linear models. Feature extraction was performed in MATLAB (MathWorks, Natick, MA).

4.5.6 Feature selection

4.5.6.1 Activity type recognition

Feature selection was based on correlation, following the assumption that a good feature set includes features highly correlated with the class, but uncorrelated to each other. This step, as well as the subsequent classification, was implemented in Java using libraries provided by the $WEKA$ machine learning toolkit (University of Waikato, Hamilton, New Zeland). The final feature set (see Section 6.6.2) was used to train the SVM.

4.5.6.2 Multiple linear regression models

Feature selection for multiple linear regression models (six activity-specific EE models, one for each cluster, the walking speed estimator and the HR normalization factor estimator) were based on how much variation in the dependent variable each feature could explain, using linear forward selection. Participant-independent models were developed for each multiple linear regression model. Additionally, anthropometrics characteristics and Resting Metabolic Rate were added to the EE models depending on the cluster [9] (see Section 6.6.4).

4.5.7 Statistics and performance measure

All analysis were performed independent of the participant. Models were derived on all the participants but one, and validated on the remaining one. This leave-one-out procedure was carried out N times ($N = number\ of\ participants$), and results were averaged. Even though performance was evaluated independent of the subject, the reported models are derived including data from al participants (Tables 4.2, 4.3 and 4.4). Performance of the activity recognition was evaluated using the percentage of correctly classified instances for each cluster. The performance measures used for EE is the Root Mean Square Error (RMSE), averaged within an activity and between participants. Results are reported only in terms of RMSE because of the large inter-individual variability that is typical for EE estimates. Normalization procedures do exist (e.g. estimating in $kcal/kg$), but do not take into account that EE during different activities is affected differently by body weight. Performance of the walking speed linear model, as well as the HR normalization factor estimates, were evaluated using the RMSE and the percentage of the explained variance of the multiple linear regression model (R^2). As statistical analysis, paired t-tests between non-normalized and normalized results were used. Significance level α was set to 0.05 for all tests.

4.6 Measurement setup and data collection

4.6.1 Participants

Twenty-nine (22 male, 7 female) healthy participants took part in the experiment. Mean age was $30.9 \pm 5.5\, years$, mean weight was $72.6 \pm 12.5\, kg$, mean height was $177 \pm 9.3\, cm$ and mean BMI was $23.0 \pm 2.6\, kg/m^2$. Our internal Ethics Committee approved the study, and each participant signed an informed consent form.

4.6.2 Instruments

4.6.2.1 ECG Necklace

The ECG Necklace [98] is a low power wireless ECG platform (see Fig. 4.4). The system relies on an ultra-low-power ASIC for ECG read-out, and it is integrated in a necklace, providing ease-of-use and comfort while allowing flexibility in lead positioning and system functionality. It achieves up to 6 days autonomy on a $175\, mAh$ Li-ion battery. For the current study, the ECG Necklace was configured to acquire one lead ECG data at $256\, Hz$, and ACC data from a three-axial accelerometer (ADXL330) at $32\, Hz$. The sensor was placed on the chest with an elastic belt. The x, y, and z axes of the accelerometer were oriented along the vertical, medio-lateral, and antero-posterior directions of the body, respectively. Two gel electrodes were placed on the participant's chest, in the lead II configuration. Data were recorded on the on-board SD card to ensure no data loss. Data were also streamed in real-time to provide visual feedback of the system functionality to the experimenter.

4.6.2.2 Indirect calorimeter

Breath-by-breath data were collected using the Cosmed $K4b^2$ indirect calorimeter. The Cosmed $K4b^2$ weights $1.5\, kg$, battery included, and showed to be a reliable measure of EE [85]. The system was manually calibrated before each experiment according to the manufacturer instructions. This process consists of allowing the system to warm-up, following a double calibration, first with ambient air and then with calibration gas values. A delay calibration was performed weekly to adjust for the lag time that occurs between the expiratory flow measurement and the gas analyzers.

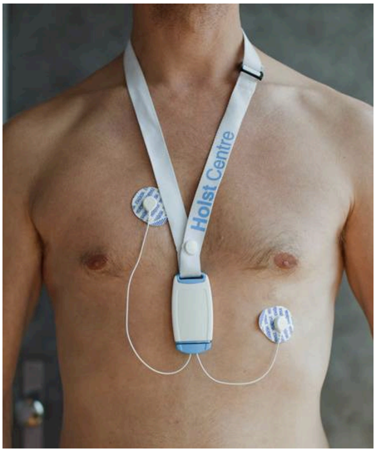

Figure 4.4: Wearable sensor used for this study (ECG Necklace). The sensor acquires 2-leads ECG and 3-axial acceleration.

4.6.3 Experimental design

Participants were invited for recordings on two separate days. They reported at the lab at $8.00\ a.m.$, after refraining from drinking (except for water), eating and smoking in the two hours before the experiment. The protocol included a wide range of lifestyle and sport activities, including sedentary and household activities. More specifically, day one consisted of activities selected as representative of common daily living of many people in industrialized countries [21]. The activities were: *lying down, resting, sitting stretching, standing stretching, desk work, reading, writing, working on a PC, watching TV, fidgeting legs, standing still, standing preparing a salad, washing dishes, stacking groceries, folding clothes, cleaning the table, washing windows, sweeping, vacuuming, walking self-paced, walking self-paced carrying books (4.5 kg), climbing stairs up, climbing stairs down*. Each sedentary and household activity was carried out for a period ranging from 4 to 12 minutes, with a 1 or 2 minutes break between the activities. Day two was carried out at the gym, where subjects performed a series of more vigorous activities, including: *walking at 3,4,5 and 6 km/h on a treadmill, walking at 4 km/h carrying a weight (5% of the subject's weight), walk- ing at 3 km/h, 5 and 10% inclination, walking at 5 km/h, 5 and 10% inclination, cycle ergometer at 60 and 80 rpm, low, medium and high resistance levels, running at 7,8,9 and 10 km/h*. Activities carried out at the gym were 4 minutes duration, except for free weights and running, which lasted for 1 to 2 minutes. Four participants did not perform all running activities and were excluded from data analysis.

4.6.4 Pre-processing

The dataset considered for this work contains about 70 hours of annotated data collected from 29 subjects, consisting of reference VO_2, VCO_2, three axial acceleration and ECG.

4.6.4.1 ECG Necklace data

Raw ECG and ACC data were downloaded from the SD card of the ECG Necklace. Raw data were exported into csv files containing time-stamped ECG and acceleration samples. A *Continuous Wavelet Transform* based beat detection algorithm was used to extract R-R intervals from ECG data, which output was manually examined to correct for missed beats that might be caused by motion artifacts [102].

4.6.4.2 Indirect calorimeter data

Breath-by-breath data acquired from the Comsed $K4b^2$ was resampled at $0.5\ Hz$. EE was calculated from O_2 consumption and CO_2 production using Weir's equation [128]. The first 1 or 2 minutes of each activity were discarded to remove non-steady-state data.

4.7 Results

4.7.1 Activity recognition

Subject-independent classification accuracy of the SVM used to select which cluster model to use in the EE estimate was 94.3%. More specifically, the accuracy was 100% for $lying$, 91% for $sedentary$, 87% for $dynamic$, 98% for $walking$, 91% for $biking$ and 99% for $running$.

4.7.2 Walking speed estimator

The walking speed multiple linear regression model could explain 94% of the variance in walking speed ($R^2 = 0.94$). RMSE of the model is $0.28 \pm 0.09\ km/h$.

4.7.3 Heart rate normalization factor estimator

The Heart Rate Normalization Factor multiple linear regression model could explain 87% of the variance ($R^2 = 0.87$). RMSE was 8.3 beats per minute (bpm) (see Fig. 4.5). Higher error was obtained with a second model, built using lower walking speeds only, since lower speeds will have higher chance to be detected in daily life (3 and $4\ km/h$, together with $height$ and age, RMSE $11.8\ bpm$). Fig. 4.5 shows an example of the normalization. By considering our normalization factor approach, the HR variance was clearly reduced. We concluded that the normalized HR can be used as part of the activity-specific EE models, reducing over or under-estimations.

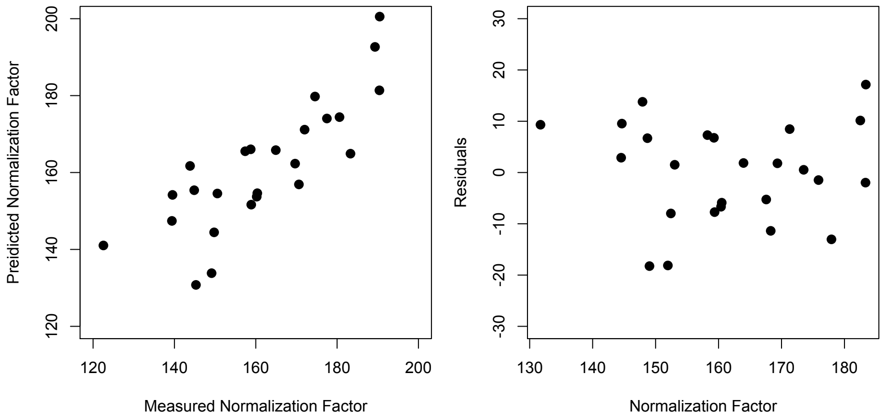

Figure 4.5: Scatterplot and residuals plot of measured (running on a treadmill) VS predicted (from age, height, HR at rest and while walking at 4,5 and 6 km/h) normalization factors (i.e. HR while running at 10 km/h).

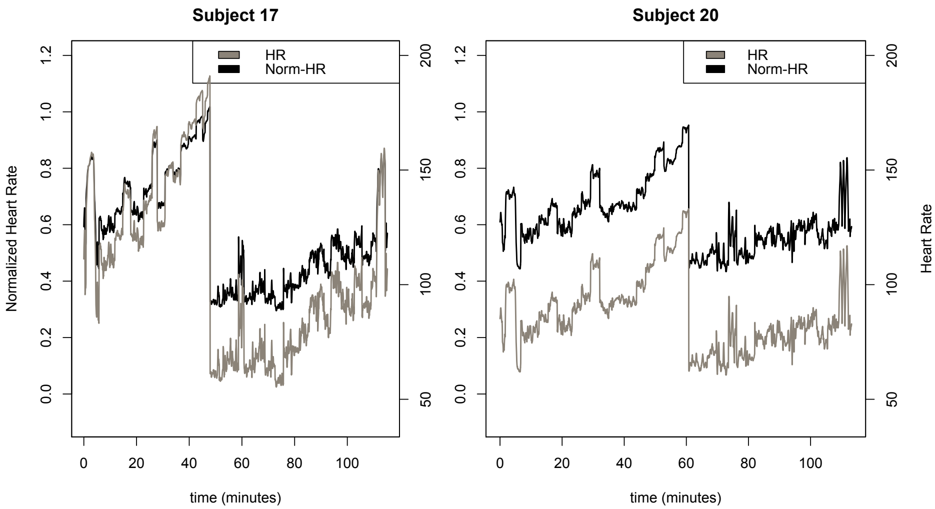

Figure 4.6: HR normalized for two participants using the approach proposed in this work. Once CRF is taken into account, absolute HR differences are significantly reduced, and Norm-HR can be used to estimate EE, reducing error. HR before normalization is shown for comparison (as in Fig. 9.1).

4.7.4 Energy expenditure estimation

RMSE for the EE estimate was $0.60 kcal/min$. More specifically, RMSE was $0.20 kcal/min$ for *lying*, 0.25 $kcal/min$ for *sedentary*, 0.58 $kcal/min$ for *dynamic*, 0.81 $kcal/min$ for *walking*, 0.92 $kcal/min$ for *biking* and 0.89 $kcal/min$ for *running*).

Fig. 4.7 shows the reduction in error for activity-specific EE models using HR (*dynamic*, *walking*, *running* and *biking*), when CRF is taken into account. RMSE was reduced from 0.60 to 0.58 $kcal/min$ for *dynamic* (3% error reduction, not significant), from 1.13 to 0.81 $kcal/min$ for *walking* (28% error reduction, $p = 0.00027 < \alpha$), from 1.38 to 0.92 $kcal/min$ for *biking* (33% error reduction, $p = 0.00037 < \alpha$) and from 1.25 to 0.89 $kcal/min$ for running (29% error reduction, $p = 0.01 < \alpha$).

4.8 Discussion and conclusions

In this paper we proposed a novel algorithm for activity-specific EE estimation based on a combination of ACC and HR data. By introducing a HR normalization factor, we were able to model the effect of CRF on HR during exercise. By normalizing HR responses from subjects with different levels of CRF, we could significantly reduce EE estimation error ($p < 0.05$ for walking, biking and running). More specifically, the proposed approach is able to reduce EE estimation error of activity-specific linear models (i.e. models developed specifically for an activity, and already including the best ACC features, as well as anthropometric characteristics) by an additional 28 to 33% compared to the best state-of-the-art models published up to date. The error reduction applies to non-sedentary clusters of activities, such as walking, biking or running at moderate intensities.

We believe this is a significant step towards personalized health and wellbeing monitoring. The proposed system uses a single monitoring device and is able to learn automatically from the user over time, collecting HR data while performing different activities (walking at different speeds, resting, etc.). The collected data is then used to determine the HR normalization factor, a coefficient representative of the CRF level of an individual.

Personalizing a system goes beyond the inclusion of the individual's anthropometric characteristics in the activity-specific EE linear models. In the future, the estimated normalization factor could be used as predictor to estimate $VO_2\ max$, using equations published in literature [62]. By doing so, a user would be aware of one of the most important health markers [81], without the burden and risks of regularly performing maximal or sub maximal tests.

We expect that our HR normalization approach will be most useful for sports training devices, where users and trainers are interested in accurate EE estimation under moderate to vigorous workloads. However, less active users willing to take up a more active lifestyle, or undergoing a physical activity intervention targeted in modifying behavior to increase level of activity, would also benefit. As a matter of fact, in the latter case CRF takes even a bigger role, since it typically changes faster in the transition from inactive to active lifestyle, while lower changes can

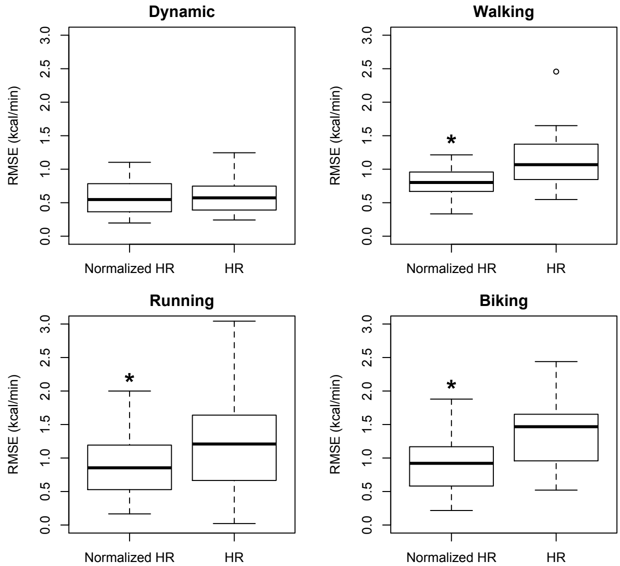

Figure 4.7: RMSE for the four moderate to vigorous clusters. Statistically significant differences are marked with * (paired t-test, $p < \alpha$, $\alpha = 0.05$) EE estimation error was significantly reduced for walking, running and biking. Error variance is reduced as well.

be expected for a continuously active lifestyle. Being able to monitor changes in CRF and HR over time would affect positively EE estimates, since EE estimates are highly dependent on HR and on the relation between CRF, HR and EE, as shown by our analysis. New opportunities for applications targeted at inducing behavioral change analyzing not only levels of PA, but also change in CRF and associated reduced risk of disease, could be developed building up on the proposed approach.

We recognize limitations in our study. Even though we developed an algorithm able to derive the HR normalization factor automatically, during regular activities, by combining rest and walking data with the subjectÕs anthropometric characteristics, we tested it using laboratory recordings only. We consider that the evaluation with lab data is a necessary first step, which can be sufficiently covered with reference measurements of EE. In particular, the approach allowed us to confirm performances of the individual estimators (activity recognition accuracy was 94.3%, walking speed RMSE was 0.28 ± 0.09 km/h). Overall, we conclude that an accurate personalized EE estimation using a single monitoring device and combining ACC and HR is feasible.

5

Automatic heart rate normalization for accurate energy expenditure estimation: an analysis of activities of daily living and heart rate features

M. Altini, J. Penders, R. Vullers, O. Amft
Adapted from: Methods of Information in Medicine 2014; 53 (5): 382-388.

Abstract

Background: Energy Expenditure (EE) estimation algorithms using Heart Rate (HR) or a combination of accelerometer and HR data suffer from large error due to inter-person differences in the relation between HR and EE. We recently introduced a methodology to reduce inter-person differences by predicting a HR normalization parameter during low intensity Activities of Daily Living (ADLs). By using the HR normalization, EE estimation performance was improved, but conditions for performing the normalization automatically in daily life need further analysis. Sedentary lifestyle of many people in western societies urge for an in-depth analysis of the specific ADLs and HR features used to perform HR normalization, and their effects on EE estimation accuracy in participants with varying Physical Activity Levels (PALs). Objectives: To determine 1) which low intensity ADLs and HR features are necessary to accurately determine HR normalization parameters, 2) whether HR variability (HRV) during ADLs can improve accuracy of the estimation of HR normalization parameters, 3) whether HR normalization parameter estimation from different ADLs and HR features is affected by the participants' PAL, and 4) what is the impact of different ADLs and HR features used to predict HR normalization parameters on EE estimation accuracy. Methods: We collected reference EE from indirect calorimetry, *accelerometer and HR data using one single sensor placed on the chest from* 36 *participants while performing a wide set of activities. We derived HR normalization parameters from individual ADLs (lying, sedentary, walking at various speeds), as well as combinations of sedentary and walking activities. HR normalization parameters were used to normalized HR and estimate EE. Results: From our analysis we derive that 1) HR normalization using resting activities alone does not reduce EE estimation error in participants with different reported PALs. 2) HRV features did not show any significant improvement in RMSE. 3) HR normalization parameter estimation was found to be biased in participants with different PALs when sedentary-only data was used for the estimation. 4) EE estimation error was not reduced when normalization was carried out using sedentary activities only. However, using data from walking at low speeds improved the results significantly* ($30-36\%$)*. Conclusion: HR normalization parameters able to reduce EE estimation error can be accurately estimated from low intensity ADLs, such as sedentary activities and walking at low speeds* ($3-4$ *km/h), regardless of reported PALs. However, sedentary activities alone, even when HRV features are used, are insufficient to estimate HR normalization parameters accurately.*

5.1 Introduction

5.1.1 Scientific background

Early epidemiological research focused on developing single models or branched equations combining accelerometer and HR data to predict EE [116, 17, 34]. These approaches are motivated by the relations between body movement and EE as well as between oxygen intake, HR and EE. The limitation of these methods include that a single accelerometer worn close to the body center of mass cannot detect low and upper body motion [29], the reduced relevance of HR during sedentary behavior and the need for individual calibration [34]. By introducing activity-specific models, consisting of a two-step process, where first an activity is recognized, and then an EE estimation model is applied, researchers were able to tackle some of these limitations [116, 17]. The relation between EE and acceleration as well as HR is peculiar of a specific context (e.g. activity), thus activity-specific models are able to capture this relation beyond what single regression models or branched models can do [29, 118, 9]. Even though algorithms including HR consistently provided improvements compared to accelerometers alone [116, 34, 9], the main limitation of HR - which is the need for individual calibration - requires a different solution. Decomposing the EE estimation process into activity-specific sub-problems is not sufficient to take into account the different relation between HR and EE in different individuals.

During moderate to vigorous PA, differences in HR between persons performing the same activity are mainly due to cardiorespiratory fitness (CRF). However, differences in CRF level do not cause different metabolic responses [104]. Nevertheless, CRF-related variance was tackled only by means of individual calibration [34] and/or by performing intense activities such as running [96]. For many practical applications personal calibration is not feasible since it would require every user to perform a suitable fitness test. We recently introduced a methodology to automatically normalize HR by estimating a normalization parameter that describes HR at a certain workload, using low intensity ADLs [13] (see Fig. 9.1). The methodology is based on the tight relation between CRF and the HR at a certain workload, which is the basis of sub-maximal CRF tests [62].

5.1.2 Rationale for the study

Practical conditions for performing the normalization automatically in daily life need further analysis. The sedentary lifestyle of many people in western societies [76] urge for an in-depth analysis of the specific ADLs required to predict HR normalization parameters, and their effects on EE estimation accuracy in persons with varying PALs. Additionally, HR variability (HRV) features from sedentary activities as well as moderate to intense ones have been shown to be linked to CRF level and PALs in past research [55, 37]. Even though this link is unclear, and results are often in disagreement [80, 56, 66], given the close relation between CRF and HR normalization parameters it is of interest to analyze if HRV features can predict HR normalization parameters and reduce EE estimation error.

5.1.3 Objectives of the study

This is the first analysis of how low intensity ADLs and HR features can be used to estimate HR normalization parameters, and their effects on EE estimation accuracy. Our objectives are: 1) To determine which ADLs and HR features are necessary to accurately determine HR normalization parameters, 2) To determine whether HRV during ADLs can improve accuracy of the estimation of HR normalization parameters, 3) To determine whether HR normalization parameters estimation from different ADLs and HR features is affected by the participants' PAL and 4) To determine what is the impact of different ADLs and HR features used to predict HR normalization parameters on EE estimation accuracy.

5.2 Methods

5.2.1 Participants

Participants were 36 (27 male, 9 female) self-reported healthy Holst Centre employees from diverse ethnic background. Mean age was 31.2 ± 5.7 years, mean weight was $73.3 \pm 11.2kg$, mean height was 176.6 ± 9.1 cm and mean BMI was 23.4 ± 2.4 kg/m^2. Imec's IRB approved the study, and each participant signed an informed consent form.

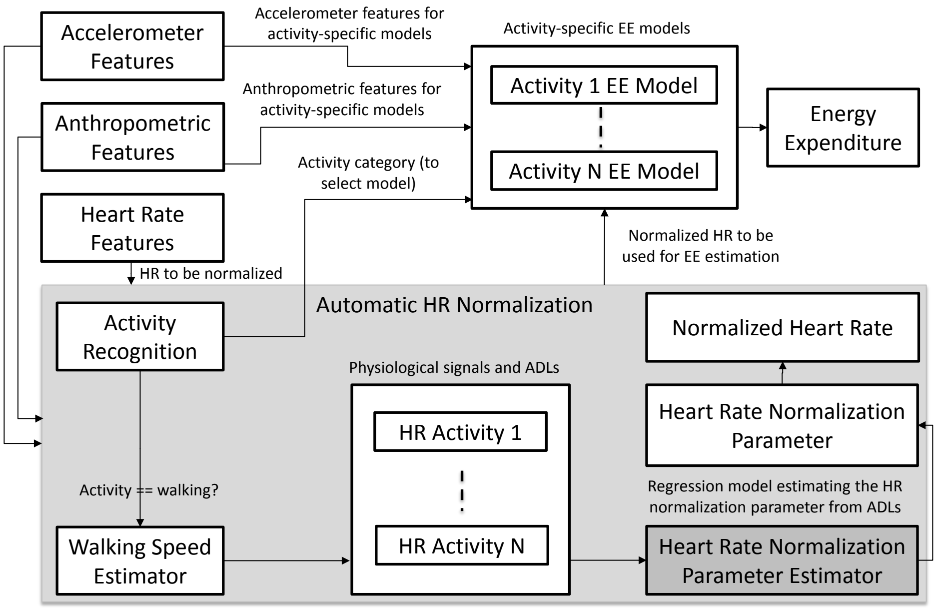

Figure 5.1: Overview on the activity-specific EE estimation and extension for automatic HR normalization using an HR normalization parameter estimated from ADLs. Accelerometer features are used for activity recognition, walking speed estimation and EE models. HR in specific activities ($1 \ldots N$, e.g. lying and walking at a certain speed) is used to estimate the HR normalization parameter. The HR normalization parameter is then used to normalize HR and predict EE with higher accuracy.

5.2.2 Study design

Participants reported at the lab after refraining from drinking (except for water), eating and smoking in the two hours before the experiment. The protocol consisted of common ADLs in industrialized countries [21], as well as intense activities. Activities were grouped into six clusters to be used for activity classification. The six clusters were lying (lying down), sedentary (sitting, standing, desk work, reading, writing, PC work, watching TV), dynamic (stacking groceries, washing dishes, cooking, folding clothes, sweeping, vacuuming), walking (treadmill flat at $3, 4, 5, 6$ km/h, inclined $3 - 5\%$, $3 - 5$ km/h), biking (low medium and high resistance level at 60 and 80 rpms), running (treadmill $7, 8, 9$ and 10 km/h). Activities lasted for a period of at least 4 minutes, with the exception of running (1 to 4 minutes).

5.2.3 Outcome measures

All analyses were performed independent of the participant (leave one subject out validation). Accuracy of the HR normalization parameter estimation was evaluated using: 1) Pearson's correlation between each HR feature and the HR normalization parameter, to determine the predictive power of each single feature in each ADLs, 2) the error derived from the difference between estimated and measured normalization parameters, to determine possible bias and precision of the estimate. As the measured normalization parameter we used the actual HR while running on a treadmill. 3) The Root Mean Square Error (RMSE) between estimated and measured normalization parameters, to determine the accuracy of the estimate. Additionally, participants were split in active (ACT) and inactive (INA) groups, based on reported PALs in order to determine possible PAL-induced bias in the estimation procedure. The performance measure used for EE was the RMSE, averaged within an activity and between participants. A one-way repeated-measures within-subjects ANOVA with five levels was used to compare RMSE between EE models. The Tukey test was used as a post hoc test to perform pairwise comparisons and identify significant differences. In addition, unpaired t-tests were used to compare INA and ACT groups. Significance was assessed at $\alpha < 0.05$ for all analyses.

5.2.4 Methods for data acquisition and measurement

5.2.4.1 ECG Necklace

The ECG Necklace [8] is a low power wireless ECG platform which was configured to acquire one lead ECG data at 256 Hz, and accelerometer data from a three-axial accelerometer at 32 Hz (see Fig. 9.2). The sensor was placed on the chest with an elastic belt. Two gel electrodes were placed on the participant's chest, in the lead II configuration. A *Continuous Wavelet Transform* based beat detection algorithm was used to extract R-R intervals from ECG data, which output was manually examined to correct for missed beats that might be caused by motion artifacts [102].

5.2.4.2 Indirect Calorimeter

Breath-by-breath data were collected using the Cosmed K4b2 indirect calorimeter. The Cosmed $K4b^2$ weights 1.5 kg including battery and showed to be a reliable measure of EE [85].

5.2.5 Methods for data analysis

Accelerometer and HR features were used to derive activity recognition models, walking speed, HR normalization parameter estimation models and EE estimation linear models (see Fig. 9.1). To estimate walking speed, we deployed multiple regression models using accelerometer-only features as predictors according

Figure 5.2: ECG Necklace. The device was used to acquire ECG and accelerometer data.

to [10, 63]. Details on the accelerometer features and on the implementation of the models have been widely covered elsewhere [9, 10]. Here, we will focus on the HR features and ADLs used for the estimation of the HR normalization parameter.

5.2.5.1 HR Features

HR features were extracted from R-R intervals, computed over 2 minutes windows to ensure sufficient frequency resolution in the Low Frequency band [23]. Time domain features included mean HR (meanHR), standard deviation of beat-to-beat intervals (SDNN), square root of the mean squared difference of successive R-Rs (rMSSD) and number of pairs of successive R-Rs that differ by more than 50 ms (pNN50). Frequency domain features included low (LF, 0.04-0.15 Hz) and high frequency power (HF, 0.15-0.40 Hz).

5.2.5.2 Automatic HR Normalization Factor Estimation from ADLs

Multiple linear regression models were built to analyze individual ADLs that can be recognized with high recognition rates (e.g. lying 100%, sedentary 91% and walking 98%, together with walking speed - RMSE 0.28 ± 0.09 km/h [10]), as well as combinations of such ADLs. For each ADL we built a multiple linear regression model using as predictors HR and/or HRV features during such ADL, and as dependent variable the HR normalization factor. As HR normalization factor we selected running at 9 km/h, since no performance improvement in EE estimation accuracy was shown in our dataset when using the HR at more intense workloads. The activities and combinations of activities selected were the following:

- Lying: lying down resting
- Sed: sedentary activities
- Walk 3-4-5-6: walking at 3-4-5 or 6 km/h
- Comb A: Lying, Sed, Walk3 and Walk4
- Comb B: Lying, Sed, Walk3, Walk4, Walk5 and Walk6

To analyze the impact of HRV features, two multiple regression models were built for each activity and combination, one including HR only, and one including HR and HRV features.

5.2.5.3 EE Estimation

EE was estimated by first classifying the activity performed using accelerometer features and then applying an activity-specific EE linear regression model. The activity-specific EE linear models use anthropometric characteristics, accelerometer and HR features. Thus, we developed six multiple linear regression models, one for each cluster. Normalized HRs (i.e. HR divided by the estimated HR normalization parameter) obtained from different sets of ADLs were used as predictors in the multiple regression models for moderate to vigorous clusters (dynamic, walking, running and biking).

5.3 Results

5.3.1 Automatic HR Normalization Factor Estimation from ADLs

Mean HR showed significant correlation with the HR normalization factor during all ADLs (lying 0.50, sed 0.50, walk3 km/h 0.86, walk4 km/h 0.86, walk5 km/h 0.88 and walk6 km/h 0.90, $p < \alpha$). No HRV feature was found significantly correlated to the HR normalization factor, in any ADL analyzed ($p > \alpha$ for each HRV feature in each ADL). Additionally, no HRV feature was able to discriminate between participants groups divided by PALs (INA vs ACT), in any activity except for low speed walking ($p < \alpha$ for SDNN, pNN50, LF and HF during walk3). Mean HR could discriminate between INA and ACT in all activities ($p < \alpha$).

Fig 6.3.a-c shows the density plot of the difference between estimated and measured HR normalization factors. The spread of the distribution reduced by 47% from lying to walk6. Fig 6.3.d-i show the difference distribution for models where HR or HR+HRV features were predictors, for single ADLs. No significant difference was found when including HRV features in any activity. RMSE was 17.6 bpm for lying data, 17.6 for sed, 10.5 for walk3, 10.3 for walk4, 10.4 for walk5, 9.4 for walk6, 11.8 for CombA and 9.0 for CombB. No differences in RMSE were found when HR and HRV features were combined ($p > \alpha$ for all activities). Fig 6.3.j shows the HR normalization factor estimation error when different ADLs are used as predictors, divided per PAL of the participants. When only resting data is used (e.g. lying), the HR normalization factor is overestimated for ACT participants, while it is underestimated for INA ones. No difference in the estimation accuracy between ACT and INA participants was found when walking data was included in the models as well (CombA and CombB), with higher walking speeds (CombB) showing higher precision (spread further reduced by 24%).

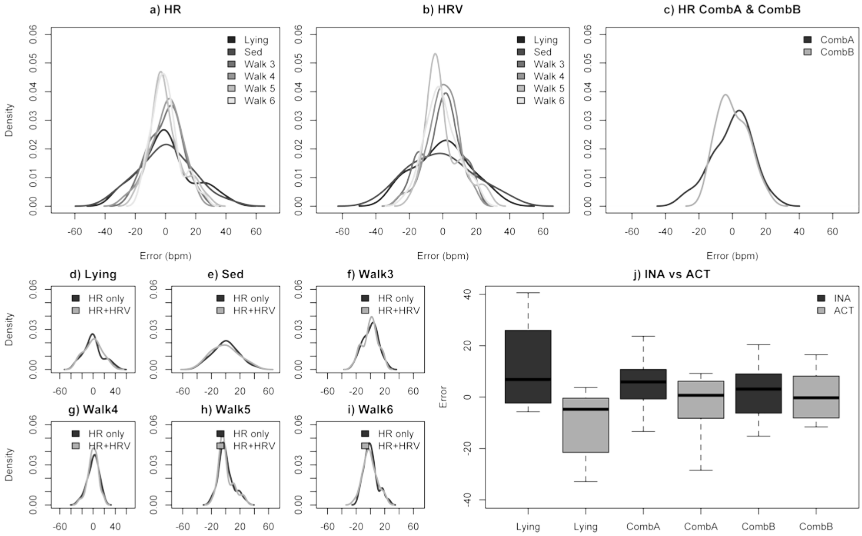

Figure 5.3: Difference between HR normalization parameter measured in the lab while the participants were running at 9 km/h and estimated HR normalization parameter as predicted from a) HR features only and b) HRV features, during a,b,d-i) single ADLs and c) combinations of ADLs. j) Prediction error divided by PAL.

5.3.2 EE Estimation

Fig. 8.2 shows the results of the HR normalization on EE estimation. The results of three different normalizations (from lying data only and using combined lying and walking speed data, CombA and CombB), is compared to the cases of no normalization (No Norm) and normalization using the measured HR normalization factor (Opt Norm). RMSE is reduced between 14 and 17% for dynamic activities, between 10 and 37% for walking activities, between 6 and 38% for biking activities and between 6 and 42% for running activities. No significant error reduction was shown when the HR normalization factor estimated using lying data only was used (6 to 17%, $p > \alpha$). Error reduction when walking data was included was significant for walking activities (36-37%, $p < \alpha$), biking activities (30-38%, $p < \alpha$) and running activities (31-40%, $p < \alpha$), but not for dynamic activities (14-15%, $p > \alpha$). CombA and CombB could reduce RMSE at the same extent the optimal HR normalization could (difference between ComA, CombB and Opt Norm was not statistically significant, $p > \alpha$).

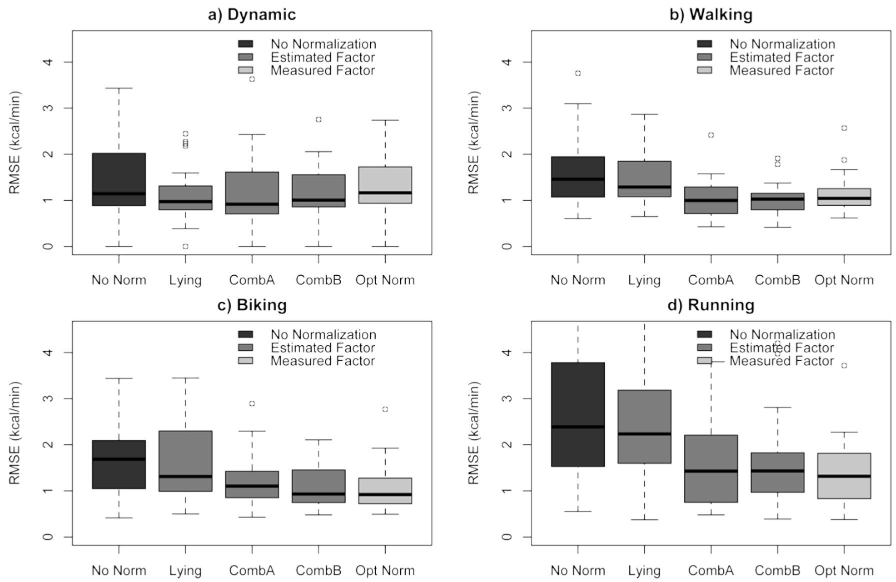

Figure 5.4: Algorithm performance in terms of RMSE of EE estimations during different moderate to vigorous activity clusters (a to d show dynamic activities, e.g. household, and walking, biking and running activities) where HR is not normalized (No Norm), normalized using lying data only, and normalized using ADLs included in CombA and ComB. Normalization performed using the measured HR normalization parameter (Opt Norm) is also shown for comparison. The first column of each subplot shows performance of state of the art activity-specific EE models combining accelerometer and heart rate features, but without HR normalization.

5.4 Discussion

5.4.1 Answers to study questions

We report the main findings of our analysis, in relation to the four objectives of this study. 1) To determine which ADLs and HR features are necessary to accurately determine HR normalization parameters: from our analysis we derive that resting activities alone are not sufficient to estimate HR normalization parameter, even if there is positive correlation between HR at rest and the HR normalization parameter. Thus, resting activities alone are unable to reduce EE estimation error in participants with different reported PALs. However, results obtained using data at rest and while walking at low speeds (e.g. < 4 km/h), showed results comparable to the ones obtained when including data while walking at higher speeds. Hence, ADL and HR features support estimating the HR normalization parameter in typical mixed lifestyle. 2) To determine whether HRV during ADLs can improve accuracy of the estimation of HR normalization parameters: from our analysis HRV features were unable to provide additional information and therefore improve the estimate accuracy of the HR normalization parameter (see Fig. 6.3.d-i). We attribute this finding to a weaker inter-personal relation between HRV and CRF. 3) Whether HR normalization parameter estimation from different ADLs and HR features is affected by the participants' PAL: our analysis showed that the normalization procedure works equally well in participants with different PALs, provided that walking data is included in the HR normalization parameter multiple linear regression models (see Fig. 6.3.j). Estimating precision is improved when data while walking at higher speeds is included in the HR normalization parameter multiple linear regression models (see Fig 6.3.a-c). 4) To determine what is the impact of different ADLs and HR features used to predict HR normalization parameters on EE estimation accuracy: our analysis showed that EE estimation accuracy when the HR normalization parameter is estimated from ADLs including walking (CombA and CombB) reaches the same accuracy of the optimal normalization that could be performed measuring the HR normalization parameter during a treadmill test (see Fig. 8.2).

5.4.2 Strength and weaknesses of the study

To the best of our knowledge, this is the first time that HR and HRV features are investigated during ADLs as predictors of a HR normalization parameter, together with the impact of such normalization procedure on EE estimation accuracy and participants with different PALs. Using the proposed personalization approach, it is possible to significantly reduce EE estimation error by automatically normalizing HR using low intensity ADLs, such as sedentary activities and walking at low speeds. However, we recognize limitations in our study. Even though we developed algorithms able to derive the HR normalization parameter automatically during ADLs, we tested it using laboratory recordings only. We consider that evaluation with lab data is a necessary first step. In particular, the approach allowed us to establish the accuracy of EE estimation models derived with ADLs and HR features. Further investigations should explore the relation between specific contexts and physiological parameters beyond linear models. The analysis should also be extended to a wider population consisting of participants with varying cardiorespiratory fitness level.

5.4.3 Results in relation to other studies

Previous work by our group [10] as well as others [118, 96] showed that normalizing the HR using a normalization parameter representative of CRF, such as the HR at a certain workload, can significantly reduce inter-person differences and consequently improve EE estimation accuracy. However, to determine the HR normalization parameter for an individual, required personal calibration (e.g. performing a treadmill test), which is not practical. Moreover, the calibration would need to be repeated frequently. In this study we investigated the possibility to determine the HR normalization parameter from different combinations of ADLs, including rest only activities (e.g. lying or sedentary). Additionally, we analyzed HRV features during ADLs, in the context of EE estimation.

Given the tight relation between CRF and the HR normalization parameter, which is the basis of sub-maximal CRF tests [62], it is of interest to review previous research on the relation between HRV and CRF. Many studies investigated the relation between HR and CRF during cross-sectional studies [80, 56, 66], as well as interventions [83, 32], and showed reductions in HR due to higher CRF levels, but no changes in HRV. Our results are in agreement with those, where HRV features could explain very little of the differences in fitness level, and mean HR was the best predictor of such differences. Since differences in HR and HRV features at rest are mainly driven by age, while feature differences during exercise are mainly driven by fitness [121], we investigated HRV during low intensity ADLs as well. However, we could not find a relation between HRV features while walking and the HR normalization parameter. Other authors did report a significant increase in HRV features and CRF following a physical activity intervention [87], however it is not clear if HRV features could be used as predictors of CRF.

5.5 Conclusions

We analyzed the impact of HR and HRV features in different ADLs as predictors of a HR normalization parameter necessary in order to reduce inter-person differences in HR and improve EE estimation accuracy. Using HR and HRV features during ADLs as predictors, we aimed at providing a normalization procedure able to automatically normalize HR without requiring any specific test. Overall, we conclude that an accurate personalized EE estimation is feasible, even when only data at rest and from walking at low speeds is available, as frequently occurring in today's lifestyle.

6

Personalizing energy expenditure estimation using physiological signals normalization during activities of daily living

M. Altini, J. Penders, R. Vullers, O. Amft
Adapted from: Physiological Measurements, vol. 35, no. 9, p. 1797, September 2014.

Abstract

In this paper we propose a generic approach to reduce inter-individual variability of different physiological signals (HR, GSR and respiration) by automatically estimating normalization parameters (e.g. baseline *and* range*). The proposed normalization procedure does not require a dedicated personal calibration during system setup. On the other hand, normalization parameters are estimated at system runtime from sedentary and low intensity Activities of Daily Living (ADLs), such as* lying *and* walking*. When combined with activity-specific EE models, our normalization procedure improved EE estimation by* 15 *to* 33% *in a study group of* 18 *participants, compared to state of the art activity-specific EE models combining accelerometer and non-normalized physiological signals.*

6.1 Introduction

The inclusion of physiological signals such as HR, Galvanic Skin Response (GSR), respiration, skin temperature or humidity, in combination with accelerometers, consistently provided better EE estimation results than accelerometers alone [9, 34, 129, 114]. However, inter-individual differences in physiology, as well as the consequent need for individual calibration, limit accuracy and practical applicability of such systems [11, 34, 43]. Breaking down the EE estimation process into activity-specific sub-problems is not sufficient to take into account the different relation between physiological signals and EE in different individuals. A method is needed to automatically normalize physiological signals without requiring individual calibration and fully exploit the relation between such signals and EE.

In this paper, we introduce a generic method to personalize EE estimates, by normalizing physiological signals from Activities of Daily Living (ADLs). Our contribution is two-fold:

1. We introduce a method able to normalize multiple physiological signals (HR, GSR and respiration) by automatically estimating *normalization parameters* (i.e. *baseline* and *range*). The proposed methodology uses low intensity ADLs, such as *lying* down and *walking* and is independent of the underlying physiological process driving inter-individual differences.

2. We evaluate the benefit of the proposed normalization methodology for activity-specific EE estimation. We implemented activity-specific models combining accelerometer and physiological data from two wearable sensors, located at the chest and wrist. In a study group of 18 participants, we show error reductions between 15% and 33% when normalized physiological signals are used, compared to state of the art activity-specific EE models without normalized physiological signals.

6.2 Related work

6.2.1 EE estimation in epidemiological research

Typically, accelerometer based methods use *activity counts*, a unit-less measure representative of whole body motion, as independent variable in the regression model developed to predict EE [52]. The main shortcoming is that a single model does not fit all the activities, since the slope and intercept of the regression model changes according to the activity performed. EE estimation based on HR suffers from different problems. First, HR based estimations are inaccurate during sedentary behavior, given that HR is also affected by non-activity related factors, such as stress and emotions. Secondly, HR based models need individual calibration to perform accurately [34]. The highly correlated relation between HR and EE within one individual changes substantially between individuals [10].

6.2.2 Machine learning methods for EE estimation

The latest algorithms for EE estimation use machine learning techniques. Some authors applied machine learning methods to directly estimate EE from accelerometer features, using for example neural networks [60, 103]. However these approaches suffer from the same limitations of the *activity counts*-based approaches, being unable to capture the peculiarities of the relation between accelerometers features and EE during different activities [29, 105]. Others extended the single model approach, performing activity recognition over a pre-defined set of activities, and then applying different methods to predict EE [9, 29, 118, 107]. These models are typically called *activity-specific*. Additionally, some hybrid approaches have been developed. Unsupervised clustering was used to avoid time consuming activity labeling during data collection, still dividing the EE estimation problem into sub-problems [45]. However, this approach also showed sub-optimal performance compared to activity-specific models.

Given the substantial amount of work using activity-specific models and the consistent improvements obtained compared to other methods, as reported by [9, 29, 105], we believe that activity-specific models are presently the best methodology to follow when developing EE estimation algorithms. However, inter-individual differences in physiology, as well as the resulting need for individual calibration, limit the accuracy and practical applicability of EE models using physiological signals [11, 34]. Partitioning the EE estimation into activity-specific sub-problems is not sufficient to address the relation between physiological signals and EE in different individuals.

6.2.3 Normalization of physiological signals

During moderate to vigorous PA, differences in physiological signals between individuals performing the same activities can be due to a variety of factors. While cardiorespiratory fitness (CRF) is the main factor driving changes in HR during physical exercise [121], differences in respiration, skin temperature or GSR might be caused by different underlying processes or characteristics of the person [109]. We recently investigated the relation between multiple physiological signals (HR, respiration rate, GSR and skin humidity) and EE for activity-specific EE estimation models [11]. Physiological signals showed higher correlation with EE compared to accelerometer data. However, subject-independent models including physiological signals performed sub-optimally, confirming the need for individual calibration. Individual calibration limits practical applicability, since the individual relation between a physiological signal and EE needs to be determined for the algorithm to be accurate. To the best of our knowledge, the only attempt to automatically normalize physiological signals without requiring individual calibration was reported by our group. In [10], we normalized HR from Activities of Daily Living (ADLs) exploiting the known relation between HR, CRF and EE.

In this work, we propose a generic methodology to automatically normalize different physiological signals at runtime, independently from the causes driving inter-individual differences in such signals. The proposed normalization methodology uses low intensity ADLs to avoid individual calibration in laboratory or supervised settings.

6.3 Relation between EE, accelerometer and physiological data

In this section, we introduce the problem of inter-individual differences in physiological signals when estimating EE. Figure 9.1.a shows the correlation between different signals and EE. Even though physiological signals show higher correlation with EE compared to accelerometer data, subject-independent models including physiological signals perform sub-optimally, confirming the need for individual calibration (see figure 9.1.b). Figure 9.1.b shows the larger individual errors obtained when using physiological signals in subject-independent models, compared to accelerometer only models (A-C and A-W). HR-based estimates still report the lowest error, but with the highest variability. When comparing subject-independent and subject dependent models, little difference is found for accelerometer-based models (3-4%), while physiological signals-based models showed error increase up to 50% (see figure 9.1.c). Figure 9.2 highlights the inter-individual differences peculiar of physiological signals, for the cases of HR and GSR. For two subjects with similar body size, EE and accelerometer data is similar during different activities, however large inter-individual differences in physiology (both GSR and HR) can be seen. Clearly, if these signals are used to estimate EE, underestimations and overestimations will occur.

6.4 Methodology overview

Our approach is to estimate normalization parameters of physiological variables during ADLs, and use normalized physiological variables for activity-specific EE estimation. When determining the signal range, we are interested in estimating the physiological signal value at *rest* ($X_{phy_{base}}$), as well as the value that an individual would reach when performing a *high intensity activity* ($X_{phy_{high}}$).

We hypothesize that the physiological signal value during a *high intensity activity* ($X_{phy_{high}}$) can be estimated from ADLs, such as *resting* and *walking*, thus without requiring any specific calibration test. Figure 6.3 shows a block diagram of the normalization methodology and its three main logical blocks: *a) the recognition of type and intensity of ADLs*, such as *lying*, *walking* and *walking speed*, *b) the estimation of normalization parameters using ADLs* and *c) the normalization of physiological signals*.

As in standard activity-specific models, we divided the EE estimation process into activity recognition and activity-specific regression models. Physiological signals are normalized using the estimated *normalization parameters* (i.e. *baseline* and *range*), before being used in the activity-specific models. Assuming n clusters of activities c_i:

$$C = \{c_1, \ldots, c_n\}, \forall c_i \in C, \quad \exists \quad y_{act_i} = X_{act_i}\beta_{act_i} + \epsilon \tag{6.1}$$

y_{act_i} is the vector of actual EE values for a specific cluster of activities, β_{act_i} is the vector of regression coefficients, and X_{act_i} is the vector of m input features.

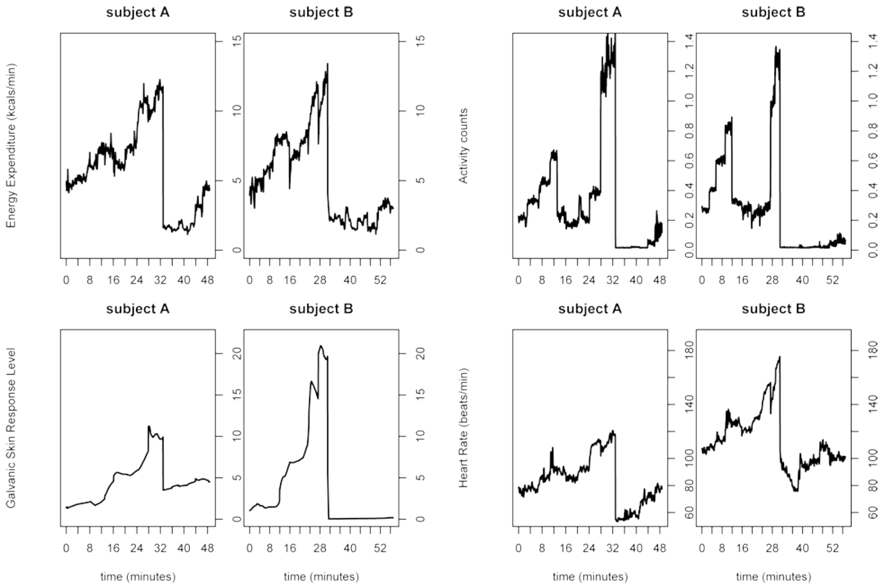

Figure 6.1: Reference EE, accelerometer and physiological data during a series of physical activities for two subjects with similar body size. While EE and accelerometer data show similar results and low inter-individual variability, big differences are found in both GSR and HR, highlighting the need for normalization of these parameters before their use for EE estimation.

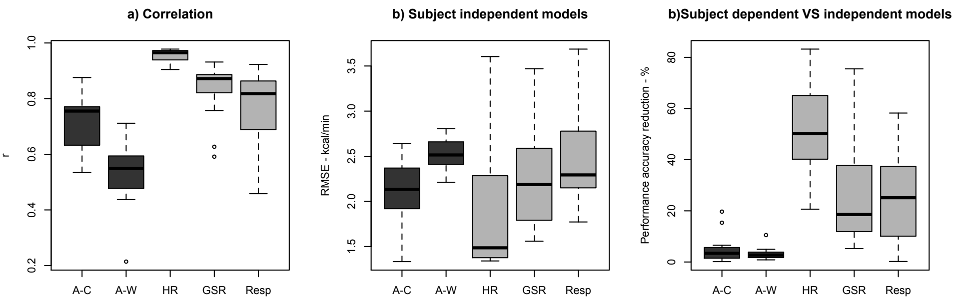

Figure 6.2: a) Correlation between accelerometer and physiological data with EE, b) Root Mean Square Error of subject independent EE models developed using accelerometer or physiological data, c) performance accuracy reduction when moving from subject dependent to subject independent models. A-C is accelerometer data at the chest, A-W is accelerometer data at the wrist, HR is heart rate, GSR is galvanic skin response, Resp is respiration.

Features can be grouped into accelerometer features (X_{acc_i}), anthropometric characteristics (X_{ant}) and normalized physiological signals (X_{phy_n}).

The normalized physiological signals (X_{phy_n}, block *c* in figure 6.3) are derived using the normalization parameters (i.e. the *baseline* - $X_{phy_{base}}$ - and *range* - $X_{phy_{range}}$ - of a certain signal for a specific individual), according to the following equation:

$$X_{phy_n} = (X_{phy} - X_{phy_{base}})/X_{phy_{range}} \tag{6.2}$$

Where X_{phy} are the non-normalized physiological signals. $X_{phy_{base}}$ and $X_{phy_{range}}$ are determined automatically from ADLs. More specifically, $X_{phy_{base}}$ is the value of the physiological signal X_{phy} when the user is *lying down resting*, while $X_{phy_{range}}$ is:

$$X_{phy_{range}} = X_{phy_{high}} - X_{phy_{base}} \tag{6.3}$$

$X_{phy_{high}}$ is the estimated physiological value for a particular user during a *high intensity activity* (e.g. *running at 8 km/h*). Instead of using a high intensity activity or calibration test, we estimate $X_{phy_{high}}$ using a multiple linear regression model. The regression maps physiological signals during various ADLs (X_{ADL}) to the physiological signals value during a *high intensity activity* ($X_{phy_{high}}$, see figure 6.3, blocks *a,b*):

$$X_{phy_{high}} = X_{ADL}\beta_{ADL} + \epsilon \tag{6.4}$$

where X_{ADL} is the vector of physiological signals values in pre-defined ADLs, such as *lying* down resting and *walking* at certain speeds (e.g. *4* to *6 km/h*), while β is the vector of regression coefficients.

6.5 Measurement setup and data collection

6.5.1 Participants

Eighteen (14 male, 4 female) healthy adults took part in the experiment. Mean age was $32.1 \pm 5.8 years$, mean weight was $73.6 \pm 9.4 kg$, mean height was $176.3 \pm 9.5 cm$ and mean BMI was $23.62 \pm 1.66 kg/m^2$. Imec's internal Ethics Committee approved the study. Each participant signed an informed consent form.

6.5.2 Instruments

Two wearable sensors were used for data collection, imec's ECG Neckalce and Wristband (see figure 6.4). The ECG Necklace was configured to acquire one lead ECG data at 256 Hz, and accelerometer data at 32 Hz. Two gel electrodes were placed on the participant's chest. Imec's Wristband was configured to acquire phasic and tonic GSR data at $128 Hz$ and accelerometer data at $32 Hz$. A *Continuous Wavelet Transform* based beat detection algorithm was used to extract R-R intervals

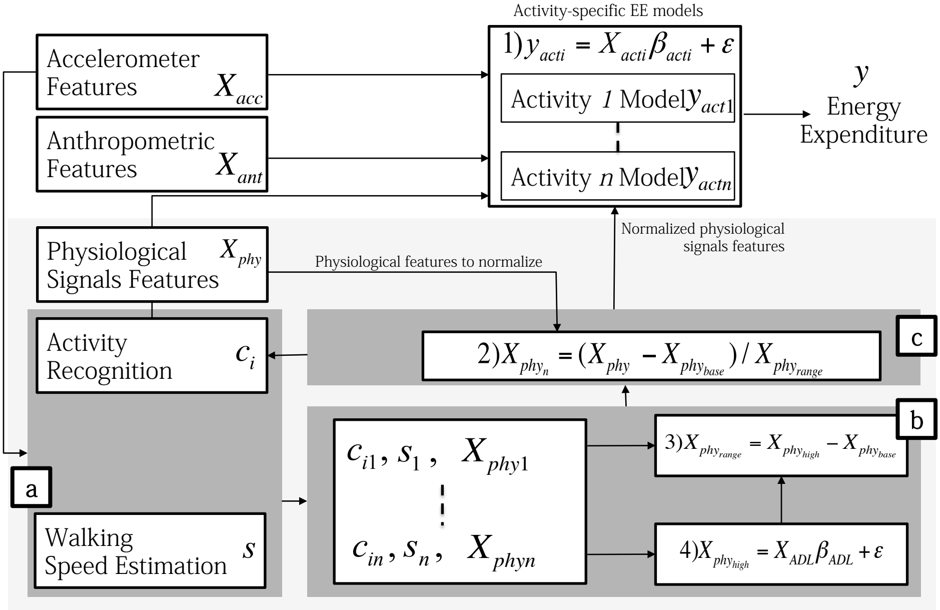

Figure 6.3: Overview of our method to normalize physiological signals and estimate EE. Normalized physiological signals are used for activity recognition and EE estimation models. a) components required for the recognition of type and intensity of ADLs, b) components for the estimation of normalization parameters and c) equation used to normalize physiological signals.

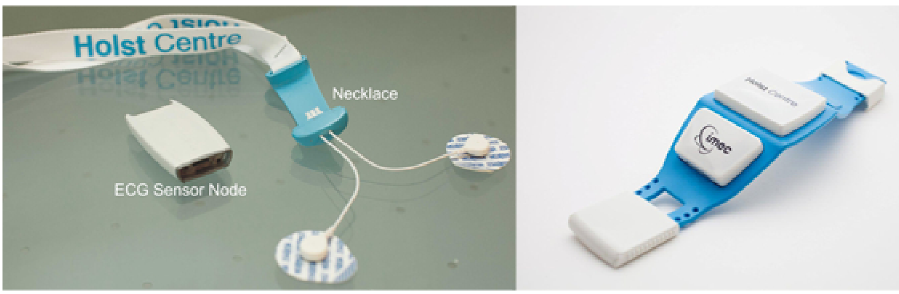

Figure 6.4: The two wearable sensors used in this experiment, ECG Necklace (left) and Wristband (right).

from ECG data, which output was manually examined to correct for missed beats that might be caused by motion artifacts [102]. Additionally, reference EE was collected using the *Cosmed* $K4b^2$ indirect calorimeter [85].

6.5.3 Experimental design

Participants were invited for recordings and reported to the lab after refraining from drinking (except for water), eating and smoking in the two hours before the experiment. The first part of the protocol consisted of activities selected as representative of common daily leaving of many people in industrialized countries [21]. The activities were: *lying down, resting, desk work, writing, working on a PC, standing still, washing dishes, stacking groceries, cleaning the table, vacuuming, walking self-paced, climbing stairs up, climbing stairs down*. Each sedentary and household activity was carried out for a period ranging from 4 to 12 minutes. The second part of the protocol was carried out at the gym, where participants performed a series of more vigorous activities, including: *walking at 3,4,5 and 6 km/h on a treadmill, walking at 3 km/h, 10% inclination, cycle ergometer at 60 and 80 rpm, low, medium and high resistance levels, running at 7,8,9 and 10 km/h*. Activities carried out at the gym were 4 minutes duration, except for running, which lasted between 1 and 4 minutes.

6.5.4 Statistics and performance measures

All analysis were performed independent of the participant (leave one subject out cross-validation). Performance of the activity recognition was evaluated using the percentage of correctly classified instances for each cluster. The performance measure used for EE were the Root Mean Square Error (RMSE), as commonly used to report EE estimation errors, and the Mean Absolute Percentage Error (MAPE), which provides an indication of the error in relation to the EE required by the performed activity. Performance of the *normalization parameters* estimation and walking speed estimation models were evaluated using the RMSE and the percentage of the explained variance of the multiple linear regression model (R2). As statistical analysis, paired t-tests between non-normalized and normalized results were used. Significance level α was set to 0.05. To allow for comparisons between methodologies and sensor locations, we implemented six configurations (three for the Necklace and three for the Wristband): $1)$ accelerometer data only, $2)$ combined accelerometer and non-normalized physiological data, $3)$ combined accelerometer and normalized physiological data. To evaluate the accuracy of the normalization parameters estimation against the ideal case of individual calibration, single regression models were built using as predictors the physiological signals only (HR, GSR level and respiration rate), and EE as dependent variable. Two models were implemented for each signal. One model included physiological signals normalized using the actual $X_{phy_{high}}$, determined while subjects were running on a treadmill (individual calibration). The second model included physiological signals normalized using the estimated normalization parameters.

Table 6.1: Distribution of the activities into the six clusters used for activity recognition.

Cluster name	**Original activities**
Lying	Lying down resting
Sedentary	Sitting resting, desk work, writing, working on a PC, standing still
HWBM/Dynamic	Stacking groceries, washing dishes, cleaning and scrubbing, vacuuming
Walking	Treadmill (flat: $3, 4, 5, 6$ km/h, incline: $3km/h$ 10%, self-paced, stairs up and down)
Biking	Cycle ergometer, low, medium and high resistance level at 80 rpm
Running	$7, 8, 9, 10$ km/h on a treadmill

These models were also compared against single regression models using non-normalized physiological signals as dependent variables, to evaluate the impact of the normalization procedure.

6.6 Methods implementation

6.6.1 Pre-processing

The dataset acquired in this work consists of reference VO_2, VCO_2, three axial acceleration from chest (A-C) and wrist (A-W), ECG, respiration rate and GSR. EE was calculated from O_2 and CO_2 (Weir 1949). Two subjects were unable to perform all activities, while data from one subject had to be discarded due to sensor failure.

6.6.1.1 Activity type clusters.

We manually grouped the activities into six clusters related to the activity type and involved motion patterns (see table 1). We included *lying* and *sedentary* as inactive clusters. Additionally, we included four active clusters, one representative of household activities and dynamic transitions between activities, namely the *high whole body motion cluster (HWBM or Dynamic)* and three related to locomotion and active transportation, namely *walking*, *biking* and *running*. The HWBM cluster is useful in distinguishing sedentary behavior and non-sedentary daily life activities even when only one sensor is used [9, 29].

6.6.1.2 Feature extraction and selection.

Accelerometer data from both sensors were segmented in 4 second windows, band-pass filtered between 0.1 and 10 Hz, to isolate the dynamic component, and low-pass filtered at 1 Hz, to isolate the static component. The feature set includes; *mean of the absolute band-passed signal, magnitude and inter-quartile range*, *median*, *variance and standard deviation* and *main frequency peak and amplitude of the main frequency peak*. Feature selection for activity type recognition was based on mutual information [22], while feature selection for activity-specific EE models was automated using linear forward selection. Anthropometrics features were added depending on the activity cluster, following the methodology of [9]. Features derived from physiological signals were used for both activity recognition and EE models. The most discriminative features were selected based on correlation. Selected features were; *mean HR*, *mean skin conductance level* and *respiration rate*. Features were extracted over 15 seconds windows.

6.6.2 Activity recognition

Given the positive results in past research on activity recognition, we selected Support Vector Machines (SVMs) as classifiers. For the SVMs, we used a polynomial kernel with degree 5 ($\lambda = 10$, $C = 1$). Activity recognition was used for EE estimation, and as part of the automatic physiological signals normalization system.

6.6.3 Automatic physiological signals normalization using ADLs

Two *normalization parameters* are required to perform the physiological signals normalization, *baseline* and *range*. While the *baseline* is determined as the physiological signals value while *lying*, a multiple linear regression model is built to predict the physiological signals values while performing a *high intensity activity* ($X_{phy_{high}}$ i.e. an individual's physiological signal while *running at 8 km/h*) from physiological signals values while $walking$. We selected *lying* and *walking* as the ADLs to use given the low intensity and high accessibility of such activities. We chose the range between 4 and 6 km/h for *walking speeds*, since speeds close to this range were often reported as the average walking speeds in healthy individuals (5.3 km/h in [36] and 5 ± 0.8 km/h in [89]). The walking speed estimator is a multiple linear regression model using as predictors the individual's *height* and the following accelerometer features: *main frequency peak on the X axis, mean absolute value of the band-passed signal, sum of the variance on the three axis, inter-quartile range on the X and Y axis* and *high frequency band signal power on the X and Z axis*.

The vector X_{ADL} in equation 4, was implemented as:

$$X_{ADL} = [X_{phyLying}, X_{phyWalking4}, X_{phyWalking5}, X_{phyWalking6}] \tag{6.5}$$

Where $X_{phyLying}$ and $X_{phyWalkingN}$ are the *means* of the physiological signals values while *lying* and *walking* at *N km/h*, for a certain user. $N = 4, 5, 6$. Actual physiological signal values are finally normalized according to equation 2 in section 6.4, removing the *baseline* and dividing by the estimated *range*.

6.6.4 Personalized activity-specific EE estimation

Following the methodology applied in current state of the art EE estimation algorithm, EE is estimated by first classifying the activity performed and then applying an activity-specific EE linear regression model. The activity-specific EE linear models use anthropometric characteristics, accelerometer and physiological signals features.

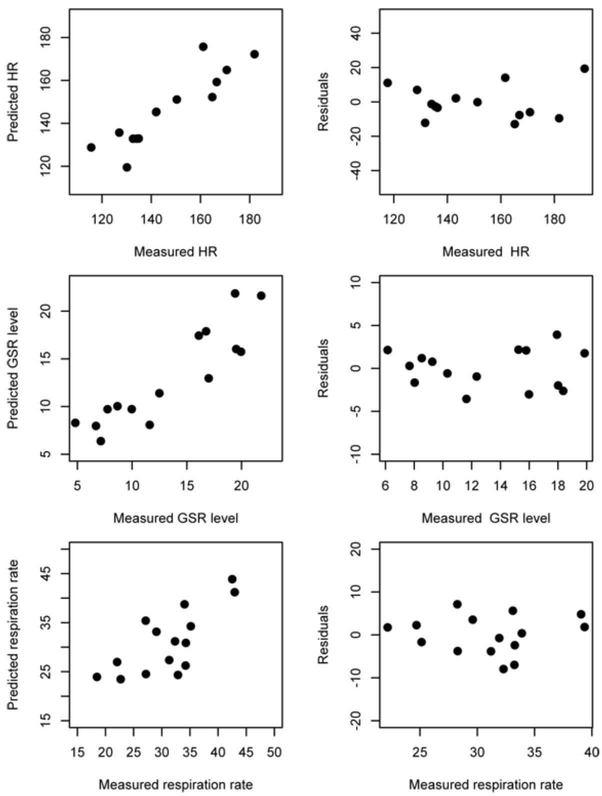

Figure 6.5: Scatterplot and residuals per study participant of measured (running on a treadmill) vs. predicted (from physiological signals during ADLs) physiological signals values during a *high intensity activity* ($X_{phy_{high}}$). $X_{phy_{high}}$ is used together with $X_{phy_{base}}$ to determine the *range* and normalize the signals.

6.7 Results

6.7.1 Automatic physiological signals normalization using ADLs

Activity recognition accuracy for the ADLs used by the normalization methodology was 100% for *lying* and 98% for *walking*. The walking speed multiple linear regression model could explain 94% of the variance in walking speed ($R^2 = 0.94$). RMSE of the model was 028 ± 0.09 km/h. Both models were previously reported in [10]. The multiple linear regression models used to estimate $X_{phy_{high}}$ could explain 90% of the variance for HR, 88% of the variance for GSR and 72% of the variance for respiration rate (R^2). RMSE was 9.3 beats per minute for HR, 2.4 μS for GSR and 4.8 breaths per minute for respiration rate. Figure 6.5 shows the relation between the measured and estimated $X_{phy_{high}}$. RMSE for single EE estimation models using physiological data only was 1.91 kcal/min for HR, 2.29 kcal/min for GSR and 2.49 kcal/min for respiration. RMSE for single EE estimation models using estimated *normalization parameters* was 1.18 kcal/min for HR, 1.96 kcal/min for GSR and 2.14 kcal/min for respiration. No difference was found when comparing the models to single EE estimation models using measured *normalization parameters* (i.e. performing individual calibration) - $p = 0.89 > \alpha$ for HR, $p = 0.08 > \alpha$ for GSR and $p = 0.68 > \alpha$ for respiration rate. EE estimation error was reduced by $60\%, 25\%$ and 18% for HR, GSR and respiration rate respectively, when compared to non-normalized models.

6.7.2 Personalized activity-specific EE estimation

6.7.2.1 Activity cluster classification

Subject independent classification accuracy of activity type for the ECG Necklace using accelerometer features only was 93%. Performance was improved by 1% when physiological signals were included in the model, and by 3% when normalized physiological signals were included ($p = 0.08 > \alpha$, not significant). Accuracy for the Wristband was 76%. Accuracy increased by 4% when physiological signals were included in the model, and by 6% when normalized physiological signals were included ($p = 0.04 < \alpha$).

6.7.2.2 Activity-specific EE estimation.

RMSE for the ECG Necklace EE estimation models - average of the six clusters - was 1.26 kcal/min when accelerometer-only data was used, 1.11 kcal/min when combining accelerometer and physiological data, and 0.83 kcal/min when combining accelerometer and normalized physiological data ($p = 0.02 < \alpha$). RMSE for the Wristband EE estimation models - average of the six clusters - was 2.47 kcal/min when accelerometer-only data was used, 1.42 kcal/min when combining accelerometer and physiological data, and 1.23 kcal/min when combining accelerometer and normalized physiological data ($p = 0.01 < \alpha$). Normalized physiological signals could reduced EE RMSE by 33% for the ECG Necklace and

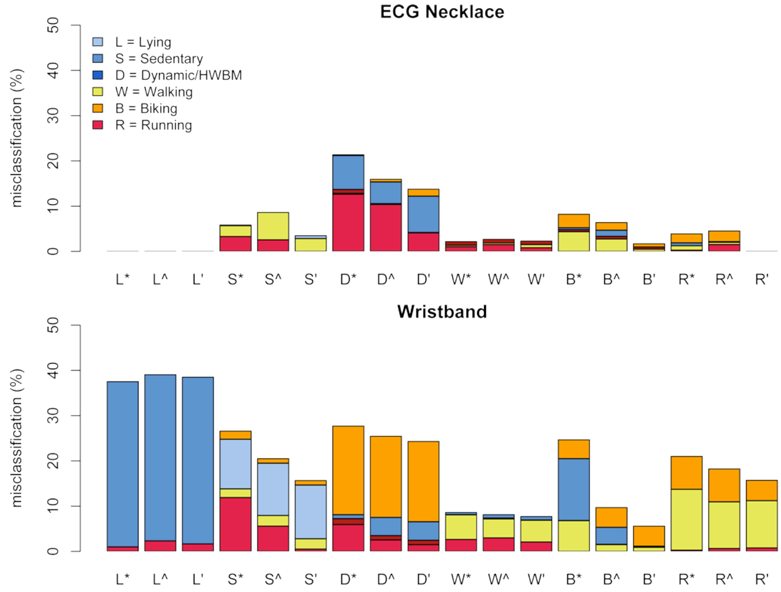

Figure 6.6: Misclassification of the activity recognition models per cluster of activities. *indicate accelerometer-only models, Λ indicate models combining accelerometer and physiological data, $'$ indicate models combining accelerometer and normalized physiological data.

by 15% for the Wristband. Misclassification effect (i.e. increased RMSE due to the application of the wrong EE model) when no physiological signals were used was 20% for the ECG Necklace and 125% for the Wristband (due to the high confusion between active and inactive clusters). Including physiological signals reduced the misclassification effect to 10% for the ECG Necklace and 29% for the Wristband. Normalized physiological signals could further reduce the misclassification effect, which was 4% for the ECG Necklace and 19% for the Wristband. Details for each model and activity are listed in table 6.2.

6.8 Discussion

In this paper we introduced a method to normalize multiple physiological signals (HR, GSR and respiration) by automatically estimating *normalization parameters*. The proposed method uses low intensity ADLs such as *lying* down resting and *walking at different speeds* to estimate the *normalization parameters*, and it is independent of the underlying physiological process driving inter-individual differences. To validate our methodology, we implemented activity-specific models combining accelerometer and physiological data from two wearable sensors, located at the chest and wrist. We evaluated the impact of the proposed normaliza-

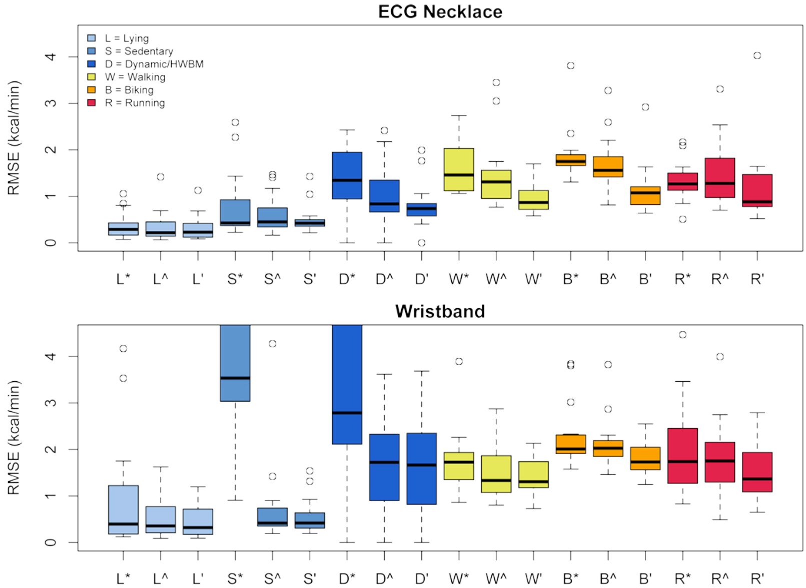

Figure 6.7: RMSE of activity-specific EE estimation models for the ECG Necklace and Wristband sensors, including misclassification effects. *indicate accelerometer-only models, Λ indicate models combining accelerometer and physiological data, ′ indicate models combining accelerometer and normalized physiological data.

tion methodology for activity-specific EE estimation, analyzed on the same subjects and activities.

To the best of our knowledge, this is the first work which aims at defining a generic method able to automatically normalize physiological signals. By applying the proposed normalization method, we could significantly reduce estimation errors for activity recognition and EE estimation. Other advantages that emerge from our normalization method: by normalizing physiological signals from data acquired during ADLs over a recent period of time (e.g. 2 weeks), the system could adapt to changes in physiological or environmental factors. Changes in physiology (e.g. CRF) would for example affect HR, while changes in environmental factor (e.g. temperature) would affect GSR, requiring a new individual calibration. However by estimating the *normalization parameters* from ADLs, the system could automatically adapt to such changes, without requiring repeated individual calibrations.

6.8.1 Automatic physiological signals normalization using ADLs

We estimated *normalization parameters* from ADLs, by modeling the relation between the physiological signals values during *lying* down resting, *walking at dif-*

Table 6.2: RMSE and (MAPE) for all clusters of activities evaluated in this work. Necklace refers to accelerometers only models, Necklace+Physio combines accelerometer, HR and respiration rate data, while Necklace+Physio Norm combines accelerometer, normalized HR and respiration rate data. Wristband refers to accelerometers only models, Wristband+Physio combines accelerometer and GSR data, while Wristband+Physio Norm combines accelerometer and normalized GSR data.

	Lying	Sedentary	HWBM	Walking	Biking	Running	Avg
Necklace	0.38 (20)	0.80 (38)	1.35 (34)	1.56 (22)	1.90 (21)	1.59 (12)	**1.26 (22)**
Necklace + Physio	0.34 (18)	0.62 (32)	1.01 (29)	1.47 (21)	1.72 (20)	1.51 (12)	**1.11 (20)**
Necklace + Physio Norm	0.33 (18)	0.51 (28)	0.77 (22)	0.98 (14)	1.16 (14)	1.24 (9)	**0.83 (15)**
Wristband	1.01 (32)	4.38 (136)	3.16 (70)	1.74 (24)	2.30 (27)	2.21 (16)	**2.47 (48)**
Wristband + Physio	0.55 (26)	0.78 (43)	1.73 (48)	1.53 (22)	2.12 (26)	1.80 (13)	**1.42 (28)**
Wristband + Physio Norm	0.45 (22)	0.57 (29)	1.62 (45)	1.42 (21)	1.79 (20)	1.54 (11)	**1.23 (22)**

ferent speeds and the physiological signals value during a *high intensity activity* ($X_{phy_{high}}$). $X_{phy_{high}}$ could be estimated with high accuracy for HR ($R^2 = 0.90$), while the relation between the measured and estimated $X_{phy_{high}}$ was weaker for GSR ($R^2 = 0.88$) and respiration ($R^2 = 0.72$). We speculate that these differences are mainly due to two factors: *a)* the tighter relation between HR and EE, due to the direct link between HR and oxygen intake, which makes HR a better predictor of EE compared to GSR and respiration rate. *b)* The higher responsiveness of HR, which is almost instantaneously affected by changes in activity type and intensity, while GSR changes were slower. However, all models were able to significantly improve EE estimation results compared to non-normalized signals.

RMSE for single EE estimation models using physiological data only was reduced by $60\%, 25\%$ and 18% for HR, GSR and respiration rate respectively, when compared to non-normalized models. Most importantly, all EE estimation models using normalization showed no differences when compared to models developed using individual calibration, confirming the feasibility of our normalization method. While single models were useful to determine the effectiveness of the physiological signals normalization, accelerometer data is required since the estimation of the normalization parameters relies on the user context (activity and walking speed), which is derived from accelerometer data.

6.8.2 Personalized activity-specific EE estimation

The proposed method reduced error in activity recognition, impact of misclassification on EE estimation (by reducing misclassification between active and inactive clusters) and EE estimation. While activity recognition is improved by only 2% when physiological signals were normalized (compared to non-normalized physiological signals), the impact of the error reduction on EE is larger. Activity misclassification of the Wristband is due to the fact that not only movement at the wrist is weakly related to EE, but also to activity type (high intensity of wrist movement can be detected even at rest, while e.g. *writing*). By combining accelerometer and physiological signals, the misclassification error between inactive and active clusters could be significantly reduced. Thus, avoiding high EE estimation errors due to the application of the wrong activity-specific model. For example figure 6.6, shows that *sedentary* activities misclassification as *biking* was reduced from 11% to 5%, while *biking* misclassification rates as *sedentary* were reduced from 14% to 4%. Misclassification rates are significantly further reduced when normalized physiological data was employed. Misclassification of *sedentary* activities as *biking* dropped to 0.4%, while misclassification of *biking* as *sedentary* dropped to 0%. These improvements are due to the fact that normalized physiological signals are more representative of the activity performed, while non-normalized physiological signals are more representative of the underlying physiological differences in different persons (e.g. level of CRF). Previous research underestimated the importance of physiological signals in activity type recognition, since multiple accelerometer were used (Tapia 2008). Single sensor estimation approaches, as used in this work, could improve user comfort over multi-devices solutions. When dealing with single sensor devices, physiological data can provide significant improvements, especially when normalized. Finally, we showed error reductions in EE estimation between 15 and 33%, compared to state of the art activity-specific EE models combining accelerometer and non-normalized physiological signals. Especially when the sensor is located where motion is weakly related to activity type and EE, combining accelerometers and normalized physiological signals showed the most substantial improvements.

We recognize limitations in our study. Even though we developed an algorithm to derive the *normalization parameters* automatically, during ADLs, we evaluated it using laboratory recordings only. We consider that the evaluation with lab data is a necessary first step, as during lab recordings sufficient reference measurements of EE could be acquired. In particular, out methodology allowed us to confirm performances of the individual estimators (activity, walking speed, normalization parameters, EE) during different PAs. Activities were chosen that are often occurring in free living situations (e.g. lying and walking).

6.8.3 Conclusion and further work

In this work, we introduced a methodology to normalize physiological signals using ADLs, in order to reduce inter-individual differences in physiological signals between individuals and improve EE estimation accuracy. We believe that our method is a significant step towards personalized physical activity monitoring, and to fully exploit the tight individual relation between physiological signals and EE. In this work, we confirmed that a relationship between physiological data during low intensity ADLs and the normalization parameters exists. As future work, we are currently investigating the practical applicability of the proposed methodology in free-living situations and on a bigger sample size, as well as the possibility to combine multiple sensors to further improve the estimate.

Part III: VO_2max estimation using wearable sensor data

7

Personalized cardiorespiratory fitness and energy expenditure estimation using hierarchical Bayesian models

M. Altini, P. Casale, J. Penders, O. Amft
Adapted from: Journal of Biomedical Informatics, 56:195-204, 2015.

Abstract

Accurate estimation of Energy Expenditure (EE) and cardiorespiratory fitness (CRF) is a key element in determining the causal relation between aspects of human behavior related to physical activity and health. In this paper we estimate CRF without requiring laboratory protocols and personalize energy expenditure (EE) estimation models that rely on heart rate data, using CRF. CRF influences the relation between heart rate and EE. Thus, EE estimation based on heart rate typically requires individual calibration. Our modeling technique relies on a hierarchical approach using Bayesian modeling for both CRF and EE estimation models. By including CRF level in a hierarchical Bayesian model, we avoid the need for individual calibration or explicit heart rate normalization since CRF accounts for the different relation between heart rate and EE in different individuals. Our method first estimates CRF level from heart rate during low intensity activities of daily living, showing that CRF can be determined without specific protocols. Reference VO_2max *and EE were collected on a sample of 32 participants with varying CRF level. CRF estimation error could be reduced up to 27.0% compared to other models. Secondly, we show that including CRF as a group level predictor in a hierarchical model for EE estimation accounts for the relation between CRF, heart rate and EE. Thus, reducing EE estimation error by 18.2% on average. Our results provide evidence that hierarchical modeling is a promising technique for generalized CRF estimation from activities of daily living and personalized EE estimation.*

7.1 Introduction

In the recent past, wearable sensing technologies have been used to objectively monitor human behavior, and started to provide unprecedented insights into the relation between physical activity and health. While energy expenditure (EE) is the most commonly used single metric to quantify physical activity, with many algorithms proposed in the recent past [118, 29, 9, 67], cardiorespiratory fitness (CRF) is not only an objective measure of habitual physical activity, but also a useful diagnostic and prognostic health indicator for patients in clinical settings, as well as healthy individuals [81].

Additionally, EE and CRF are tightly coupled when EE estimation is performed based on heart rate data acquired using wearable sensors. The inverse relation between heart rate and CRF is one of the main causes behind the need for individual calibration of heart rate monitors, since differences in CRF cause differences in heart rate but not in metabolic responses [104]. Thus, CRF estimation could both provide a relevant health marker and be used to personalize EE estimation models, improving estimation accuracy.

To date, the most commonly used measure for CRF level is the maximal oxygen uptake, or VO_2max. However, measures of VO_2max are rare in healthcare, due to safety concerns and laboratory infrastructure requirements. To tackle some limitations of VO_2max tests, *submaximal* test have been developed. Submaximal tests rely on the relation between heart rate and VO_2 at a certain exercise intensity, which is fixed by the strict exercise protocol that has to be executed [18, 48, 54]. Instead of performing a specific test that specifies exercise intensity at which heart rate is measured, we propose to use wearable sensor data to determine specific contexts (e.g. activity type and walking speed) and model the relation between heart rate in a specific context and CRF.

State of the art EE estimation models subdivide the estimation procedure into two steps. First, an activity is recognized. Secondly, an activity-specific regression model is applied to estimate EE [29, 118]. Recent work showed that including physiological data and normalizing heart rate can further improve results [9, 10]. Others, modeled the relation between EE and sensor data (e.g. accelerometer) while capturing commonalities across users of differing anthropometric characteristics [125, 126] using a hierarchical approach. Thus, structuring sensor data at the first level of a hierarchical structure, and anthropometric data at the second level of a hierarchical structure.

In this work, we hypothesized that using hierarchical Bayesian regression we could model both the influence of anthropometric characteristics and CRF level on accelerometer and heart rate data, and the variation in parameters depending on the performed activity, as in activity-specific models for EE estimation. Thus, the flexibility of a hierarchical regression framework was used to estimate CRF and effectively personalize EE estimation models without the need for explicit heart rate normalization. In particular, this paper provides the following contributions:

1. We propose a hierarchical Bayesian model to estimate CRF level from ac-

celerometer and heart rate data acquired using a single body-worn sensor during low intensity activities of daily living. Thus, the proposed model does not require specific laboratory tests or individual calibration. We show that low intensity activities of daily living (e.g. walking at 4 km/h) and heart rate data are sufficient to reduce CRF estimation errors by 27.0% compared to a model including anthropometric characteristics alone as predictors.

2. We extend previous work on EE estimation by proposing a hierarchical Bayesian model including non-nested group level parameters to simultaneously model the relation between activity type and EE, as well as between anthropometric characteristics, CRF and EE. Grouping by activity allows the model parameters to change as in activity-specific models. By including CRF among the group level parameters, we are able to account for the relation between CRF and heart rate and therefore personalize EE models. We show reductions in EE estimation error by 18.2% on average.

7.2 Related work

7.2.1 Maximal oxygen uptake

CRF is a well established and robust indicator of cardiovascular health and predictor of premature all cause mortality [26, 41]. The most commonly used measure for CRF level is VO_2max. VO_2max is the maximal capacity of the individual's body to transport and use oxygen (O_2) during exercise. Direct measurement of VO_2 using gas analysis during maximal exercise is regarded as the most precise method for determining VO_2max [124]. Despite the indubitable importance of CRF for health, measurements of $VO2$max in healthcare are rare, for different reasons. The test is time consuming, has to be performed by specialized personnel in a lab environment and expensive equipment is needed. The high motivation demand and exertion of subjects makes the test unfeasible in many patients groups [94].

7.2.2 Submaximal CRF estimation

To overcome these problems, many submaximal tests have been developed. Some are *non-exercise* CRF models, others are specific lab protocols performed while monitoring heart rate at predefined speeds (e.g. treadmill tests) or output powers (e.g. bike tests) [18, 48, 54], without requiring maximal effort. Several *non-exercise* models of CRF have been developed using easily accessible measures such as age, sex, self reported physical activity level, body composition [69, 73]. Results typically provide decent accuracy at the group level [93]. However significant limitations apply at the individual level, since each individual is assumed to be equal to group averaged characteristics. Limited accuracy at the individual level is a common problem when physiological variables are not measured. Most *submaximal* exercise tests rely on the relation between heart rate and VO_2 at a certain exercise intensity, which is fixed by the strict exercise protocol that has to be sustained. Submaximal exercise tests should be re-performed every time CRF needs to be assessed and often require laboratory infrastructure.

7.2.3 CRF estimation in free living

Both maximal and submaximal tests to estimate CRF are affected by important limitations. A more ideal solution, which possibly would be applicable to a larger population, is to estimate VO_2max during activities of daily living, without the need for a predefined exercise protocol. Towards this direction, Plasqui et al. [100] showed that a combination of average heart rate and activity level over a period of 7 days correlates significantly with VO_2max. However, by averaging over several days, the relation between average heart rate and activity counts depends on the amount of activity performed by the participants. Tonis et al. [120] explored different parameters to estimate CRF from heart rate and accelerometer data in laboratory settings. However, no models to extract these parameters in daily life (e.g activity type to detect walking or walking speed estimation models) are presented. In their work, VO_2max reference was not collected, but also estimated from walking data.

7.2.4 EE estimation

Recent work on EE estimation relying on wearable sensor data proposed activity-specific models as an improvement to previously used single or branched regression models [29, 9, 118]. Activity-specific EE estimation models consist of a two-step process, where first an activity is recognized, and then an EE estimation model is applied. Algorithms combining accelerometer and heart rate data consistently provided improvements compared to accelerometers alone [9, 118]. However, decomposing the EE estimation process into activity-specific sub-problems is not sufficient to take into account the different relation between heart rate and EE in different individuals. During moderate to vigorous physical activity, differences in heart rate between persons performing the same activity are mainly due to CRF. However, differences in CRF level do not cause different metabolic responses [104] (see Fig. 9.1). Thus, when estimating EE using heart rate data, individual calibration is necessary to deal with CRF-related differences between individuals [34]. For many practical applications personal calibration is not feasible since it would require every user to perform a suitable fitness test, and other personalization techniques would be preferable. We recently introduced a methodology to automatically normalize heart rate. By estimating a normalization parameter that describes heart rate at a certain workload during low intensity activities of daily living [10, 13, 15] we could personalize EE estimates. The methodology was based on the tight relation between CRF and heart rate at a certain workload, which is also the basis of sub-maximal CRF tests. In our previous work we required explicit heart rate normalization by estimating a normalization parameter representative of CRF, such as the heart rate while running at a certain intensity.

In this current work, we propose a novel model in which, instead of normalizing heart rate, we take the source of between-individual variability in heart rate, i.e. actual CRF into account. To this aim, we collected reference CRF as measured by a VO_2max test and developed a model for personalizing EE estimation without the need for explicit heart rate normalization. We hypothesized that CRF could account for the varying relation between heart rate and EE in different individuals by acting as a group level predictor in a hierarchical Bayesian model.

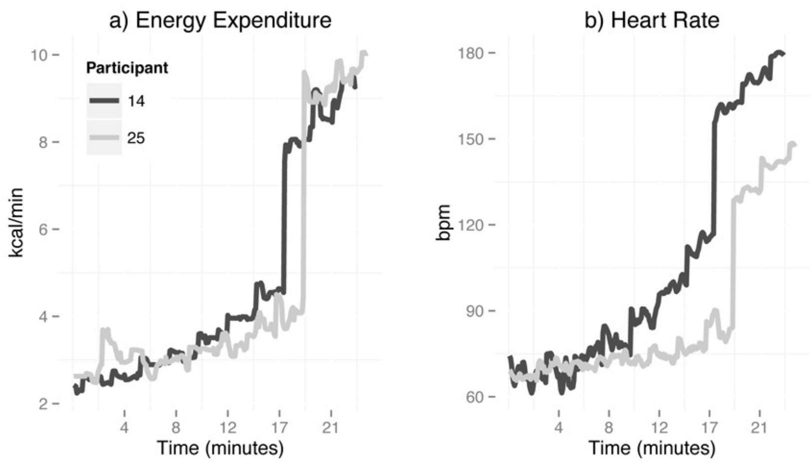

Figure 7.1: Relation between EE and heart rate in different participants during a sequence of different physical activities. a) Absolute EE levels are similar due to similar body weight. b) heart rate differs significantly between participants due to different CRF level (VO_2max participant 14 is 2104 ml/min, VO_2max participant 25 is 3130 ml/min).

7.2.4.1 Hierarchical models

Activity-specific EE models are typically implemented using linear regression models. Linear regression can be extended to capture commonalities across a population using a hierarchical linear model [61]. Hierarchical techniques use linear models at levels within (*individual* level) and across (*group* level) participants. In the remaining of this paper, we use the term *group level* parameters to indicate parameters at the second level of a hierarchical structure. These parameters are the ones influencing the relation between predictors at the first level of a hierarchical structure and the outcome variable. We refer to parameters at the first level of a hierarchical structure as *individual level* parameters [61]. These models were introduced in EE literature by Vathsangam et al. [126]. At one level the authors included participant specific parameters relating inertial sensor features to EE. At a second level they captured the inter-dependence of different person-specific parameters (e.g. anthropometric characteristics) using a (second) regression model. However [126], the authors limited their analysis to walking activities and accelerometer data, for EE estimation.

In this work, we hypothesized that hierarchical Bayesian models could be used to accurately model individual and group level differences in CRF level from wearable sensor data during activities of daily living. We expected that estimated CRF could be used to personalize heart rate-based EE estimations in order to improve the estimate accuracy. Additionally, we use the flexibility of a hierarchical regression framework to model both the influence of anthropometric characteristics and CRF level parameters on accelerometer and heart rate data, as well as the variation in parameters depending on the performed activity, as in activity-specific models.

7.3 Methods

In this section we describe our approach to CRF and EE estimation, as illustrated in Fig. 7.2. We use wearable sensor data, accelerometer X_{acc} and heart rate X_{hr}, together with anthropometric characteristics X_{ant} (e.g. height, body weight, etc.) as input to our models. CRF y_c is estimated from heart rate X_{hr} during low intensity activities of daily living, i.e. contexts s, simulated in the lab. For example, a context s can be walking at 4 or 6 km/h. Heart rate measured during a specific context is used together with anthropometric characteristics X_{ant} in a Bayesian regression model to estimate CRF y_c. Subsequently, we use the predicted CRF y_c as input for the second level of a hierarchical Bayesian model, to estimate EE y_{ee}. The hierarchical modeling accounts for variance in CRF between individuals and allows for more accurate EE estimation.

We introduce three hierarchical regression models, to estimate walking speed, CRF and EE, as shown in Fig. 9.3. Details on the notation and modeling technique are provided in Appendix A. We indicate group level predictors as U and individual level predictors as X.

Following a top down approach, we propose a hierarchical Bayesian model to estimate EE (see Fig. 9.3.c). We consider $i = 1, \ldots, n$ sensor data samples, $p = 1, \ldots, np$ participants and $a = 1, \ldots, T$ activities. Individual level parameters β_{pa} are influenced by both activity type a (which is the nature of activity-specific models) and the participants' anthropometric characteristics X_{ant} and CRF y_c, however the grouping by activity and by participant are non-nested:

$$y_{ee_i} \sim N(X_{ee_i}\beta_{i[pa]}, \sigma^2_{ee}), \tag{7.1}$$

$$i = 1, \ldots, n \quad a = 1, \ldots, T \quad p = 1, \ldots, np$$

$$\beta_{pa} \sim N(U_{ee_p}\,\gamma_{pa}, \Sigma_{pa}) \tag{7.2}$$

$$a = 1, \ldots, T \quad p = 1, \ldots, np$$

$$X_{ee_i} = [1, X_{acc_i}, X_{hr_i}] \in \mathbb{R}^{n \times (K+1)} \tag{7.3}$$

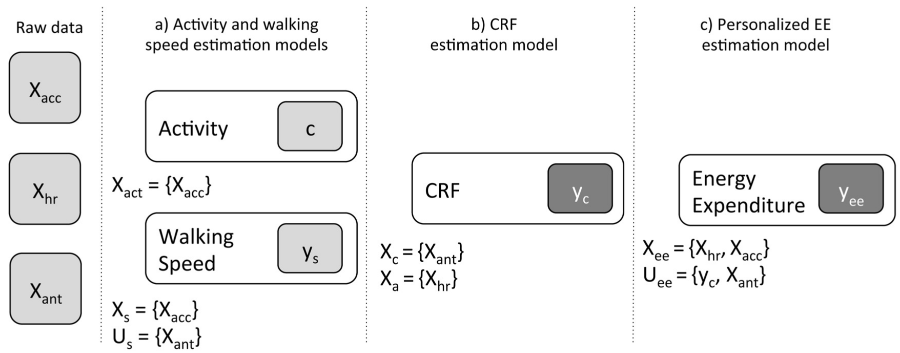

Figure 7.2: Block diagram of the proposed CRF and EE estimation approach. $a)$ Activity type and walking speed are estimated from sensor data of a wearable device (X_{acc} and X_{hr}). $b)$ CRF is estimated from heart rate during low intensity activities of daily living, such as walking, as derived from models $a)$, together with anthropometric characteristics X_{ant}. X_a consists in heart rate during predefined contexts (for example walking at 4 km/h), and therefore requires activity c and speed y_s information. $c)$ EE is derived by combining X_{acc}, X_{hr}, X_{ant} and CRF y_c in a hierarchical model, as shown in Fig. 9.3. Data flow is left to right. At each processing block, indicated by vertical dashed lines, we indicated which data streams are received from the previous processing blocks.

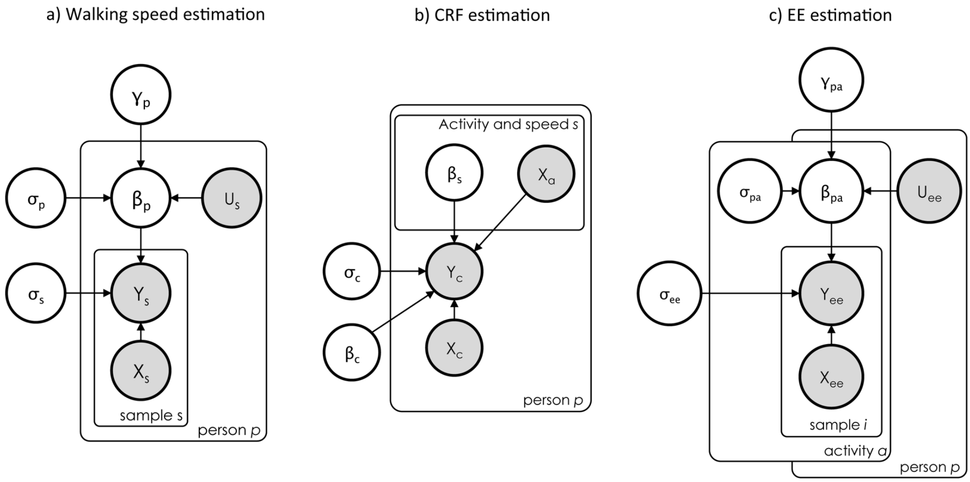

Figure 7.3: Proposed hierarchical models in plate notation. a) Walking speed estimation. Parameters β_p vary depending on the person's anthropometric characteristics U_s b) CRF estimation. Parameters β_s vary by activity and speed. c) EE estimation. Parameters β_{pa} are allowed to vary depending on the performed activity as well as on the persons' anthropometric characteristics and CRF U_{ee}. The two groupings are non-nested. Estimated CRF y_c from model b) is used as group level parameter U_c for model c). Hyperparameters are not shown for clarity.

$$i = 1, \ldots, n$$

$$U_{ee_p} = [1, X_{ant_p}, y_{c_p}] \in \mathbb{R}^{np\times(L+1)} \tag{7.4}$$

$$p = 1, \ldots, np$$

$$\gamma_{pa} \sim N(\mu_{\gamma_{pa}}, \sigma^2_{\gamma_{pa}}) \tag{7.5}$$

where the matrix X_{ee} is of dimension $n \times (K+1)$ and include K individual-level predictors such as heart rate X_{hr} and accelerometer features X_{acc}, over n data samples. U_{ee} is the matrix of dimension $np \times (L+1)$ and include L group level predictors controlling the individual level parameters β_{pa}. The predictors U_{ee} include anthropometric characteristics X_{ant} (e.g. *body weight*) and the estimated CRF y_c, for np participants. The hyperparameter matrix γ_{pa} is of dimension $(L+1)\times(K+1)\times T$, where T is the number of activities. Σ_{pa} is the $(K+1)\times(K+1)$ covariance matrix representing the variation of intercepts and slopes in the different groups. $\mu_{\gamma_{pa}}$ and $\sigma_{\gamma_{pa}}$ indicate hyperparameters for group level parameters γ_{pa}.

Estimated CRF y_c as included in the EE estimation model is shown in Fig. 9.3.b and consists in a hierarchical Bayesian model allowing only heart rate coefficients to vary by group, and can be described as:

$$y_{c_p} \sim N(X_{c_p}\beta_c + X_{a_p}\beta_{s[p]}, \sigma^2_c), \quad p = 1, \ldots, np \tag{7.6}$$

$$s = 1, \ldots, R \quad p = 1, \ldots, np$$

$$X_{c_p} = [1, X_{ant_p}] \in \mathbb{R}^{np \times (D+1)} \tag{7.7}$$

$$p = 1, \ldots, np$$

$$X_{a_p} = [X_{hra_p}] \in \mathbb{R}^{np \times 1} \tag{7.8}$$

$$p = 1, \ldots, np$$

$$\beta_s \sim N(\mu_{\beta_s}, \sigma^2_{\beta_s}) \tag{7.9}$$

where the matrix X_{c_p} of individual level attributes is of dimension $np \times (D + 1)$ (i.e. *body weight*, *height*, *age*, *sex*). The associated parameters β_c do not vary. Contexts s are a set of combined activity types and walking speeds (e.g. *walking at 4 km/h*, etc), which control the parameters β_s for the attributes X_a. X_a consists of heart rate during predefined contexts s (indicated as X_{hra_p}), and is of dimension $np \times 1$. μ_{β_s} and σ_{β_s} indicate hyperparameters for group level parameters β_s.

Activity type a is recognized from a set of T activities $A = a_1, \ldots, a_t$, using Support Vector Machines (SVM). Implementation details can be found in Sec. 9.5. Walking speed estimation y_s is shown in Fig. 9.3.a and consists of a hierarchical Bayesian model allowing accelerometer features X_{acc} to vary depending on anthropometric characteristics X_{ant}:

$$y_{s_i} \sim N(X_{s_i} \beta_{i[p]}, \sigma^2_s), \tag{7.10}$$

$$i = 1, \ldots, n \quad p = 1, \ldots, np$$

$$\beta_p \sim N(U_s \, \gamma_p, \Sigma_p) \tag{7.11}$$

$$p = 1, \ldots, np$$

$$X_s = [1, X_{acc}] \in \mathbb{R}^{n \times (K+1)} \tag{7.12}$$

$$i = 1, \ldots, n$$

$$U_s = [1, X_{ant}] \in \mathbb{R}^{np \times (L+1)} \tag{7.13}$$

$$p = 1, \ldots, np$$

$$\gamma_p \sim N(\mu_{\gamma_p}, \sigma^2_{\gamma_p}) \tag{7.14}$$

where the matrix X_s is of dimension $n \times (K + 1)$ and includes K individual-level accelerometer features X_{acc}, over n data samples. U_s is the matrix of dimension $np \times (L+1)$ and includes L group level predictors controlling the individual level parameters $\beta_{i[p]}$. The predictors U_s are the anthropometric characteristics X_{ant} such as *body weight* and *height*. The hyperparameter matrix γ_p is of dimension $(L+1) \times (K+1)$. Σ_p is the $(K+1) \times (K+1)$ covariance matrix representing the variation of intercepts and slopes in the different groups. μ_{γ_p} and σ_{γ_p} indicate hyperparameters for group level parameters γ_p.

7.4 Evaluation study

7.4.1 Participants and data acquisition

Participants were 32 self-reported healthy individuals. Characteristics are reported in Table 7.1. Written informed consent was obtained, and the study was approved by the ethics committee of Maastricht University. Participants were selected to have a wide range in physical activity levels and CRF. Measurements were obtained using an ECG Necklace, a low power wireless platform which was configured to acquire one lead ECG data at 256 Hz, and three-axial accelerometer data at 32 Hz. The ECG Necklace was worn on the chest, during all recordings, since the chest showed to be an optimal location for EE estimation in previous research comparing multiple on body sensor locations [14]. A *Continuous Wavelet Transform* based beat detection algorithm was used to extract R-R intervals from ECG data, which output was manually examined to correct for missed beats that might be caused by motion artifacts [102]. Participants were equipped with an indirect calorimeter consisting of a mouthpiece and nose clip. Expired air was continuously analyzed for O_2 consumption and CO_2 production (Oxycon-β), from which EE was derived [128].

Table 7.1: Participants characteristics, mean and standard deviation (SD)

Characteristic	Female Mean $\pm$ SD	Male Mean $\pm$ SD	All Mean $\pm$ SD
Number	17	15	32
Age (y)	24.6 $\pm$ 2.5	23.7 $\pm$ 1.6	24.2 $\pm$ 2.1
Height (cm)	167.1 $\pm$ 5.9	177.0 $\pm$ 6.3	171.8 $\pm$ 7.8
Weight (kg)	60.4 $\pm$ 6.8	72.5 $\pm$ 11.1	66.1 $\pm$ 10.8
BMI (kg/m^2)	21.6 $\pm$ 2.4	23.1 $\pm$ 3.5	22.4 $\pm$ 3.7
VO_2max (ml/min)	2534.2 $\pm$ 488.5	3518.6 $\pm$ 401.2	2995.6 $\pm$ 667.0

7.4.2 Experiment protocol

Participants reported at the lab on three separate days and after refraining from drinking (except for water), eating and smoking in the two hours before the recordings. Two laboratory protocols were performed. The first protocol included simulated activities of daily living performed while wearing a portable indirect calorimeter. Activities included: lying down resting, sitting, sitting writing, standing, cleaning a table, sweeping the floor, walking at different speeds (treadmill flat at 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6 km/h) and running at different speeds (treadmill flat at 8, 9, 10 km/h). The second protocol was a VO_2max test. VO_2max was determined during an incremental test on a cycle ergometer [78], thus providing reference data for CRF level, biking activity and EE while biking. Finally, anthropometric measurements including the participant's body weight, height and body fat were performed. Body fat was assessed using doubly labelled water [131]. Depending on lab and participant availability, the protocols were carried out in different days. Activities were carried out for a period of at least four minutes, always in the same order. Participants were allowed to rest between activities. Rest periods were normally between one and two minutes.

7.4.3 Statistics and performance measures

CRF estimation models were compared against models including anthropometric characteristics to describe individual variability. Our hierarchical EE estimation approach including estimated CRF as group level predictor was compared against two other estimation methods. First, we compared against state of the art activity-specific EE models including accelerometer and heart rate features but without CRF estimation. In literature, activity-specific EE models showed performance superior to other linear and non-linear EE estimation methods [9, 30] and therefore were selected as baseline for our proposed method. Secondly, we compared against hierarchical models including actual CRF (referred to as *CRF measured*) as a predictor. Hierarchical models including actual CRF serve as a lower boundary indicating the theoretical RMSE that is achievable.

Models were derived using data from all but one participants, and validated on the remaining one (leave-one-participant-out cross validation). The same training set, consisting of data from all participants but one, was used to build feature selection, activity recognition, walking speed estimation, CRF estimation and EE estimation models. The remaining data (from one participant) was used for validation. This procedure was repeated n (n = number of participants) times, and results were averaged. Performance of the activity recognition models was evaluated using the class-normalized accuracy $= \frac{1}{N_a}\sum_{c=1}^{N_a}\frac{recognized_a}{relevant_a}$, where N_a is the total number of classes, and $recognized_a$ and $relevant_a$ are the number of correctly identified and total instances for activity a, respectively. Results for walking speed are reported in terms of Root-mean-square error (RMSE) where the outcome variables was speed in km/h. Results for CRF and EE estimates are reported in terms of RMSE, mean absolute percentage error (MAPE) and explained variation (R^2), where the outcome variables were VO_2 in ml/min and EE in kcal/min respectively. Paired t-tests were used to compare RMSE between models. Significance was assessed at $\alpha < 0.05$.

7.5 Implementation

7.5.1 Pre-processing

The dataset considered for this work contains about 88.6 hours of annotated data collected from 32 participants, consisting of reference VO_2, VCO_2, three axial acceleration, ECG and VO_2max during laboratory recordings. A continuous wavelet transform based beat detection algorithm was used to extract RR intervals from ECG data, which output was manually examined to correct for missed beats. Breath-by-breath data acquired from the indirect calorimeter was resampled at 0.2 Hz. EE was calculated from O_2 consumption and CO_2 production using Weir's equation [128]. The first 1 or 2 minutes of each activity were discarded to remove non-steady-state data. Activities were grouped into six clusters to be used for activity classification. The six clusters were *lying* (lying down), *sedentary* (sitting, sitting writing, standing), *dynamic* (cleaning the table, sweeping the floor), *walking* (treadmill flat at different speeds), *biking* (cycle ergometer) and *running* (treadmill flat at different speeds).

7.5.2 Features extraction and selection

Features extracted from the sensors' raw data were used to derive all models. Activity recognition was performed to classify the six activity clusters previously introduced. Accelerometer data from the three axes were segmented in 5 s windows, band-pass filtered between 0.1 and 10 Hz, to isolate the dynamic component caused by body motion, and low-pass filtered at 1 Hz, to isolate the static component, due to gravity. Feature selection for activity type recognition was based on mutual information [22] and features were derived and selected from our previous work [9], using a different dataset. The complete feature set can be found in [9]. Selected features were: *mean of the absolute signal, inter-quartile range, median, variance, standard deviation, main frequency peak (i.e. mode of the frequency spectra), low and high frequency band signal power*. Heart rate was extracted from ECG data over 15 seconds windows. Anthropometric characteristics (*body weight*, *height*, *age*, and *sex*) were included in walking speed, CRF and EE estimation models.

7.5.3 Activity recognition

We implemented an activity recognition algorithm to classify the following clusters of activities: *lying*, *sedentary*, *dynamic*, *walking*, *running* and *biking*. Given the promising results in past research on activity recognition [9], we selected SVM as classifier. For the SVM, we used a gaussian radial basis kernel (C = 1).

7.5.4 Hierarchical Bayesian regression models

Hierarchical Bayesian models introduced in Sec. 7.3 were implemented using R and JAGS. Posterior estimations were performed by Gibbs sampling with 3 chains and 10000 iterations. The first 500 iterations were discarded (burn-in period). Anthropometric characteristics were *height* for the walking speed model, *height*, *weight*, *age* and *sex* for the CRF estimation model and *weight* for the EE model. Additionally, EE models included CRF as group level parameter. Individual level features were accelerometer only for walking speed estimation models, accelerometer and heart rate for EE estimation models and heart rate features for CRF estimation models. Prior distributions for all parameters and hyperparameters were non-informative uniform distributions with $\mu = 0$ and $\sigma = 100$.

7.6 Results

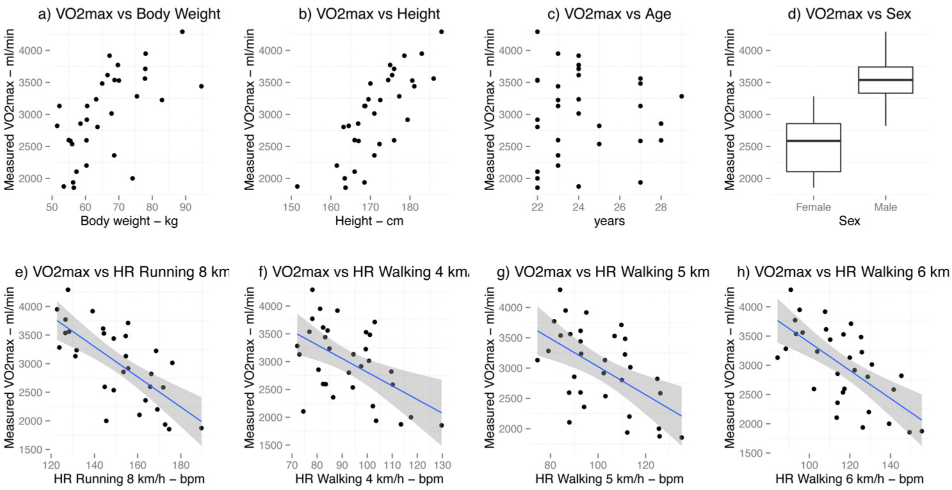

Figure 7.4: $a)$ to $d)$ relation between anthropometric characteristics and VO_2max. $e)$ to $h)$ relation between heart rate and VO_2max for different activities. Regression line and 95% confidence intervals highlighting the inverse relation between heart rate at different activities intensities and VO_2max are shown in plots $e)$ to $h)$.

Subject-independent class-normalized accuracy of the SVM was 92.7%. More specifically, the accuracy was 95.4% for *lying*, 95.2% for *sedentary*, 81.9% for *dynamic*, 96.3% for *walking*, 87.5% for *biking* and 99.7% for *running*. Walking speed estimation RMSE was 0.53 km/h.

Fig. 7.4 shows the relation between anthropometric characteristics and VO_2max, as well as the relation between heart rate and VO_2max for different activities. Correlation between heart rate and VO_2max was highest for running activities ($r = -0.71$), as shown in Fig. 7.4.e. Correlation increased between $r = -0.52$ and $r = -0.66$ for increases in low intensity activities of daily living, e.g. for walking between 4 to 6 km/h, regardless of anthropometric characteristics.

Fig. 7.5 shows results of the CRF estimation model for three conditions. As a CRF estimation baseline we considered anthropometric characteristics (model referred to as *Ant* in Fig. 7.5) as predictors, resulting in RMSE of 382.3 ml/min ($R^2 = 0.56$). RMSE was reduced to 279.5 ml/min (26.9% error reduction, $p = 0.02 < \alpha$, $R^2 = 0.73$) and 279.2 ml/min (27.0% error reduction, $p = 0.02 < \alpha$, $R^2 =$

0.74) when including heart rate while walking at 4 km/h and 6 km/h respectively as predictors. Detailed results for men and women are shown in Table 7.2. While error is relatively higher for women, differences are not significant ($p = 0.25 > \alpha$ for *Ant*, $p = 0.73 > \alpha$ for *Walk 4 km/h*, $p = 0.30 > \alpha$ for *Walk 6 km/h*).

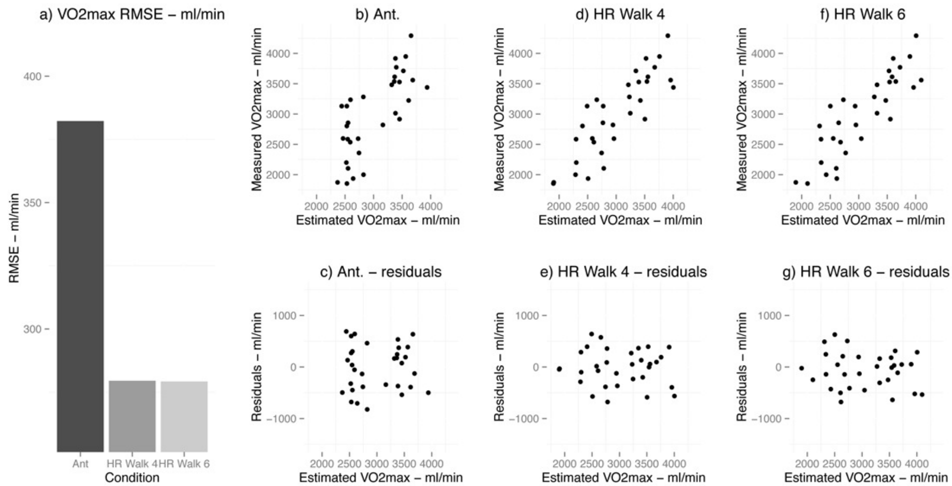

Figure 7.5: a) VO_2max estimation accuracy for different models over all participants. *Ant* does not include heart rate data but anthropometric characteristics only, *HR Walk 4* and *HR Walk 6* include heart rate while walking at 4 km/h and 6 km/h respectively. b-g) Scatterplots and residuals plot for measured and estimated VO_2max according to the three models of plot a).

Table 7.2: CRF estimation results.

CRF model	**Sex**	**RMSE ml/min**	**MAPE %**
Ant	all	382.3	14.0
	male	336.5	9.7
	female	422.7	17.8
Walk 4 km/h	all	279.5	9.8
	male	266.3	7.7
	female	291.2	11.7
Walk 6 km/h	all	279.2	10.2
	male	238.6	7.1
	female	315.1	12.9

EE estimation results are shown in Fig. 7.6. For an EE estimation baseline we considered for this analysis state of the art activity-specific EE estimation models.

Activity-specific EE estimation models (*no CRF*) included accelerometer and heart rate data as predictors and resulted in RMSE of 0.88 kcal/min ($R^2 = 0.94$). Additionally, we compared results obtained with the proposed hierarchical model (*CRF estimated*) to the theoretical case where actual CRF is available, instead of being estimated by our architecture (*CRF measured*). RMSE was reduced from the *no CRF* condition to 0.72 kcal/min (18.2% error reduction, $p = 0.003 < \alpha$, $R^2 = 0.95$) for *CRF estimated* and to 0.69 kcal/min (21.8% error reduction, $p = 0.002 < \alpha$, $R^2 = 0.96$) for *CRF measured*. In Table 7.3 we provide detailed results for moderate to vigorous activities only, since personalizing the relation between heart rate and EE during sedentary activities is not beneficial [13]. When including estimated CRF (Fig. 7.6, *no CRF* vs *CRF estimated*), EE RMSE was reduced from 0.61 kcal/min to 0.56 kcal/min for *dynamic* (8.9% error reduction), from 0.60 kcal/min to 0.55 kcal/min for *walking* (8.2% error reduction), from 2.18 kcal/min to 1.62 kcal/min for *biking* (25.5% error reduction) and from 1.36 kcal/min to 1.11 kcal/min for *running* (18.4% error reduction).

Table 7.3: EE estimation results. Activity *All* includes lying, sedentary, dynamic, walking, biking and running activities. Reference EE is shown as mean ± standard deviation and was collected by indirect calorimeter. EE and RMSE are reported in kcal/min.

EE model	Activity	EE	RMSE	MAPE %
No CRF	all	4.98 ± 4.14	0.88	18.4
	dynamic	2.64 ± 0.71	0.61	23.7
	walking	3.96 ± 0.59	0.60	13.6
	biking	10.50 ± 2.38	2.18	22.1
	running	10.35 ± 1.85	1.36	13.2
CRF estimated	all	4.98 ± 4.14	0.72	16.2
	dynamic	2.64 ± 0.71	0.56	21.2
	walking	3.96 ± 0.59	0.55	12.0
	biking	10.50 ± 2.38	1.62	15.6
	running	10.35 ± 1.85	1.11	10.6
CRF measured	all	4.98 ± 4.14	0.69	15.7
	dynamic	2.64 ± 0.71	0.54	20.6
	walking	3.96 ± 0.59	0.52	11.3
	biking	10.50 ± 2.38	1.48	14.3
	running	10.35 ± 1.85	1.10	10.6

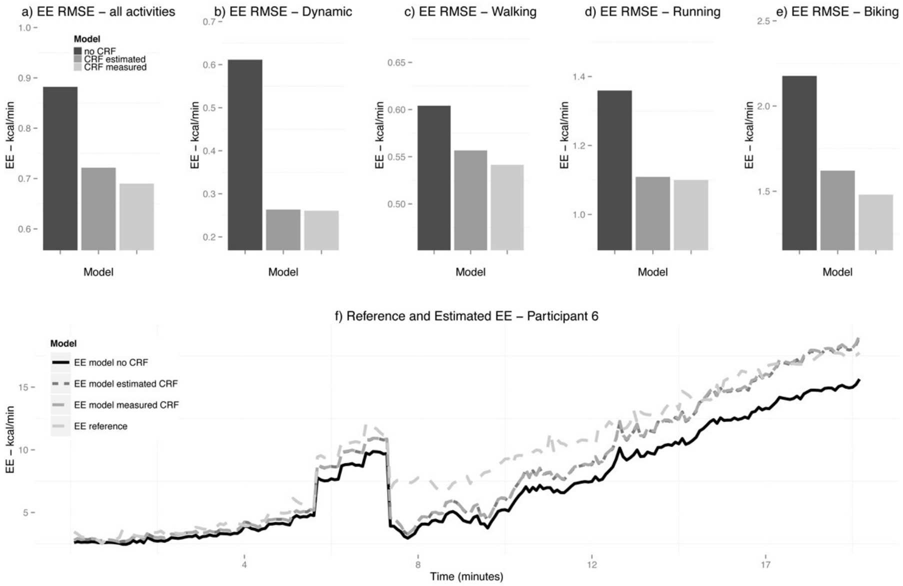

Figure 7.6: EE estimation RMSE for a) all activities averaged, b) dynamic, c) walking d) running and e) biking. Three models are compared, *no CRF*, indicating state of the art activity-specific EE models, *CRF estimated* indicating the proposed hierarchical Bayesian approach, where estimated CRF is used as group level parameter, and *CRF measured* which consists in the same model as *CRF estimated*, but including actual VO_2max as measured in the lab. *CRF measured* serves as a lower limit to the RMSE achievable by the proposed approach. f) Effect of CRF in EE estimation for an unfit participant. Without CRF, EE was underestimated, due to the higher heart rate. Including CRF increased the EE estimate, therefore reducing RMSE. No difference appeared between EE estimation models using estimated or measured CRF for this participant. Reference EE is shown in light gray.

7.7 Discussion

In this work, we demonstrated that hierarchical Bayesian regression could be used to accurately model individual and group level differences in CRF estimated from heart rate data during low intensity activities of daily living. We also validated our hypothesis that such estimated CRF could be used to personalize heart rate-based EE estimation models in order to improve the estimate accuracy for different activities. We adopted hierarchical Bayesian models as a powerful and flexible extension to conventional regression frameworks, structuring our models into groups which are both nested and non-nested.

To estimate CRF, we relied on the relation between CRF and heart rate at a certain submaximal intensity (e.g. while walking). Heart rate parameters were allowed to vary by activity type and speed, in order to let the proposed CRF model provide estimates without constraining the participant in performing specific activities or walking at predefined speeds, as normally done in submaximal laboratory tests. By using a hierarchical approach where parameters vary based on the activities performed by the participant, we also allow the CRF estimation model to use different parameters based on the participant activities, thus potentially increasing accuracy when more intense activities are performed. We chose walking speeds of 4 to 6 km/h for our analysis, since speeds close to this range were often reported as the average walking speeds in healthy individuals (5.3 km/h in [36] and 5 ± 0.8 km/h in [89]). We analyzed the impact of different features on CRF estimation, such as anthropometric characteristics, and the relation between heart rate while walking at different speeds. Anthropometric characteristics alone were shown to estimate CRF with in past research [69, 73]. Our models confirm these findings, due to the high correlation between VO_2max and most anthropometric characteristics, such as *body weight*, *height* and *gender* (see Fig. 7.4). However, only when including in the models physiological data such as the heart rate, differences between participants with similar anthropometric characteristics can be estimated. By including heart rate data while walking at 4 km/h to 6 km/h we could reduce RMSE up to 27.0%. Since CRF is a strong and independent predictor of all-cause and cardiovascular mortality, the proposed CRF estimation model could be used to provide accurate information about an individual's health without the need for laboratory infrastructure or specific tests.

As a second contribution, we developed a two level hierarchical Bayesian model, where accelerometer and heart rate parameters were allowed to vary by activity-type, as in activity-specific EE models, and by anthropometric characteristics as well as CRF level. Previous work by our group [10, 13] as well as others [118] showed that normalizing heart rate using a normalization parameter representative of CRF, such as the heart rate at a certain workload, could significantly reduce inter-person differences in heart rate and consequently improve EE estimation accuracy. The proposed hierarchical structure goes to the root of the problem, including estimated CRF level as a group level parameter able to control the relation between heart rate and EE. Since CRF is estimated from activities of daily living of varying intensity, no predefined test or laboratory calibration is necessary in order to improve EE estimation models. EE estimation RMSE was reduced by 8.9%, 8.2%, 25.5% and 18.4% for dynamic, walking, biking and running activities respectively. In our models, we excluded sedentary activities. In previous work heart rate during sedentary behavior was not found to be beneficial in estimating EE. During sedentary behavior, heart rate is affected by other factors such as stress, emotions, etc., and is typically weakly correlated with EE, and therefore often omitted for EE estimation [9, 43]. While RMSE for biking and running is relatively high compared to other activities, larger errors are expected for intense activities. Nevertheless, we believe that the RMSE reductions compared to current state of the art methods (up to 25%) are practically relevant, especially as the proposed method does not require laborious individual calibration. Moreover, the estimation performance obtained in this work is close to the theoretical performance estimate using actual CRF level data as shown in Table 7.3. Thus, measuring heart rate in low intensity activities of daily living is sufficient to estimate CRF at sufficient accuracy to obtain optimal EE estimation results.

Personalizing a system goes beyond the inclusion of the individual's anthropometric characteristics in CRF or EE models, as shown by the increased accuracy of the proposed models. While our CRF estimation model could be applied to a wide population and provide feedback on health status, we expect that our EE normalization approach will be most useful for individuals having a moderately active lifestyle. Sports training devices, where users and trainers are interested in accurate EE estimation under moderate to vigorous workloads, could benefit from inclusion of CRF in the EE estimation models. Additionally, less active individuals willing to take up a more active lifestyle, or undergoing a physical activity intervention targeted in modifying behavior to increase level of activity, would also benefit. As a matter of fact, in the latter case CRF takes even a bigger role, since it typically changes faster in the transition from inactive to active lifestyle, and failing to capture these changes would result in higher errors in EE estimation. New opportunities for applications targeted at inducing behavioral change by creating a feedback loop involving objectively measured physical activity level and EE, as well as change in CRF and associated reduced risk of disease, could be developed building up on the proposed approach.

We recognize limitations in our study. Even though we developed an algorithm able to derive CRF during regular activities, by combining walking heart rate data with the subjects anthropometric characteristics, we tested it using laboratory recordings only. We consider that the evaluation with lab data is a necessary first step, which can be sufficiently covered with reference measurements of CRF and EE. We proposed activity recognition and walking speed estimation models to detect activity type and walking speed such that the proposed model could be deployed in free living. Some activities (e.g. dynamic and biking) were recognized with suboptimal accuracy, due to sensor positioning and high variability in movement involved, for example, during household activities. Nevertheless, activity recognition performance for walking activities used by our models was sufficiently high to obtain useful EE estimation performance. We consider these results promising for free-living deployment in further research. Additionally, while our participants population included a wide range of weight, height, BMI and was balanced between male and female, their limited age range prevents us from generalizing the results to other age groups. However, our CRF models provide RMSE comparable with ordinary submaximal tests [111] without requiring specific exercises or individual calibration. Another point of attention is the difference in accuracy of our CRF estimation model in men and women. The slightly higher error for women might be due to a combination of factors. For example, the higher VO_2max standard deviation suggests higher variability in the female population. Adding explanatory variables such as body fat, which is known to have an important role in VO_2max estimation [100] might reduce this error. However, our goal was to use basic anthropometrics that can be easily acquired without laboratory tests. Therefore we limited our analysis to body weight. However, given the small difference in RMSE for EE estimation models using either CRF estimated by our procedure (CRF estimated) or actual VO_2max (CRF measured), the higher error found for CRF estimation in women does not seem to negatively affect personalization of EE estimation models.

In this work, CRF estimation was used to model the relation between heart rate and EE in participants of different fitness level, effectively reducing EE estimation error during moderate to vigorous physical activities. No intense activities or laboratory tests were used for CRF estimation and EE personalization. Instead, heart rate during low intensity activities of daily living was used as a predictor in our models, which provides for the practical applicability of the proposed method. Additionally, we used only simple anthropometrics data, excluding body fat, to allow for the development of models which do not require parameters acquired under laboratory conditions. Our methodology could be applied to other problems in which the relation between physiological parameters (e.g. heart rate, galvanic skin response, respiration, etc.) and an outcome variable (e.g. energy expenditure, mental stress, disease progression, etc.) varies between individuals. By modeling the source of variation, in our case CRF, at the second level of a hierarchical structure, the relation between physiological data and the outcome variable is modeled. Consequently, no explicit normalization is needed that would require individual calibration.

Acknowledgment

The authors would like to thank Guy Plasqui, Giuseppina Schiavone, Gabrielle ten Velde and Stefan Camps for their support during data collection.

Appendix A

In this section we clarify the mathematical notation used in this manuscript. We adopted the notation of [61]. Each sample, is indicated by an index. In the following equations we will use i to indicate the index. The classical linear regression model where the predicted variable is indicated by y_i and the array of K predictors is indicated by X_i, can be written in mathematical form as:

$$y_i = X_{i1}\beta_1 + \dots X_{ik}\beta_k + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, n \tag{7.15}$$

X_{i1} is the constant term, while X_{i2} to X_{ik} are features, for example accelerometer or heart rate data. We assume independent normal distribution with mean 0 and standard deviation σ for ϵ. Equation 1 can be written in compact form as:

$$y_i = X_i\beta + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, n \tag{7.16}$$

Equivalently to equation 2, we can express the relation between predictors and predicted variables as:

$$y_i \sim N(X_i\beta, \sigma^2), \quad i = 1, \dots, n \tag{7.17}$$

We adopted the latter notation for simplicity and reduced verbosity, especially when multiple parameters and levels are included in the models.

Hierarchical models are a generalization of linear regression models such as the one described in equation 3 in which parameters act at two levels. We use the term *group level* parameters to indicate parameters at the second level of a hierarchical structure. These parameters are the ones influencing the relation between predictors X and the outcome variable y. In the context of hierarchical modeling, parameters β are indicated as *individual level* parameters [61]. Individual level parameters β, (i.e. the slopes and intercepts) are allowed to vary by group. Thus, additionally to the variables already introduced, we introduce the group index j and represent group membership as $j[i]$. We also introduce group level parameters as γ. Thus, we define a hierarchical model in which parameters β vary by group as:

$$y_i \sim N(X_i \beta_{j[i]}, \sigma^2), \quad i = 1, \ldots, n \tag{7.18}$$

In 4, individual level parameters β vary depending on the group j. If there are no group level predictors, β acts similar to indicator variables in standard regression. This is the case for example of our activity-specific EE models, where different coefficients are derived for each activity class, however there is no group level predictors. Individual level parameters β in this case can be expressed as:

$$\beta_j \sim N(\mu_\beta, \sigma_\beta^2), \quad j = 1, \ldots, J \tag{7.19}$$

Where μ_β and σ_β^2 are hyperparameters. Alternatively, parameters β can also be estimated by higher level regression models, including group level parameters γ and a set L of group level predictors U. This is the case of EE estimation models where anthropometric characteristics such as body weight, height, etc., are used as group level predictors U. The notation used in this case is the following:

$$\beta_j \sim N(U_j \gamma, \sigma_\beta^2), \quad j = 1, \ldots, J \tag{7.20}$$

$$\gamma \sim N(\mu_\gamma, \sigma_\gamma^2), \quad j = 1, \ldots, np \tag{7.21}$$

Where γ is of dimension $K + 1 \times L + 1$, μ_γ and σ_γ^2 are hyperparameters.

Part IV: Personalized EE estimation and VO_2max estimation in free-living

8

Personalization of energy expenditure estimation in free living using topic models

M. Altini, P. Casale, J. Penders, O. Amft
Adapted from: IEEE Journal of Biomedical and Health Informatics, 19.5 (2015): 1577-1586.
Featured as research highlight on JBHI's homepage.

Abstract

We introduce an approach to personalize energy expenditure (EE) estimates in free living. First we use Topic Models (TM) to discover activity composites from recognized activity primitives and stay regions in daily living data. Subsequently, we determine activity composites that are relevant to contextualize heart rate (HR). Activity composites were ranked and analyzed to optimize the correlation to HR normalization parameters. Finally, individual-specific HR normalization parameters were used to normalize HR. Normalized HR was then included in activity-specific regression models to estimate EE. Our HR normalization minimizes the effect of individual fitness differences from entering in EE regression models. By estimating HR normalization parameters in free living, our approach avoids dedicated individual calibration or laboratory tests. In a combined free-living and laboratory study dataset, including 34 *healthy volunteers, we show that HR normalization in 14-day free living data improves accuracy compared to no normalization and normalization based on activity primitives only (*29.4% *and* 19.8% *error reduction against lab reference). Based on acceleration and HR, both recorded from a necklace, and GPS acquired from a smartphone, EE estimation error was reduced by* 10.7% *in a leave-one-participant-out analysis.*

8.1 Introduction

Wearable technology can provide novel insights on the relation of physical activity (PA) and health [108]. Energy expenditure (EE) is the most common parameter used to quantify PA [125], and is typically estimated using acceleration and heart rate (HR) sensors [43, 52]. Acceleration reflects a relation between motion and EE while HR shows a strong correlation with EE via the relation of EE and oxygen consumption. State-of-the-art EE estimation methods first classify user activity and subsequently apply activity-specific regression equations, to estimate EE [29, 118, 14]. Using HR in activity-specific regression equations showed consistent improvements in EE estimation compared to using acceleration only [10, 34]. However, HR during an activity is specific to a person since it depends on the individual's cardiorespiratory fitness (CRF) level [104]. To derive a reliable EE estimate, it is therefore necessary to normalize HR according to an individual's fitness. In turn, the normalized HR could serve as independent variable in EE regression models. Normalizing HR requires information on the individuals' fitness level, as fitness and HR are tightly related for a given workload [18]. Thus, in our previous work we predicted a surrogate of fitness, i.e. the HR while running at 9 km/h, and used it as HR normalization parameter to reduce EE estimation error [10]. As a proof of concept for HR normalization which does not require intense activities to be performed in laboratory settings, we estimated the HR while running at 9 km/h from the HR during low intensity activities. In particular, we defined a regression model using as predictors the HR while walking at a certain speed. However, our validation was performed in laboratory settings.

HR interpretation in free living is more complex. While for an individual any specific lab-performed activity may show little variation in HR, HR in free living is likely depending on context. The presence of various daily life stressors requires a novel estimation approach compared to laboratory studies. In particular, we assume that HR in free living is not only affected by activity primitives such as walking, but by a combination of activity primitives and more abstract activity composites such as social interactions, doing sport, etc. Thus, to exploit HR normalization for EE estimation in free living, activities must be recognized and interpreted according to the situation in which they were performed.

In this work, we present a method to derive HR normalization parameters during free living and personalize population based EE estimation models accordingly. In particular, our contribution is three-fold:

1. We define HR normalization parameters as surrogates of fitness levels estimated by contextualized HR. We contextualize HR in free living with a combination of activity primitives, activity composites and walking speeds. We use HR normalization parameters to normalize HR and estimate EE more accurately at the individual level.

2. We present a framework to discover activity composites in free living, and determine which activity composites are more suitable for HR normaliza-

tion. To discover activity composites we first utilize topic models (TM). Secondly, we determine *relevant activity composites* by ranking activity composites and analyzing the relation between ranked activity composites and HR normalization parameters across individuals.

3. We evaluate our approach in a combined free-living and laboratory study, including 34 participants. A laboratory protocol was used to obtain reference data for activity primitives and HR normalization. A 14-day free-living protocol was used to evaluate the estimation performance for HR normalization and personalization of EE estimation, yielding a 10.7% error reduction in EE estimation.

8.2 Related work

Accelerometer and HR monitors are the most commonly used devices for EE estimation [43, 52]. The latest EE estimation algorithms extend approaches based on simple linear regression models by splitting the estimation process into two phases. First, an activity is recognized. Secondly, an activity-specific regression model is used to predict EE [29, 118, 14]. Including HR data in the activity-specific linear models showed consistent improvements in EE estimation accuracy compared to algorithms using accelerometer only data [10, 34]. However, breaking down the EE estimation process into activity-specific sub-problems is not sufficient to take into account the different relation between HR and EE in different individuals [10]. Fig. 9.1 shows how participants with similar body weight consume similar amounts of energy. However, the different CRF level results in very different HR, but no difference in metabolic responses [104]. Thus, estimating EE based on HR results in under and overestimations [10, 13].

8.2.1 Personalized EE estimation

HR showed higher correlation with EE compared to accelerometer data [13]. However, subject-independent models including HR performed sub-optimally, confirming the need for individual calibration [13]. Individual calibration limits practical applicability, since the individual relation between HR and EE needs to be determined for the algorithm to be accurate. To the best of our knowledge, the only attempt to automatically normalize HR without requiring individual calibration was reported by our group. In [10], we introduced an approach to normalize HR by estimating a *HR normalization parameter*. A regression model including HR measured during activities of daily living simulated in the lab (e.g. *walking*) was used to estimate HR during intense exercise, such as running at 9 km/h. The estimated HR was used as the *HR normalization parameter*. While EE estimation error was reduced by the proposed methodology, we used laboratory recordings only to build our models. Supervised recordings allowed us to acquire data free of artifacts due to other daily life stressors, which was a necessary first step to prove the effectiveness of our approach. However, the presence of a multitude of stressors in free living urges for a different solution.

8.2.2 Context recognition

Our assumption is that physiological data, for example HR, in free living settings is not only affected by activity primitives, but by both activity primitives and activity composites. Incorporating contextual information beyond activity primitives

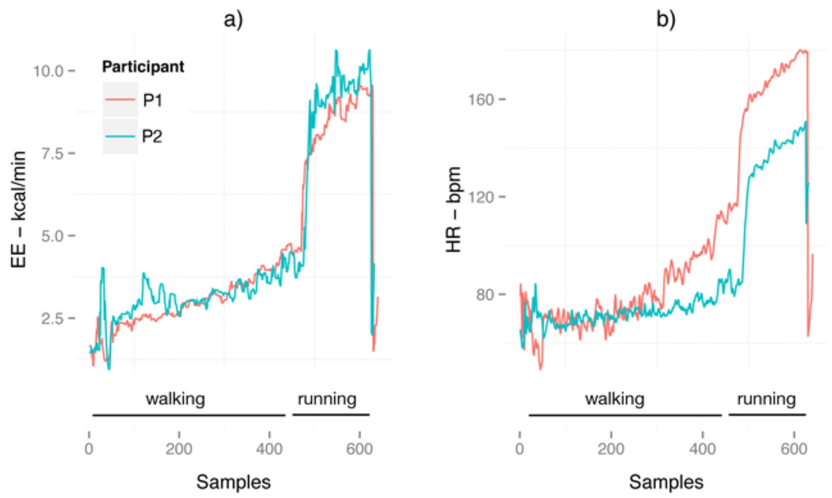

Figure 8.1: Relation between EE and HR in two participants during walking and running activity primitives. a) Absolute EE levels are similar due to similar body weight. b) HR differs between participants due to different CRF level (VO_2max participant 1 is 2104 ml/min, VO_2max participant 2 is 3130 ml/min). Thus, EE estimation based on HR would cause large individual error.

could potentially improve interpretation of HR or other physiological data in free living. Fig. 8.2 shows HR during activity primitives and activity composites performed in free living by one participant. HR during the same activity primitives changes depending on the activity composites. For example, HR during *social interactions* (plot *b*) is higher than during *work* (plot *a*) for both *sedentary* and *walking* activity primitives. Variations in HR can be noticed in different activity composites, and motivate the need for additional contextual information when interpreting HR data. Activities are often thought of in a hierarchical manner, starting from low level activity primitives, and building up to more complex activity composites [72]. Activity primitives are typically considered as a set of atomic activities that can be determined on a short time window [118], directly from low level raw sensor data. Atomic activities can be obtained using supervised machine learning methods, across a wide population. An example of activity primitives can be a set of postures and locomotion activities, such as: *lying down, sedentary, dynamic, walking, biking* and *running*, as adopted in previous research [29, 14]. On the contrary, higher level contextual information, such as activity composites, can benefit from a different recognition approach. Activity composites (e.g. social interactions, commuting, etc.) are personal and need unsupervised methods able to discover different patterns in each individual, depending on their behavior. A possible solution is the use of TMs. TMs were initially introduced by the text mining community, to discover topics from corpus of documents, starting from words [27]. For activity recognition, the same concept was applied to discover activity composites from activity primitives [72]. Recent work investigated the impact of multiple latent Dirichlet allocation (LDA) parameters for activity composites discovery, showing promising results [112]. In this work, we identified activity composites that are representative of HR normalization parameters in a unsupervised manner. To this aim, we introduced the concept of *relevant activity composites*.

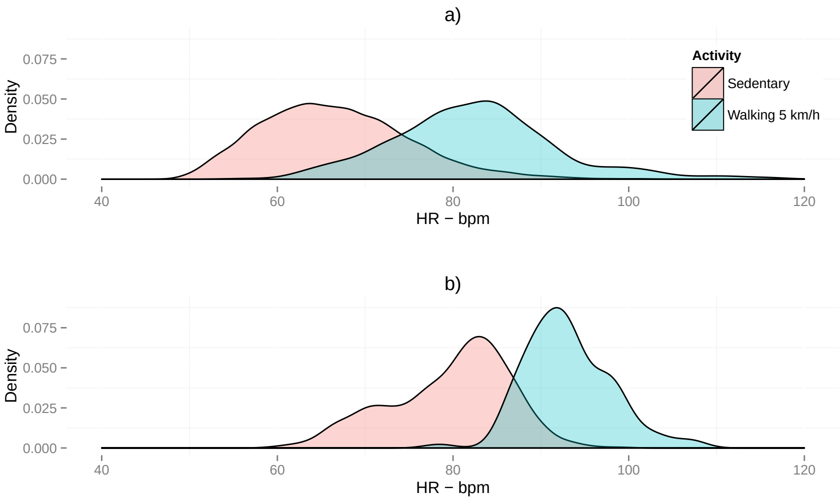

Figure 8.2: HR during activity primitives and activity composites performed in free living by one participant. Activity composites were manually annotated by the participant. HR during the same activity primitives changes substantially depending on the activity composites: a) work, b) social interactions.

8.3 Methods

We include HR in activity-specific EE estimation models after being normalized by the *HR normalization parameter*, HR_{np}. HR_{np} was predicted from HR while walking at a recognized speed, only during *relevant activity composites*. *Relevant* *activity composites* are activity composites in which HR is representative of HR_{np}, and were derived during training phase. We first utilized topic models (TM) to derive activity composites. Then, we determined *relevant activity composites* by ranking activity composites and analyzing the relation between ranked activity composites and HR normalization parameters across individuals, as described in Sec. 8.3.2. Following a top down approach, EE was estimated by activity-specific models (see Fig. 8.3). For each activity primitive c_i, a regression model is defined:

$$\begin{aligned} C = \{c_1, \ldots, c_{cn}\}, \quad \forall c_i \in C, \\ \exists \quad y_{act_i} = X_{act_i}\beta_{act_i} + \epsilon \\ X_{act_i} = \{X_{acc_i}, X_{ant_i}, X_{hr_i}\} \end{aligned} \tag{8.1}$$

where we assumed cn activity primitives C, recognized by a combined Support Vector Machine (SVM) classifier and Hidden Markov Models (HMM). Input for the SVM classifier are accelerometer features X_{acc}. The HMM is used to smooth transitions over the SVM output by defining the hidden states as the actual activity primitives c_i. For an activity primitive c_i, y_{act_i} is the dependent variable, the vector of target EE values, β is the vector of regression coefficients, and X_{act_i} is the vector of input features. Features X_{act_i} used in the activity-specific regression models can be grouped into accelerometer features X_{acc_i}, anthropometric characteristics X_{ant_i}, and normalized HR, X_{hr_i}, as shown in Fig. 8.3.

8.3.1 HR normalization parameter estimation

Normalized HR was obtained as shown in Fig. 8.3 by dividing HR by person-specific HR normalization parameters HR_{np}. In turn, HR_{np} was estimated from contextualized HR data $\overline{HR}_{ctx*}$ in free living:

$$X_{hr} = \frac{HR}{HR_{np}} \tag{8.2}$$

$$HR_{np} = \overline{HR}_{ctx*}\beta_{np} + \epsilon \tag{8.3}$$

$\overline{HR}_{ctx*}$ refers to HR data in a specific context, e.g. HR while walking at a certain speed during *relevant activity composites*. Activity composites were discovered using LDA. LDA is a generative probabilistic model which discovers K *activity composites*, from S time windows of N words y_n. Words y_n were stay regions and activity primitives (see Sec. 9.5). According to the generative process, for each word y_n, we first draw the activity composite z_n. Each assigned activity composite $z \in 1 : K$ is derived from a multinomial distribution defined by the parameter θ_s. θ_s is the distribution over activity composites for time window s:

$$\theta_s \sim \mathrm{Dir}(\alpha) \quad 1 \le s \le S \tag{8.4}$$

$$z_n \sim \mathrm{Mult}(\theta_s) \quad 1 \le s \le S, \quad 1 \le n \le N \tag{8.5}$$

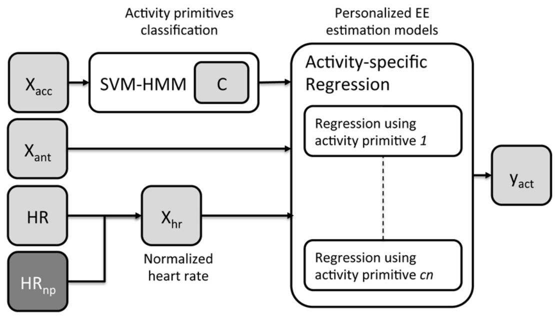

Figure 8.3: Proposed approach to personalized EE estimation. HR data HR were normalized by the HR normalization parameter HR_{np}, resulting in the normalized HR X_{hr}, before being used in activity-specific EE models.

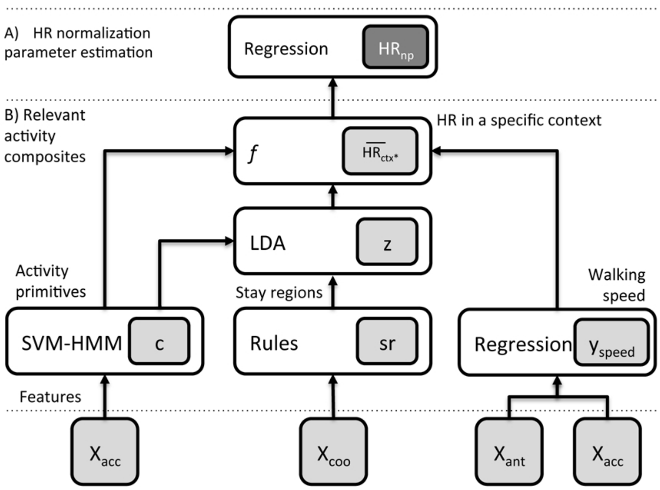

Figure 8.4: Proposed approach to determine the HR in a specific context $\overline{HR}_{ctx*}$, i.e. HR while walking at a certain speed during *relevant activity composites*, and estimate the HR normalization parameter HR_{np}. Activity primitives c and stay regions sr are determined from accelerometer features X_{acc} and GPS coordinates X_{coo}. LDA uses activity primitives and stay regions to discover a set of activity composites, which are ranked, determining *relevant activity composites*. Finally, a regression model is used to estimate the HR normalization parameter HR_{np} from contextualized HR, $\overline{HR}_{ctx*}$.

LDA defines θ_s as a Dirichlet distribution with hyperparameter α. Then, another multinomial is used to choose a word y_n, conditioned on the activity composite z_n, $p(y_n|z_n)$:

$$y_n \sim \text{Mult}(\beta_{z_n}) \quad 1 \leq n \leq N \tag{8.6}$$

Where β is defined as the probability of each word $n \in 1 : N$ for topic z. The joint distribution can be specified as:

$$p(y, z, \theta, \phi|\alpha, \beta) =$$

$$\prod_{s=1}^{S} \int p(\theta_s, \alpha) \prod_{n=1}^{N} \sum_{z=1}^{K} p(z_{sn|\theta_s}) p(y_{sn}|z_{sn}, \beta) d\theta_s \tag{8.7}$$

We were interested in estimating the distributions of the parameter θ_s. Multiple activity composites were derived by LDA in each time window s, each activity composite being assigned a probability. For each time window we considered only the activity composite maximizing θ_s, which we selected as the window's main activity composite z_s.

8.3.2 Relevant activities composites

During the training phase, we defined a feature selection method to determine which activities composites to use as *relevant activity composites*. The HR while walking at a certain speed was computed for each main activity composite z_s and participant par, resulting in the matrix HR_{ctx}. HR_{ctx} is of dimension $K \times npar$, where K is the number of activity composites z, and $npar$ is the number of participants par. One column of the matrix HR_{ctx}, i.e. contextualized HR for one participant across activity composites, is shown in Fig. 8.6.b. LDA-derived activity composites do not include semantics and cannot be compared across participants. To overcome the problem of comparing activity composites, our feature selection method ranks activity composites using a features set T. For example, $T_1 \in T$ could be *the total time spent in each activity composite*, as shown in Fig. 8.6. Then, HR_{ctx} is ranked by feature T_1, allowing us to investigate the relation between the HR in different activity composites and HR_{np}, across participants. The ranking orders HR_{ctx} by values of T from maximum to minimum, as shown in Fig. 8.6.c. Since we were interested in highlighting commonalities between activity composites, ranked HR_{ctx} were smoothed by a moving average of m elements over activity composites, resulting in $\overline{HR}_{ctx}$ (see Fig. 8.6.d). We conclude the training phase by determining which feature in T maximizes Pearson's correlation between $\overline{HR}_{ctx}$ and HR_{np}. We define the vector of correlations r_T for a set of TN features:

$$r_T = \{r_{rank_{T1}}, \ldots, r_{rank_{TN}}\}, \tag{8.8}$$

$$r_{rank_i} = r(\overline{HR}_{ctx_{par=\{1,\ldots,npar\},i}}, HR_{np_{par=\{1,\ldots,npar\}}}) \tag{8.9}$$

Where r_{rank_i} is the correlation between the vector $\overline{HR}_{ctx}$ and HR_{np}, among all participants par for a feature T_i. The feature $T_i = \max r_T$ showing the highest correlation between $\overline{HR}_{ctx}$ and HR_{np} was chosen as indicative of which activity composites are such that HR is more representative of fitness levels, i.e. *relevant activities composites*. For new participants, the function f in Fig. 9.4 ranks HR_{ctx} based on the feature T_i maximizing the correlation on our training set, and determines $\overline{HR}_{ctx*}$. $\overline{HR}_{ctx*}$ is the HR while walking at a certain speed during *relevant activities composites*. Thus, $\overline{HR}_{ctx*}$ is the first element of the vector of ranked and smoothed HR, $\overline{HR}_{ctx}$. Once determined, $\overline{HR}_{ctx*}$ is used to estimate the HR normalization parameter HR_{np} and normalized HR, as shown in Eq. 2 and 3.

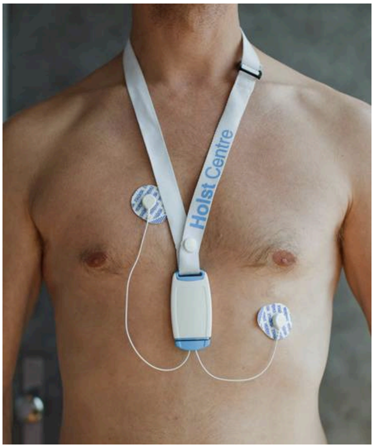

Figure 8.5: ECG Necklace, the wearable sensor used to collect accelerometer and ECG data in this study. The ECG Necklace was worn during laboratory protocols and free living recordings close to the body's center of mass

8.4 Evaluation study

8.4.1 Participants and data acquisition

Participants were 34 (14 male, 20 female) self-reported healthy individuals, mean age 23.7 ± 2.5 years, mean weight 66.3 ± 10.6 kg, mean height 172.4 ± 8.3 cm, mean BMI 22.2 ± 2.5 kg/m^2 and mean VO_2max 3002.9 ± 665.0 ml/min. Written informed consent was obtained, and the study was approved by the ethics committee of Maastricht University. The sensor platform used was the ECG Necklace, which was configured to acquire one lead ECG data at 256 Hz, and three-axial accelerometer data at 32 Hz (see Fig. 9.6). The ECG Necklace was worn close to the body's center of mass, thus in an ideal location for EE estimation, as reported in literature [14]. The ECG Necklace was worn during laboratory protocols and free living recordings. A *Continuous Wavelet Transform* based beat detection algorithm was used to extract R-R intervals from ECG data recorded under laboratory conditions, which output was manually examined to correct for missed beats that

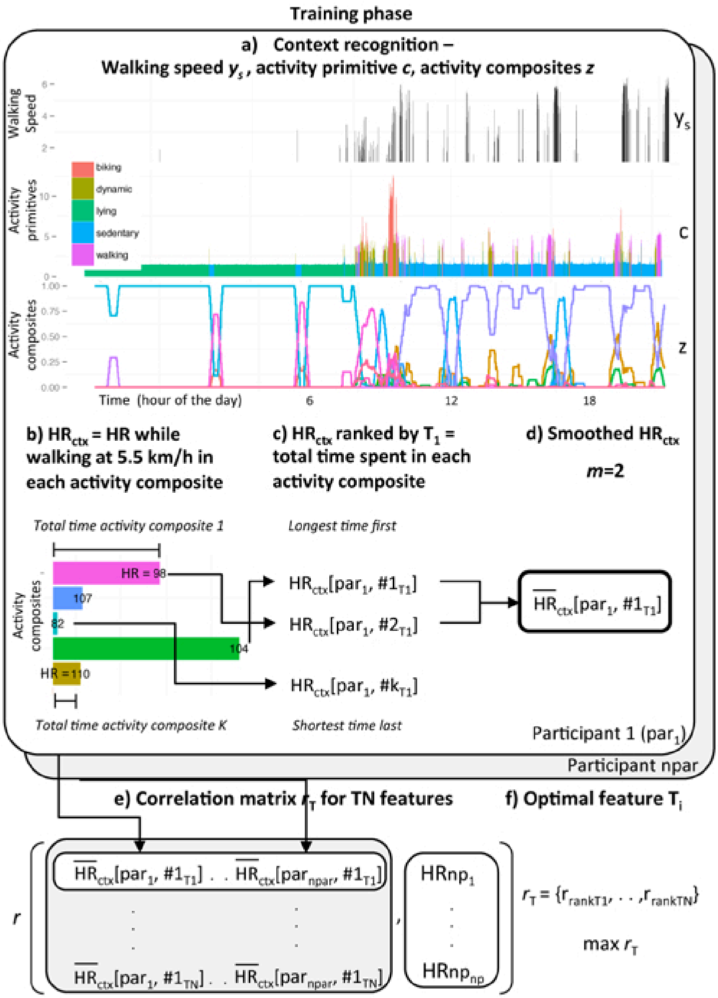

Figure 8.6: Exemplary diagram of our approach to discover *relevant activity composites* for the case of walking at 5.5 km/h. $a-b$) Walking speed y_s, activity primitives c and activity composites z are used to determine HR in specific contexts, HR_{ctx}. c) HR_{ctx} are ranked by activity composite feature T_1, the *total time spent in each activity composite*. Bars in plot b) indicate values for T_1 in each activity composite, while numbers are average HR while walking at 5.5 km/h for each activity composite. d) Ranked HR_{ctx} were smoothed by a moving average of $m = 2$ elements. e) $\overline{HR}_{ctx}$ across participants are correlated with the HR normalization parameter HR_{np} in the training dataset. f). The feature T_i maximizing the correlation is chosen to select *relevant activity composites*.

might be caused by motion artifacts [102]. For ECG data recorded under unsupervised free-living conditions, we selected high quality data by discarding periods in which more than 15% errors were detected in a time window. Errors were defined as consecutive RR intervals differing more than 20%, as typically reported in clinical practice. Additionally, during free living each participant carried a Samsung Galaxy S3 used to record GPS coordinates at 5 minutes intervals. During laboratory recordings participants were equipped with a indirect calorimeter analyzing O_2 consumption and CO_2 production (Oxycon-β), from which EE was derived [128]. VO_2max was determined during an incremental test on a cycle ergometer [78]. Activity composites were manually annotated by the participants on a diary, while activity primitives were annotated during laboratory protocols by the experimenter. The dataset acquired contains about 363 days of data collected from 34 subjects in free living, including accelerometer, ECG and GPS data plus 72 hours of laboratory recordings including reference VO_2 and VCO_2 for validation of EE estimation.

8.4.2 Experimental design and validation procedure

We collected data in free living and laboratory settings. Free living data was used to learn the normalization parameter HR_{np} using the proposed method, which combines activity primitives and *relevant activity composites* to contextualize HR. The proposed approach is referred to as *combined*. Then, activity-specific EE estimation models including normalized HR as a predictor were validated in laboratory settings using reference calorimeter data.

We evaluated the proposed approach in estimating HR_{np} against two other approaches: a) *no-context*: HR in free living is used directly to estimate HR_{np}, b) *low level*: HR in free living is contextualized using activity primitives and walking speed and used to estimate HR_{np}.

EE estimation using HR normalized by the proposed approach was also evaluated against two other approaches: a) *no-normalization*: EE was estimated by activity-specific models using as predictors non-normalized HR, accelerometer and anthropometrics data, b) *low level*: EE was estimated by activity-specific models using as predictors normalized HR, accelerometer and anthropometrics data. For the *low-level* approach HR was normalized by HR_{np} and HR_{np} was determined using activity primitives and walking speed only, but no activity composites.

Two laboratory protocols were designed and implemented for each participant on two separate days to avoid the maximal fitness test to affect physiological parameters during less intense activities and vice versa.

8.4.2.1 Laboratory protocols

Participants reported at the lab on three separate days and after refraining from drinking, eating and smoking in the two hours before the experiment. Two laboratory protocols were performed. The first protocol included simulated activity primitives performed while wearing a portable indirect calorimeter, to acquire reference EE data. Activities included: lying down, sitting, sit and write, standing, cleaning a table, sweeping the floor, walking (treadmill flat at 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6 km/h) and running (treadmill flat at 7, 8, 9, 10 km/h). Activities were carried out for a period of at least 4 minutes. The second protocol was a VO_2max test providing reference data for biking and EE while biking. The third day was used for anthropometric measurements including the participant's body weight, height and body fat. Body fat was assessed using doubly labelled water [131].

8.4.2.2 Free living protocol

Participants worn the ECG necklace for 14 consecutive days in free living and manually annotated their activities composites (high level activities such as going to work, sleeping, etc.). Participants carried a Samsung S3 phone and were instructed to charge both the ECG Necklace and phone and to change electrodes daily.

8.4.2.3 Statistics and performance measures

Models were validated using leave one participant out cross-validation. The procedure was repeated for each participant and results were averaged. Thus, data used for model building was not used for model validation. LDA parameters were derived on data from each participant to be validated, since no reference or training set are necessary. Performance of the activity recognition models was evaluated using the class-normalized accuracy. Results for HR normalization parameters estimation, walking speed estimation and EE estimation are reported in terms of Root-mean-square error (RMSE), where the outcome variables were HR in bpm, speed in km/h and EE in kcal/min respectively.

8.5 Implementation

8.5.1 Features extraction and selection

Accelerometer data were segmented in 5 s windows, band-pass filtered between 0.1 and 10 Hz, to isolate the dynamic component due to body motion, and low-pass filtered at 1 Hz, to isolate the static component, due to gravity. Features X_{acc}were derived and selected based our previous work [9], using a different dataset. Selected features were: *mean of the absolute signal, inter-quartile range, median, variance, standard deviation, main frequency peak, low and high frequency band signal power*. HR was extracted from RR intervals, computed over 15 seconds.

8.5.2 Activity primitives

Laboratory activities were grouped into six clusters c_i to be used for classification of activity primitives. The six clusters were *lying* (lying down), *sedentary* (sitting, sit and write, standing), *dynamic* (cleaning the table, sweeping the floor), *walking*, *biking* and *running*. Activity primitives were derived combining a SVM and HMM. For the SVMs, we used a gaussian radial basis kernel (C = 1). The HMM is defined by parameters $\lambda = (\pi, A, B)$. π is the vector of probabilities of each state (i.e. low level atomic activity) to be the initial state, A is the transition probability matrix, defining the probability of transitioning between one activity to the other at time interval t. Thus, the HMM states correspond to activity primitives. B is the emission matrix, which defines the probability of getting an emission at time t, given the state. We implemented the emission matrix B as $b_{ij} = 0.5 \iff i = j$, $b_{ij} = 0.1 \iff i \neq j$, while transitions probabilities A between actual states were derived from training data. Training data was the SVM classification result obtained with reference activity primitives manually annotated in laboratory settings.

8.5.3 Walking speed

Features for the linear regression model used to estimate walking speed were: *mean of the absolute signal, inter-quartile range, variance, main frequency peak, high frequency band signal power* and *height*, as derived by linear forward selection [10]. Free living walking speeds used to contextualize HR were 4.5 km/h (4 to 5 km/h

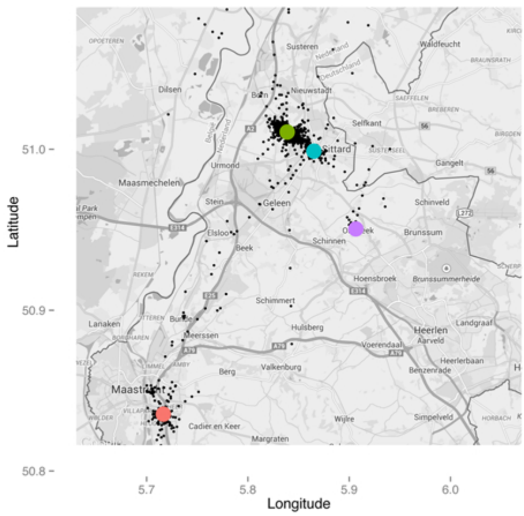

Figure 8.7: Exemplary stay regions detection from noisy GPS data for one participant. Small dots represent all recorded GPS data points, while bigger dots represent detected stay regions.

range) and 5.5 km/h (5 to 6 km/h range) since speeds close to this values were reported in healthy individuals (5.3 km/h in [36] and 5 ± 0.8 km/h in [89]).

8.5.4 Stay regions

Stay regions were computed from GPS data X_{coo}according to time and distance thresholds, which were set to 60 minutes and 1 km according to previous literature [135]. The time threshold ensures that each stay region is a location where the participants spent a substantial amount of time, while the distance threshold ensures that noisy recordings do not result into a multitude of stay regions being detected. GPS data was collected at 5 minutes intervals to conserve battery power. The relatively wide distance and time thresholds were chosen due to the low frequency of the GPS recordings. An example of stay region detection for one participant is shown in Fig. 8.7.

8.5.5 Relevant activity composites discovery

Input primitives for LDA were occurrences histograms of stay regions and activity primitives in time windows s. LDA hyperparameter α was set to 0.01, while segment size and number of activity composites K were set to 15 minutes and 20 topics respectively, based on results obtained in previous research [112]. Parameters were optimized using an implementation of the variational expectation-maximization algorithm proposed in [27]. The function f (see Fig. 9.4) translates LDA-derived activity composites into *relevant activity composites* by first determining the most probable activity composites in each time window s, as expressed by the parameter θ. Secondly, HR during activity composites HR_{ctx} was ranked according to features T, including *amount of time spent in each activity composite*, *amount of time spent in each activity primitive with respect to the total time spent performing the activity* and *percentage of time spent in each activity primitive per activity composite*. Features were chosen to be computed across participants and activity composites regardless of the participant lifestyle or activity composite semantics, while possibly providing information about which activity composite might retain more of the relation between HR and HR_{np}. Ranking of HR_{ctx} values was smoothed by a moving average of 5 elements. Ranked and smoothed $\overline{HR}_{ctx}$ were correlated with HR_{np} to determine which activity composites features were more representative of HR_{np}.

8.5.6 HR normalization parameter estimation

We chose the HR while running at 9 km/h as the HR normalization parameter HR_{np} to estimate in free living. Our choice was motivated by previous laboratory results reported by our group [10] as well as others [118], showing that HR normalized by the HR while running at 9 km/h highly reduces variability between participants. A linear regression model was built to predict HR_{np} using as independent variable the HR while walking at 4.5 km/h or 5.5 km/h during *relevant activity composites*, $\overline{HR}_{ctx*}$. We also implemented the models listed in Sec. 9.4 as benchmarks for the proposed approach (referred to as *combined*).

8.5.7 Personalized EE estimation

EE was estimated by first classifying the activity performed among the ones listed in Sec. 9.5.1.2 and then applying an activity-specific EE linear regression model. The activity-specific EE linear models used anthropometric characteristics, motion intensity and HR as predictors. For the proposed approach, HR was normalized by the HR normalization parameter HR_{np}, as estimated using HR contextualized by activity primitive and *relevant activity composites*. We also implemented the models listed in Sec. 9.4 as comparisons for the proposed approach, thus estimating EE using non-normalized HR (*no-normalization*) and HR_{np} estimated using HR contextualized by activity primitive only (*low level*).

Figure 8.8: Exemplary walking speed estimation and activity primitives recognition for one participant. Activities were manually annotated and performed sequentially. Improvements in activity primitives recognition using a combined SVM-HMM compared to a single SVM are shown in plots b and c.

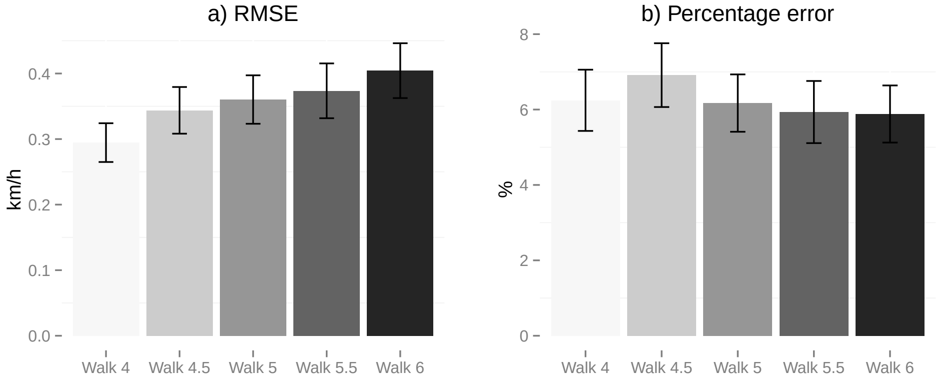

Figure 8.9: a) RMSE for walking speed models across the speed range used to contextualize HR in free living. b) Percentage error across the same speeds.

8.6 Results

8.6.1 Activity primitives and speed

Accuracy of the SVM-HMM activity recognition classifier was 92.3%. More specifically, the accuracy was 94.4% for *lying*, 96.7% for *sedentary*, 77.6% for *dynamic*, 96.3% for *walking*, 93.3% for *biking* and 95.5% for *running*. Walking speed estimation RMSE was 0.38 km/h. Results for walking speed estimation across the speeds used to contextualize HR in free living are shown in Fig. 8.9, while an exemplary output of the activity primitives recognition classifier and walking speed estimation model is shown in Fig. 8.8.

8.6.2 HR normalization parameter

An example of LDA-derived activity composites is shown in Fig. 8.6. Activities composites were ranked according to the features listed in Sec. 8.5.5. The feature T_i maximizing the relation between HR_{np} and ranked HR_{ctx} was total time spent in each activity composite, resulting in correlation $r = 0.73$. Correlation between HR_{np} and mean HR in free living (*no-context*) was $r = 0.46$ while correlation between HR_{np} and mean HR while walking in free living (*low level*) was $r = 0.53$ for walking at 4.5 km/h and $r = 0.55$ for walking at 5.5 km/h. HR_{np} estimation resulted in RMSE of 13.8 bpm for *no-context*, 13.2 bpm for *low level* when data while walking at 4.5 km/h was used, and 12.6 bpm for *low level* when data while walking at 5.5 km/h was used. For the proposed approach (*combined*), RMSE was reduced to 11.1 bpm and 10.1 bpm when using data while walking at 4.5 km/h and 5.5 km/h respectively. Thus, the proposed approach provided 29.4%

and 19.8% error reduction in estimated HR compared to *no-context* and *low level*. Including data while walking at higher speed (i.e. 5.5 km/h) provided the best results. Fig. 8.10 shows the relation between measured and predicted HR_{np} for the different cases considered in this work.

8.6.3 EE estimation

EE estimation results are shown in Fig. 8.11. Benchmark for this analysis were state of the art activity-specific EE estimation models including accelerometer and non-normalized HR data, (*no-normalization*), resulting in RMSE of 0.84 kcal/min. RMSE was reduced from the *no-normalization* condition to 0.79 kcal/min (6.4% error reduction) for *low level* and to 0.75 kcal/min (10.7% error reduction compared to *no-normalization*, $p = 0.007$ and 4.6% error reduction compared to *low level*, $p = 0.037$) for *combined*, the proposed approach. We provide detailed results for moderate to vigorous activities only, since personalizing the relation between HR and EE is mostly not useful during sedentary activities [13]. EE RMSE was reduced from 0.55 kcal/min to 0.53 kcal/min for *walking* (4.2% error reduction), from 2.34 kcal/min to 1.92 kcal/min for *biking* (18.0% error reduction) and from 1.12 kcal/min to 1.03 kcal/min for *running* (8.0% error reduction) using the proposed approach, compared to no-normalization.

8.7 Discussion

In this paper, we proposed an approach to estimate HR normalization parameters during free living. Then, we used the normalization parameters to normalize HR and reduce EE estimation error compared to population-based models obtained in laboratory conditions. The effectiveness of HR normalization parameters in reducing EE estimation error has been shown in previous literature [10, 13, 118]. However, to the best of our knowledge, this is the first work which estimates person-specific HR normalization parameters using unsupervised recordings in free living.

The presence of a multitude of stressors in free living required a different solution from what was introduced in laboratory settings. Our hypothesis was that HR in free living is not only affected by low level activity primitives - as shown in the lab - but by both activity primitives and high level activities composites. Thus, incorporating contextual information beyond activity primitives could potentially improve interpretation of HR in free living. Our results confirm the importance of activity composites in interpreting HR data in free living. HR normalization parameter estimation RMSE was reduced by 29.4% compared to average free living HR - i.e. no context - when using the HR while walking at 5.5 km/h during *relevant activity composites* as predictor. On the other hand, when HR normalization parameters were estimated using low level context information only, i.e. the HR while walking at 5.5 km/h across all activity composites, RMSE was reduced by 8.7% only compared to no context. We evaluated the proposed approach for

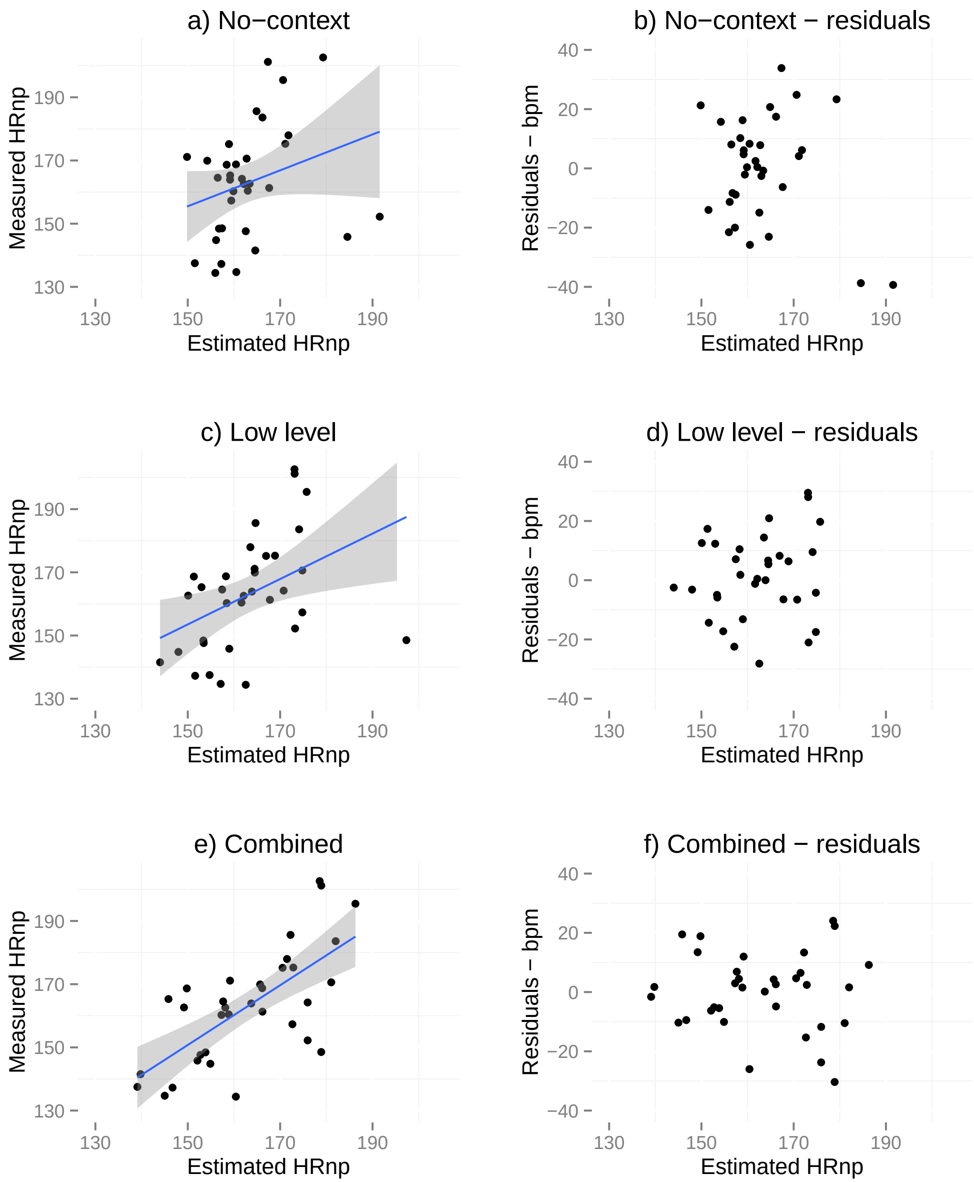

Figure 8.10: a,c,e) Relation between measured and estimated HR normalization parameters for the three conditions compared in this work: a) no-context, c) low level, e) combined. b,d,f) Residuals plots for the three conditions compared in this work: b) no-context, d) low level, f) combined. For low level and combined, only data while walking at 5.5 km/h was used, as it provided the optimal results (see Sec. 8.6.2).

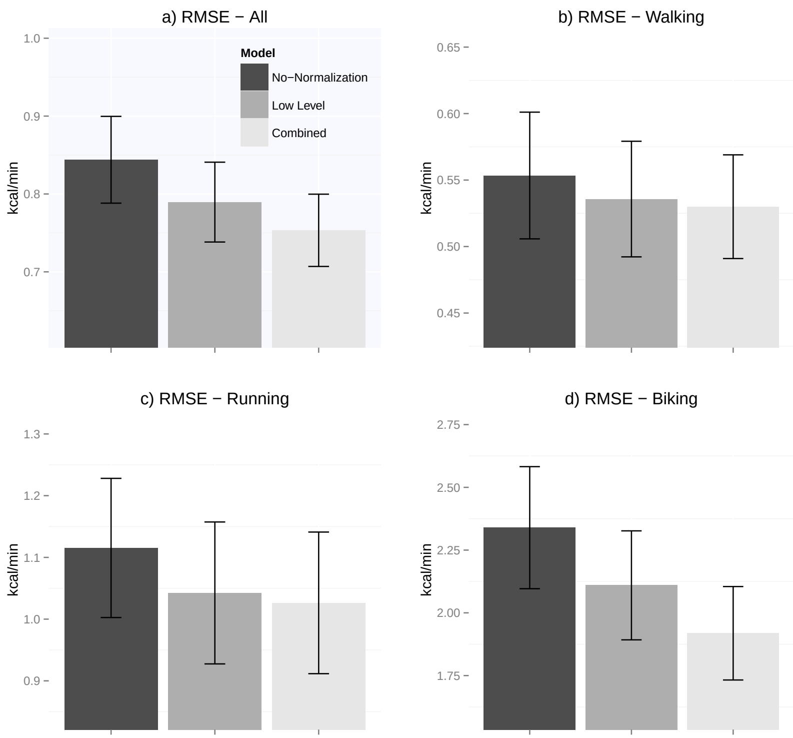

Figure 8.11: EE estimation RMSE and standard errors for a) all activities averaged, b) walking c) running and d) biking. Three models are compared, *No-normalization*, *Low Level*: HR was normalized using a normalization factor predicted from HR while walking at 5.5 km/h, and *Combined*, i.e. the proposed approach, normalizing HR using a normalization factor predicted from HR while walking at 5.5 km/h during *relevant activity composites*.

a wide range of walking speeds, from 4 to 6 km/h, and found that higher speeds resulted in better results.

We translated the need for high level contextual information into a recognition framework and introduced the concept of *relevant activity composites*. Relevant activity composites are activity composites in which HR is more representative for HR normalization parameters. While supervised methods have been introduced in literature to determine high level activity composites, these methods require to know in advance what high level activity composites will be performed by the participants, as well as sufficiently annotated data to train models. Most importantly, supervised methods assume every participant to perform the same activity composites, which is unlikely in free living. Our unsupervised approach relies on TM, in particular LDA, to discover activity composites. To determine which activity composites will be used to estimate HR normalization parameters, our method *ranks* activity composites depending on different features. Our approach thus discovers activity composites, which may differ for each participant, depending on their lifestyle. However, discovered activity composites do not provide semantics and comparison between participants is challenging. Typically, activity composite of interest are isolated and further classified using supervised methods [72, 112], thus requiring prior knowledge of the activity composites to discover, effectively limiting the unsupervised nature of the method. Ranking allowed for comparison of activity composite specific features (e.g. total time spent in each activity composite) across participants, even if activity composites were different and without semantics. Thus, making the *relevant activity composite* discovery approach unsupervised and generalizable to new participants. In particular, we found a strong relation between the total time spent in each activity composite and the HR normalization parameter. A possible explanation is that activity composites in which people spend most of their time are typically representative of a stable physiological condition, which might be more representative of their fitness level. On the contrary, infrequent or more intense activity composites might involve more physiologically stressful situations as well as intermittent HR, causing cardiovascular responses which are not reliable for HR interpretation [101]. While our method determines activities which are best suited for HR normalization, the role of other factors affecting HR, for example emotional stress or illness, could not be directly evaluated, due to lack of reference. Future work could explore the relation between relative activity composites and external factors such as stress, to further validate the effectiveness of the proposed approach in determining high level context useful for EE estimation.

Free living recordings were used to determine HR normalization parameters unsupervisedly and without requiring any individual calibration or laboratory tests. However, the effectiveness of the estimated HR normalization parameters in reducing EE estimation error were validated in laboratory settings. Double labelled water (DLW) is the only recognized method to obtain reference EE in free-living [31, 117]. However, DLW reports only total EE after a period of one or two weeks. Thus, DLW is not informative in terms of minute-by-minute EE estimation accuracy. An EE estimation model that would consistently overestime light activities and consistently underestimae intense activities could perform optimally according to DLW, due to an averaging of multiple errors. Thus, we validated our approach using laboratory data and reference indirect calorimetry, since only under these conditions we can acquire minute-by-minute EE reference for different activities, and evaluate the models' accuracy. Similarly, we could evaluate activity recognition and walking speed models only under laboratory conditions, where reference is present. The dynamic activity cluster was recognized with accuracy below average. We interpret that activities with high variability in movement and execution between participants and using a single chest-worn sensor resulted in higher classifier confusions. However, the high accuracy of walking speed estimation models and activity recognition for walking provide confidence for the free-living detection of activities used to contextualize HR.

We believe our approach is a substantial step towards personalized health and wellbeing monitoring. The proposed system learns automatically from the user over time, collecting accelerometer, HR and GPS data while performing activities of daily living unsupervisedly. Recent developments in wearable and mobile technology provided sensors and phones able to collect and process data continuously and unobtrusively [12]. Our methodology, could be applied to such systems to determine the HR normalization parameter, a coefficient representative of the fitness level of an individual. By normalizing HR using the estimated HR normalization parameter, EE estimation can be personalized. Our results show that RMSE was reduced by 10.7% on a dataset of participants with high variability in fitness level, using cross-validation.

We expect that the HR normalization approach will be most useful to individuals willing to take up a more active lifestyle, or undergoing a physical activity intervention targeted in modifying behavior to increase level of activity. The importance of CRF and its influence on HR is particularly relevant for individuals transitioning from inactive to active lifestyle. HR normalization provides optimal results for moderate to vigorous activities, especially the ones where accelerometer data is not indicative of EE due to lack of whole body movement (as shown by the highest reduction in RMSE for EE estimation when biking, 18.0%). Other activities such as rowing, walking uphill, etc. would most likely benefit as well, due to the inability of accelerometers alone to estimate EE accurately. The proposed EE estimation approach will be useful for sports training devices, where users and trainers are interested in accurate EE estimation under moderate to vigorous workloads. However, using low intensities activities, such as walking at preferred speeds in healthy individuals [36] , [89] we aim at providing accurate EE estimation in daily life across the general population. The proposed algorithms can adapt to individual fitness level and high level activity composites. The proposed approach could be used to guide in healthy lifestyle, by providing more accurate EE estimation at the individual level.

Acknowledgment

The authors would like to thank G. Plasqui, G. Schiavone, G. ten Velde and S. Camps for support during data collection.

9

Cardiorespiratory fitness estimation in free living using wearable sensors

M. Altini, P. Casale, J. Penders, O. Amft
Submitted to Artificial Intelligence in Medicine

Abstract

We propose a method to estimate cardiorespiratory fitness (CRF) in free living using wearable sensor data. Our method uses three estimation steps and does not require laboratory tests, calibration or specific exercise. Initially, we recognize activity primitives using accelerometer and GPS data. Using topic models (TMs), we group activity primitives and derive activities composites. We subsequently rank activity composites, and analyze the relation between ranked activity composites and CRF across individuals. Finally, heart rate (HR) data in specific activity primitives and composites is used as predictor in a hierarchical Bayesian regression model to estimate CRF level from the participant's habitual behavior in free living. We show that combining activity primitives and activity composites outperforms other CRF estimation models reducing estimation error between 10.3% *and* 22.6% *on a study population of* 46 *participants with varying CRF level.*

9.1 Introduction

Cardiorespiratory fitness (CRF) is the ability of the circulatory and respiratory systems to supply oxygen during sustained physical activity and is considered among the most important determinants of health and wellbeing. CRF is not only an objective measure of habitual physical activity (PA), but also a useful diagnostic and prognostic health indicator for patients in clinical settings, as well as healthy individuals [81]. Epidemiological research has shown that in both individuals affected by disease [119] and healthy individuals [127, 25] higher level of CRF resulted in better outcomes in term of slower disease progression, lower risk of cardiovascular disease as well as lower risk of all cause mortality. Thus, knowledge of CRF can be key in managing a healthy lifestyle.

Wearable sensors have great potential for accurate PA monitoring in daily life [39, 44]. However, almost all solutions to monitor PA focus on behavioral aspects such as steps, activity type and energy expenditure (EE) [33, 125]. While activity type, EE, steps, etc. are important, they reflect only individual behavior, but do not provide insights on the individual's actual health status. CRF is a marker of cardiovascular and cardiorespiratory health, and therefore is a key health parameter [81, 99, 111]. Current practice for CRF measurement is direct measurement of oxygen volume (VO_2 in ml/min) during maximal exercise (i.e.VO_2max), the gold standard. Typically, VO_2max tests consist in measuring VO_2 using an indirect calorimeter during an incremental exercise test, either on a bike or treadmill.

However, VO_2max tests are affected by multiple limitations. Medical supervision is required and the test can be risky for individuals in non-optimal healthy conditions. Less risky submaximal tests have also been developed [18] too. Submaximal tests to estimate CRF typically require measuring heart rate (HR) while running at a certain speed or biking at a certain intensity. The inverse relation between HR at a certain exercise intensity, fixed by the strict exercise protocol that has to be sustained, and fitness, is the rationale behind this approach. The need for laboratory equipment and the necessity to re-perform the test to detect changes in CRF limit practical applicability of submaximal tests.

In this work, we propose a method to estimate CRF using wearable sensor data acquired in free living. We rely on the inverse relation between HR and fitness, but without the need for laboratory tests or specific exercise protocols. We aim at using machine learning techniques to determine contexts in which HR can be interpreted usupervisedely in free living. Our hypothesis is that physiological data, for example HR, in free living settings is not only affected by activity primitives such as walking, but by a combination of activity primitives and more abstract activity composites such as social interactions, working, etc. Thus, we propose a method to determine both low level activity primitives and high level activity composites, to contextualize HR. Finally, we use contextualized HR to estimate CRF in a hierarchical Bayesian model. By using a non-nested hierarchical Bayesian model, parameters can vary depending on the activity performed, therefore being more flexible than models requiring specific activities. This paper provides the following contributions:

1. We propose a context recognition framework to contextualize HR and estimate CRF based on contextualized HR in free living. First, we use topic models (TMs) to derive activity composites. Secondly, we rank activity composites to determine which activity composites are best suited for CRF estimation. Finally, we use HR data in specific contexts (i.e. activity primitives, walking speeds and activities composites) as a predictor in a hierarchical Bayesian model to estimate CRF.

2. We show the effectiveness of the proposed approach to estimate CRF on a

dataset including 14 days of unsupervised free living recordings from 46 participants and reference VO_2max acquired in laboratory conditions. CRF estimation error was reduced between 10.3% and 22.6% compared to alternative methods.

9.2 Related work

9.2.1 Maximal and Submaximal Tests

VO_2max is regarded as the most precise method for determining CRF [124]. Despite the indubitable importance of CRF in health, measurements of VO_2max are rare since they require specialized personnel and expensive equipment. The high motivation demand and exertion of the subjects makes the test unfeasible in many patients groups [94]. As an alternative, many non-excercise and sub-

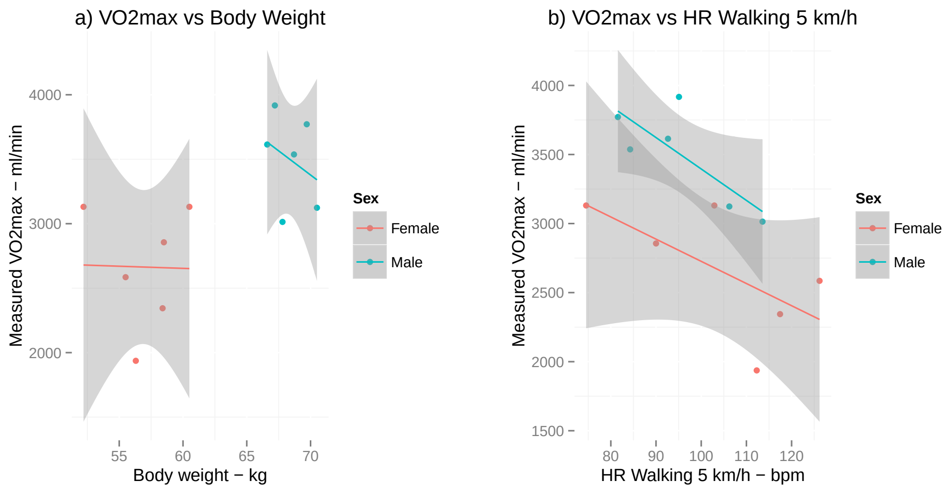

Figure 9.1: Relation between body weight, HR and CRF for participants with similar body size (weight and height) characteristics. a) Positive relation between VO_2max and body weight disappears when participants with similar body size characteristics are considered. b) Negative relation between VO_2max and HR while walking holds on a subset of participants with similar body size, and can potentially be used to discriminate CRF levels.

maximal models have been developed. Non-exercise models of CRF use easily accessible characteristics such as age, gender and self-reported PA level [73, 93]. However, for individuals with similar characteristics, CRF levels cannot be discriminated, as shown in Fig. 9.1. Submaximal tests have been developed to estimate VO_2max during specific protocols while monitoring HR at predefined workloads [18]. Contextualized HR, e.g. HR while performing a specific activity in laboratory settings, is discriminative of CRF levels between individuals with similar characteristics, due to the inverse relation between HR and CRF [100] (see Fig. 9.1). Commercial devices, for example some sport watches paired to HR monitors [55, 53], provide CRF estimation using this principle, i.e. using a regression model including HR at a predefined running speed as predictor. However, submaximal tests are still affected by limitations; the test should be re-performed every time CRF needs to be assessed, often requires laboratory infrastructure and specific activities to be performed [106].

9.2.2 CRF estimation in free living

Estimating VO_2max from parameters derived using wearable sensor data acquired in free living can potentially be applied to a larger population, compared to maximal or submaximal laboratory tests. Preliminary work explored the relation between PA level as expressed by a step counter, and CRF [40]. While PA level can provide useful insights, the relation between HR and oxygen uptake at a certain exercise intensity cannot be exploited using motion based sensors. Plasqui et al. [100] showed that a combination of average HR and activity level over a period of seven days correlates significantly with VO_2max. However, the relation between average HR and activity counts depends on the amount of activity performed [100]. Tonis et al. [120] explored different parameters to estimate CRF from HR and accelerometer data during activities of daily living simulated in laboratory settings. However VO_2max reference and free living data were not collected. When moving towards free-living settings, HR is more difficult to in-

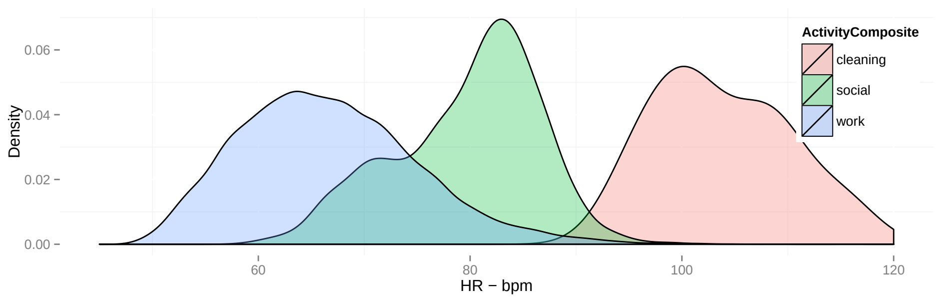

Figure 9.2: Density plot of HR data during the activity primitive *sedentary*, occurring in different *activity composites*, i.e. *cleaning*, *social*, *work*. Although the activity primitive *sedentary* occurs in all activity composites, HR differs consistently across activity composites. Thus, detecting activity composites can improve interpretation of HR in free living, and therefore provide more accurate CRF estimation. Activities composites were manually annotated.

terpret, since activities vary depending on the different lifestyles people adopt.

9.2.3 Context recognition

We hypothesize that HR in free living settings is not only affected by activity primitives but by a combination of activity primitives and more abstract activity composites. For example, HR during the activity primitive *sedentary* changes substantially depending on the context in which such activity is performed. HR during social interactions is higher than during work for sedentary activities, possibly due to the higher physiological stress involved in talking and interacting with other persons, as shown in Fig. 9.2. Thus, CRF estimation models might benefit from inclusion of activity composites, which are representative of the presence of a multitude of stressors present in free living.

Typically, activities are thought of in a hierarchical manner, starting from low level activity primitives, to more abstract activity composites [16]. An example of activity primitives can be a set of postures and locomotion activities, such as: *lying down, sedentary, dynamic, walking, biking* and *running*, as determined using supervised methods in previous research [20]. On the contrary, higher level contextual information, such as activity composites, require a different recognition approach. Such activities are personal and need unsupervised methods able to discover different patterns in each individual, depending on their behavior. A possible solution is the use of TMs [27]. In activity recognition, TMs were applied to discover activity composites from activity primitives [72]. Recent work investigated the impact of multiple TMs (in particular LDA, latent Dirichlet allocation) parameters for activity composites discovery, showing promising results [112] for recognition of abstract high level activities.

In our previous work [5], we proposed a method to determine which activity composites are better suited to interpret HR for one individual. For example, we determined in which activity composites HR was more representative of HR normalization parameters used to personalize EE estimates. Our approach consisted of ranking activity composites based on features in order to compare them across participants. In this work, we extend our method to the relation between HR during activity composites and VO_2max. We aim at finding for each individual specific contexts where HR is representative of CRF in free living, using an unsupervised approach. Then, we use contextualized HR to predict CRF without the need for laboratory tests or specific exercises.

9.3 Approach

Following a top down approach, CRF y_{CRF} was estimated from contextualized HR $\overline{HR}_{ctx*}$ and anthropometric characteristics by a hierarchical Bayesian regression model, as shown in Fig. 9.3. Contextualized HR $\overline{HR}_{ctx*}$ refers to HR during specific activity primitives, speeds and *relevant activity composites*. We used features from accelerometer X_{acc}, HR X_{hr}, location X_{coo} and anthropometrics X_{ant} as input to our context recognition and CRF estimation models. Activity primitives c were used together with stay regions sr as input for LDA topic discovery to obtain activity composites. Activity composites were ranked to find the most relevant ones for CRF estimation, referred to as *relevant activity composites* (see Sec. 9.3.3 for details). The procedure to determine activity primitives, speeds, activity composites, and therefore contextualized HR $\overline{HR}_{ctx*}$ is shown in Fig. 9.4.

In the remaining of this section, we detail the approach and provide an example. We consider *walking* at 3 and 5 km/h as exemplary activity primitives and speeds. Thus, to determine contextualized HR, we consider HR data while *walking* at 3 and 5 km/h during *relevant activity composites*.

9.3.1 CRF estimation

The CRF estimation y_{CRF} was derived by a hierarchal Bayesian regression model. Parameters modeling the relations between $\overline{HR}_{ctx*}$ and y_{CRF} vary depending on the context *ctx*. We denote the estimation model as:

$$\begin{gathered} y_{CRF_p} \sim N(X_{CRF_p}\beta_{CRF} + X_{ctx[p]}\beta_{ctx[p]}, \sigma^2_{CRF}), \\ ctx = 1,\ldots,R \quad p = 1,\ldots,np \\ X_{CRF_p} = [1, X_{ant_p}] \in \mathbb{R}^{np\times(D+1)}, \quad p = 1,\ldots,np \\ X_{ctx} = [\overline{HR}_{ctx*}] \in \mathbb{R}^{np\times 1} \quad p = 1,\ldots,np \end{gathered} \tag{9.1}$$

where matrix X_{CRF_p} is of dimension $np \times (D+1)$. np is the number of participants, while D the number of anthropometric characteristics X_{ant_p} for a person p, which includes *body weight*, *height*, *age* and *sex*. The associated parameters β_{CRF} do not vary by context *ctx* since they are relative to a person and remain the same across different activities. Contexts *ctx* are a set R representing a combination of activity primitives and speeds during *relevant and activity composites*, as shown in Fig. 9.3. In our example, contexts are $R = 2$, i.e. walking at 3 or 5 km/h during *relevant activity composites*, and control the parameters β_{ctx} for the predictor $\overline{HR}_{ctx*}$. By letting the parameters β_{ctx} vary, users are not constrained to one specific activity. Instead, the model will provide a CRF estimate y_{CRF} depending on the available activity primitives and speeds. Details on the model parameters estimation procedure are reported in Sec. 9.5. For validation purposes, we used leave one participant out cross-validation, therefore building a model for $np - 1$ participants, and then evaluating in on the remaining one. The procedure was repeated np times, as further detailed in Sec. 9.4.

9.3.2 Context recognition

In this section we introduce our context recognition architecture to determine contextualized HR $\overline{HR}_{ctx*}$, as shown in Fig. 9.4. Activity composites were discovered using LDA. LDA is a generative probabilistic model which discovers K *activity composites*, from S time windows of N words y_n. For activity recognition, words y_n are typically basic building blocks for higher level activities, such as low level atomic activities. In our implementation we used stay regions and activity primitives (see Sec. 9.5) as words y_n. Accelerometer features X_{acc} were used to derive activity primitives c_i combining a Support Vector Machines (SVM) classifier

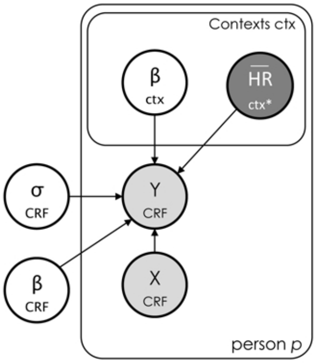

Figure 9.3: Hierarchical Bayesian model in plate notation. Parameters β_{ctx} vary by context *ctx* and model the relation between contextualized HR $\overline{HR}_{ctx*}$ and CRF y_{CRF}.

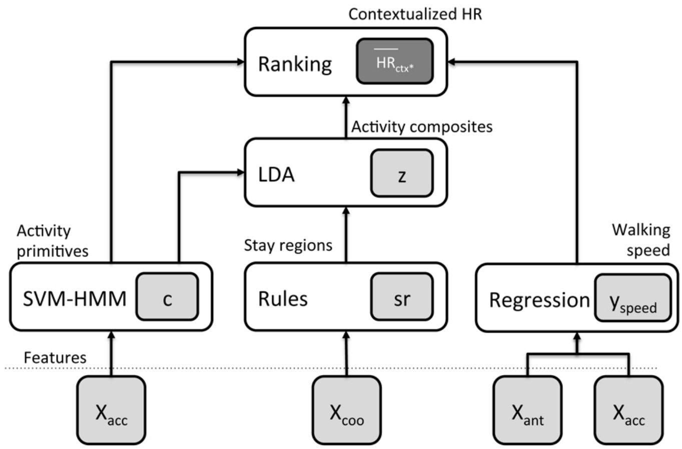

Figure 9.4: Proposed approach to determine contextualized HR $\overline{HR}_{ctx*}$. LDA uses histograms of activity primitives c and stay regions sr to discover a set of activity composites, which are ranked to determine *relevant activity composites*. Contextualized HR $\overline{HR}_{ctx*}$ is shown in the top block, and is determined by combining activity primitives, activity composites and speed. $\overline{HR}_{ctx*}$ is used as input for the CRF estimation model detailed in Fig. 9.3.

and subsequent Hidden Markov Models (HMM) used to smooth transitions between activities. The hidden states corresponded to the real activity composites, c_i, while the observable states are the ones recognized by the SVM. *Stay regions* were derived from GPS coordinates X_{coo} using time and distance thresholds (see

Sec. 9.5). According to the generative process, for each word y_n, we first draw the activity composite z_n. z_n is a scalar $z \in 1 : K$ indicating the activity composite for time window n. Each assigned activity composite z_n is derived from a multinomial distribution defined by the parameter θ_s. θ_s is the distribution over activity composites for time window s:

$$\theta_s \sim \mathrm{Dir}(\alpha) \quad 1 \leq s \leq S \tag{9.2}$$

$$z_n \sim \mathrm{Mult}(\theta_s) \quad 1 \leq s \leq S, \quad 1 \leq n \leq N \tag{9.3}$$

LDA defines θ_s as a Dirichlet distribution with hyperparameter α. Then, another multinomial is used to choose a word y_n, conditioned on the activity composite z_n, $p(y_n|z_n)$:

$$y_n \sim \mathrm{Mult}(\beta_{z_n}) \quad 1 \leq n \leq N \tag{9.4}$$

Where β is defined as the probability of each word $n \in 1 : N$ for topic z. The joint distribution can be specified as:

$$p(y, z, \theta, \phi|\alpha, \beta) =$$

$$\prod_{s=1}^{S} \int p(\theta_s, \alpha) \prod_{n=1}^{N} \sum_{z=1}^{K} p(z_{sn|\theta_s}) p(y_{sn}|z_{sn}, \beta) d\theta_s \tag{9.5}$$

We were interested in estimating the distributions of the parameter θ_s. Multiple activity composites were derived by LDA in each time window s, each activity composite being assigned a probability. For each time window we considered only the activity composite maximizing θ_s, indicated hereafter as z_s, the window's main activity composite.

9.3.3 Relevant activity composites

During the training phase, the HR for activity primitives and speeds was computed for each main activity composite z_s and participant par. Accelerometer features X_{acc} were used to estimate walking speed as $y_{speed} = X_{speed}\beta_{speed} + \epsilon, X_{speed} = \{X_{acc}, X_{ant}\}$. The resulting matrix HR_{ctx} is of dimension $K \times npar$, where K is the number of activity composites and $npar$ is the number of participants. LDA-derived activity composites do not include semantics and cannot be compared across participants. To overcome the problem of comparing activity composites, we characterized them with a set of features T which we used to rank activity composites, as in [5]. In order to provide a generalized method that is applicable to new participants, we chose features T that are independent of a person's lifestyle, for example, $T_1 \in T$ could be *the relative time spent sedentary in each activity composite* for the different participants. Regardless of what a person's lifestyle is, it will always be possible to order LDA-derived activity composites by feature T_1, e.g. *the relative time spent sedentary in each activity composite*. Then, HR_{ctx} was ranked by feature T_1, providing a way to investigate the relation between the HR in different activity composites and CRF, across participants. The ranking orders HR_{ctx} by values of T_1 from maximum to minimum. Since we are interested in highlighting commonalities across activities composites, ranked HR_{ctx} are smoothed by a moving average, resulting in $\overline{HR}_{ctx}$. As a result, we obtain an array of k ranked HR values per participant. We conclude the training phase by determining which feature in T maximizes Pearson's correlation between $\overline{HR}_{ctx}$ and CRF. We define the vector of correlations r_T for a set of TN features in a context *ctx*. Thus, for each context *ctx*, we have:

$$r_T = \{r_{rank_{T1}}, \ldots, r_{rank_{TN}}\}, \tag{9.6}$$

$$r_{rank_i} = r(\overline{HR}_{ctx_{par=\{1,\ldots,npar\},i}}, CRF_{par=\{1,\ldots,npar\}}) \tag{9.7}$$

Where r_{rank_i} is the correlation between the vector of contextualized HR $\overline{HR}_{ctx}$ and CRF, among all participants *par* for a feature T_i in a context *ctx*. The activity composite providing the highest correlation was selected, i.e. the first element of the $\overline{HR}_{ctx}$ vector across individuals and CRF, to determine which feature T_i results in activity composites most representative of CRF. Thus, the feature $T_i =$ max r_{Tctx} showing the highest correlation between $\overline{HR}_{ctx}$ and CRF is chosen to determine *relevant activities composites*.

As an example, we consider as contexts *ctx* walking at 5 km/h during activity composites with the maximum relative time spent sedentary, i.e. *relevant activity composites*, as shown in Fig. 9.5. We first determine the vector of k elements HR_{ctx}, representing the mean HR while walking at 5 km/h in each LDA-discovered activity composite. Then, HR_{ctx} are ranked based on the feature T_i maximizing the correlation on our training set (i.e. *the relative time spent sedentary in each activity composite*), to determine $\overline{HR}_{ctx*}$. The first element of the ranked and smoothed $\overline{HR}_{ctx}$ vector, is the contextualized HR $\overline{HR}_{ctx*}$, used as input for CRF estimation.

9.4 Evaluation study

9.4.1 Participants and data acquisition

Participants were 46 (21 male, 25 female) self-reported healthy individuals, age 24.7 ± 4.9 years, weight 68.6 ± 10.9 kg, height 172.8 ± 8.9 cm, BMI 22.9 ± 2.5 kg/m^2 and VO_2max 3020.8 ± 668.9 ml/min. Written informed consent was obtained, and the study was approved by the ethics committee of Maastricht University. The sensor platform used was an ECG Necklace, a platform configured to acquire one lead ECG data at 256 Hz, and three-axial accelerometer data at 32 Hz. The ECG Necklace was worn on the chest, close to the body's center of mass. The ECG Necklace was worn during laboratory protocols and free living. A *Continuous Wavelet Transform* based beat detection algorithm was used to extract R-R intervals from ECG data recorded under laboratory conditions, which output was manually examined to correct for missed beats that might be caused by motion artifacts [102]. For ECG data recorded under unsupervised free-living conditions, we selected high quality data by discarding periods in which more

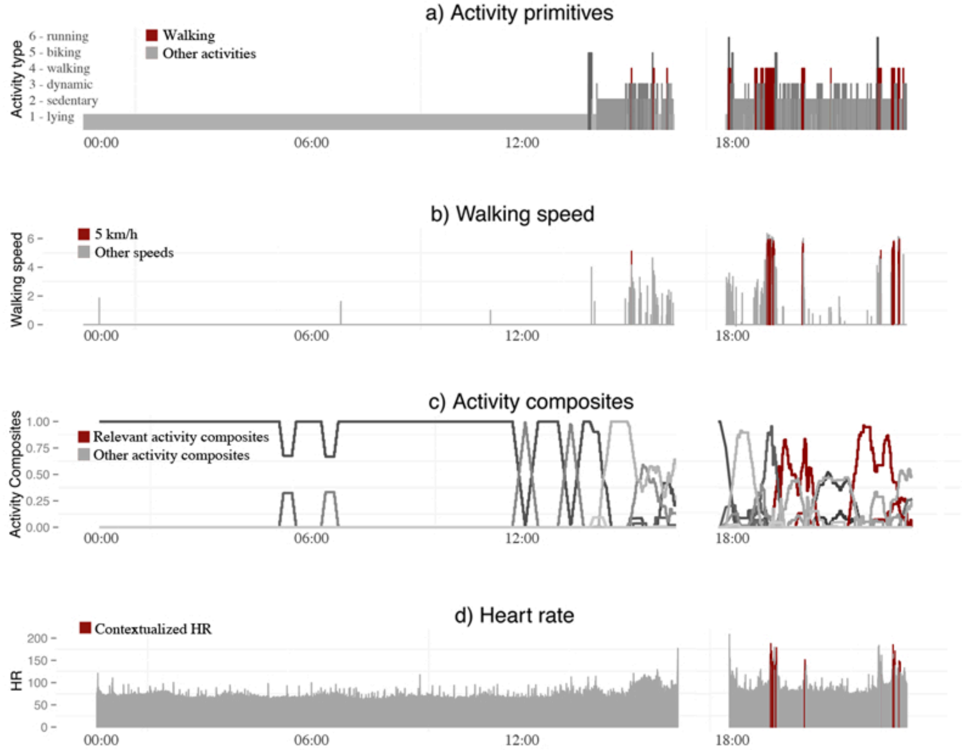

Figure 9.5: Exemplary diagram of the procedure to determine contextualized HR $\overline{HR}_{ctx*}$. Plots show 24 hours of free living data for one participant. For this illustration, we selected as activity primitive and speed *walking at* 5 *km/h during relevant activity composites*, and highlighted them in red. a) Recognized activity primitives, as detected by the combined SVM and HMM classifier. b) Walking speed y_s, determined when walking is detected, using a linear regression model. c) Activity composites determined by LDA and defined by the distribution of activity primitives and stay regions over 15 minutes windows. Relevant activity composites are determined using the procedure detailed in Sec. 9.3.3, maximizing the correlation between HR and CRF. d) Contextualized HR $\overline{HR}_{ctx*}$ is determined as the mean HR while walking at 5 km/h during *relevant activity composites* in this example, and highlighted in red. $\overline{HR}_{ctx*}$ is used to estimate CRF, as shown in Fig. 9.3. Between 17 and 18 hours no data are present since the sensor was being charged.

than 15% errors were detected in a time window. Errors were defined as consecutive RR intervals differing more than 20%, as typically reported in clinical practice. Additionally, during free living each participant carried a Samsung Galaxy S3 used to record GPS coordinates at 5 minutes intervals. Reference CRF was determined as VO_2max, by means of an incremental test on a cycle ergometer [78] using a indirect calorimeter that analyzed O_2 consumption and CO_2 production. The dataset considered for this work contains 507 days of data collected from 46 subjects in free living, thus about 11 days per participant, including accelerometer, ECG and GPS data. 75 hours of laboratory recordings including reference VO_2, VCO_2, acceleration, ECG and VO_2max were also obtained for model validation.

Figure 9.6: ECG Necklace and Samsung S3, the wearable sensor and phone used to collect accelerometer ECG and GPS data in this study. The ECG Necklace was worn during laboratory protocols and free living recordings close to the body's center of mass. The Samsung S3 was carried during free living only.

9.4.2 Experimental design and validation procedure

We collected data in free living and laboratory settings and evaluated four approaches to CRF estimation. All approaches were evaluated with respect to reference CRF measured by means of a VO_2max test carried out on a cycle ergometer. In the remaining of this paper, we will use the following terminology to characterize the four estimation conditions that were used for comparison; a) *anthropometrics*: no HR data was used, b) *no-context*: HR in free living was used directly to estimate CRF, c) *primitives*: HR in free living was contextualized using activity primitives and speed, d) *composites*: HR in free living was contextualized using activity primitives, speed and *relevant activity composites*.

Two laboratory protocols were designed and implemented for each participant on two separate days to avoid the maximal fitness test to affect physiological parameters during less intense activities and vice versa. Additionally, each participant wore the ECG Necklace in free living for 14 days. All results on CRF estimation were obtained from the free living data, whereas the laboratory data was used to derive the models, as detailed in the next Sections.

Data from laboratory protocols were used to develop supervised methods for activity type recognition and walking speed estimation. Activity type recognition and walking speed estimation models were deployed in free living and used as building blocks to contextualize HR. Additionally reference VO_2max was collected under laboratory protocols to validate the proposed CRF estimation models. Data collected in free living were used to determine contextualized HR and use contextualize HR as predictor for CRF estimation. CRF estimation models including contextualized HR as predictor relied on; laboratory-validated activity type recognition and walking speed estimation models, stay regions determined unsupervisedly in free living (see Sec. 9.5) and activity composites determined using LDA, in free living.

9.4.2.1 Laboratory protocols

Participants reported at the lab on three separate days and after refraining from drinking, eating and smoking in the two hours before the experiment. Two laboratory protocols were performed. The first protocol included simulated activities performed while wearing a portable indirect calorimeter. Activities included: lying down, sitting, sit and write, standing, cleaning a table, sweeping the floor, walking (treadmill flat at 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6 km/h) and running (treadmill flat at 7, 8, 9, 10 km/h). Activities were carried out for a period of at least 4 minutes. The second protocol was a VO_2max test providing reference data for biking and CRF. The third day was used for anthropometric measurements including the participant's body weight, height and body fat assessed using doubly labelled water [131].

9.4.2.2 Free living protocol

Participants worn the ECG necklace for 14 consecutive days in free living and manually annotated their activity composites in a paper diary. Participants were instructed to annotate activity composites as they occurred during the day and to annotate only high level activities such as going to work, sleeping, commuting, etc. Annotated activity composites were not used for model development since activity composites were derived using LDA, and therefore unsupervisedly from low level activity primitives, as detailed in Sec. 9.3 and Sec. 9.5. The annotations were only used to interpret the LDA and CRF estimation results as detailed in the discussion, Sec. 9.7. Activity composites can only be determined from free living data, since they cannot be simulated under laboratory conditions. Participants carried a Samsung S3 phone and were instructed to charge both the ECG Necklace and phone and to change electrodes daily.

9.4.2.3 Statistics and performance measures

All models were derived using leave-one-participant-out cross validation. The same training set, consisting of data from all participants but one, was used to build feature selection, activity recognition, walking speed estimation and CRF estimation models. The remaining data was used for validation. The procedure was repeated for each participant and results were averaged. LDA models were built on data from the participant to be validated, since no reference or training set are necessary. Performance of the activity recognition models was evaluated using the class-normalized accuracy, in laboratory recordings. Results for walking speed estimation and CRF estimation are reported in terms of Root-mean-square error (RMSE) and Pearson's correlation (r), where the outcome variables were speed in km/h and CRF in ml/min respectively. Paired t-tests were used to compare RMSE between models.

9.5 Implementation

9.5.1 Context recognition

9.5.1.1 Features

Accelerometer data from the three axes were segmented in 5 s windows, band-pass filtered between 0.1 and 10 Hz, to isolate the dynamic component due to body motion, and low-pass filtered at 1 Hz, to isolate the static component, due to gravity. Feature selection for activity type recognition was based on results from our previous work [9], using a different dataset. Selected features were: *mean of the absolute signal, inter-quartile range, median, variance, main frequency peak, low frequency band signal power*. Accelerometer features for walking speed estimation were: *mean of the absolute signal, inter-quartile range, variance, main frequency peak, high frequency band signal power*. HR was determined from RR intervals extracted from raw ECG data and averaged over 15 seconds windows. RR intervals were first correct for artifacts due to noise or ectopic beats, by removing consecutive intervals differing by more than 20%.

9.5.1.2 Activity primitives

Laboratory activities were grouped into six clusters to be used for classification of activity primitives. The six clusters were *lying* (lying down), *sedentary* (sitting, sit and write, standing), *dynamic* (cleaning the table, sweeping the floor), *walking*, *biking* and *running*. Activity primitives were derived combining a SVM and HMM. For the SVMs, we used a Gaussian radial basis kernel (cost function parameter C = 1). Parameters were set based on previous work from our group [6]. The HMM is defined by parameters $\lambda = (\pi, A, B)$; where π are the initial state probabilities, A is the transition probability matrix, defining the probability of transitioning between one activity to the other at time interval t. The HMM states corresponded to activity primitives. B is the emission matrix, which defines the probability of getting an emission at time t, given the state. We implemented the emission matrix B as $b_{ij} = 0.5 \iff i = j$, $b_{ij} = 0.1 \iff i \neq j$, while transitions probabilities A between actual states were derived from training data. Training data was the SVM classification result obtained with reference activity primitives manually annotated in laboratory settings.

9.5.1.3 Walking speed

Walking speed was estimated using a multiple regression model using as predictors the features listed in Sec. 9.5.1.1, together with the participant's *height*. Laboratory recordings on a treadmill while walking at different speeds were used to build subject-independent walking speed models.

9.5.1.4 Stay regions

Stay regions were computed from GPS coordinates according to time and distance thresholds, which were set to 60 minutes and 1 km according to previous literature [135]. The time threshold ensures that each stay region is a location where the participants spent a significant amount of time, while the distance threshold ensures that noisy recordings do not result into a multitude of stay regions being detected. GPS data was collected at 5 minutes intervals to conserve battery power. The relatively wide distance and time thresholds were chosen due to the low frequency of the GPS recordings.

9.5.1.5 Relevant activity composites

Input primitives for LDA were occurrences histograms of stay regions and activity primitives in each time window s. LDA hyperparameter α was set to 0.01, while segment size and number of topics k were set to 15 minutes and 20 topics respectively, based on results obtained in previous research [112]. Parameters were optimized using an implementation of the variational expectation-maximization algorithm proposed in [27]. HR during activities composites HR_{ctx} was ranked according to different features T: *amount of time spent in each activity composite*, *relative amount of time spent in each low level activity for an activity composite, with respect to the total time spent in the same low level activity across all activities composites* and *relative time spent in each low level atomic activity per activity composite*. These features were chosen since they can be computed across participants and activities composites regardless of the participant lifestyle or activity composite semantics. Ranked HR_{ctx} were correlated with CRF to determine which activities composites features were more representative of CRF. Ranking of HR_{ctx} values was smoothed by a moving average of 2 elements, i.e. over the first two ranked activity composites. The relevant activity composites discovery procedure was also evaluated independently of the participant. Contextualized HR HR_{ctx} was ranked and correlated with CRF for $np - 1$ participants. The feature resulting as the most representative of CRF, i.e. the one for which correlation was maximized, was used to determine relevant activity composites for the left out participant. The procedure was repeated np times, where np was the number of participants.

9.5.2 CRF estimation

Hierarchical Bayesian models for CRF estimation introduced in Sec. 9.3 were implemented using R and JAGS. Posterior parameters estimations were performed

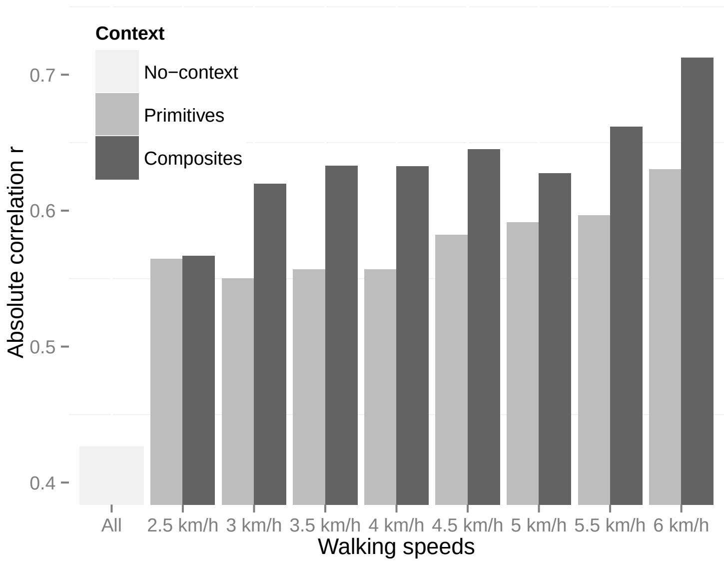

Figure 9.7: Correlation between HR and VO_2max. Correlation is lowest for *No-context* and were highest when activity composites (*Composites*) were used, compared to the condition were only activity primitives (*Primitives*) were considered. HR data during activity primitives and composites was acquired in free living settings.

by Gibbs sampling with 3 chains and 10000 iterations. The first 500 iterations were discarded (burn-in period). We consider reference VO_2max as CRF. We chose *walking* at different speeds as activity primitives normally carried out by most of the population. We evaluated our VO_2max estimation models using as predictor HR contextualized over a broad range of walking speeds, from 2.5 to 6 km/h. The hierarchical Bayesian model to estimate CRF also included the participant's weight, age, sex and height as predictors. We implemented the models listed in Sec. 9.4 for comparison, thus estimating VO_2max using anthropometric characteristics only (case *anthropometrics*), HR in free living (case *no-context*), HR while walking at a certain speed (case *primitives*), and HR while walking at a certain speed *relevant activity composites* (case *composites*).

9.6 Results

9.6.1 Activity primitives and walking speed

Activity primitives and walking speed were validated in laboratory settings. Class-normalized accuracy of the SVM-HMM activity recognition classifier was 95.8%.

More specifically, accuracy was 98.2% for *lying*, 98.9% for *sedentary*, 83.5% for *dynamic*, 99.4% for *walking*, 96.5% for *biking* and 98.4% for *running*. Walking speed estimation RMSE was 0.37 km/h.

9.6.2 Relevant activity composites

Fig. 9.7 shows the absolute value of the correlation between HR and VO_2max for different contexts. HR in free living was moderately correlated with VO_2max (comparison case *no-context*, $r = -0.43$). Correlation between HR and VO_2max in free living was stronger for *walking* activity primitives, compared to no-context, ranging from -0.55 to -0.63. Correlation had a tendency to increase as speed increased, reaching the highest value for *walking* at 6 km/h. Fig. 9.8 shows results for VO_2max estimation models. RMSE between estimated and predicted VO_2max when no HR data was used (*case anthropometrics*) was 322.5 ml/min. The relation between contextualized HR $\overline{HR}_{ctx}$ (i.e. including *relevant activity composites*) and VO_2max was maximized ranking activities composites by feature T_i = *relative time spent sedentary within an activity composite*. Correlation ranged between -0.57 and -0.71, reaching the highest value for *walking* at 6 km/h. Thus, correlation was consistently improved when a combination of activity primitives and *relevant activity composites* was used to contextualized HR, compared to no-context and activity primitives only, as shown in Fig. 9.7.

9.6.3 CRF estimation

RMSE was reduced to 286.3 ml/min (11.3% error reduction) when including free living HR as predictor but no contextual information (case *no-context*). Estimation error was further reduced for case *primitives*, i.e. using the HR while walking at a certain speed as predictors. More specifically, RMSE varied between 287.3 and 267.6 ml/min, depending on walking speed. RMSE was reduced by 17.0% and 6.5% compared to *case anthropometrics* and *no-context* respectively, when the best model was used (i.e. walking at 6 km/h). Contextualizing HR by a combination of activity primitives and activity composites provided better accuracy than any other model. RMSE varied between 268.9 ml/min and 249.5 ml/min, depending on walking speed. A combination of activity primitives and activity composites always outperformed activity primitives alone, as shown in Fig. 9.8.

Activity primitives in free living were recognized as follows: 44.5% lying, 36.4% sedentary, 9.5% dynamic, 5.4% walking, 3.8% biking and 0.4% running. The average walking speed in free living over the entire dataset was 3.5 ± 1.5 km/h. Participants spent 71 ± 27 minutes per day in walking activities, 7 ± 5.4 minutes walking at 6 km/h.

Overall, combining activity primitives and activities composites provided error reductions up to 22.6%, 12.8% and 10.3% compared to *anthropometrics*, *no-context* and *primitives* respectively. Fig. 9.9 shows estimated and measured VO_2max for the four models compared in this study. Explained variance (R^2) and RMSE are reported, showing increased R^2 and reduced error as more context is included.

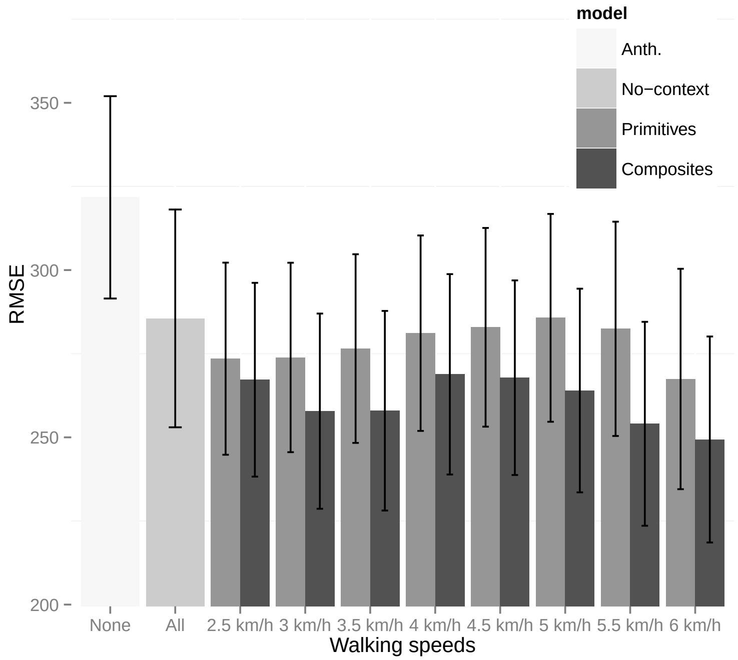

Figure 9.8: RMSE of CRF estimation in free living against VO_2max reference. Error bars represent standard error. RMSE is highest for *Anth*, followed by *No-Context*, showing that not using HR data or using HR data without context produces larger errors in VO_2max estimation. A combination of activity primitives and activity composites (condition *Composites*) shows optimal results, i.e. the lowest RMSE across different walking speeds, compared to the condition were only activity primitives (*Primitives*) were considered. HR data used as predictors was acquired during activity primitives and composites performed unsupervisedly in free living settings.

For the latter figure, only the best performing models is shown for cases *primitives* and *composites*.

9.7 Discussion

Many methods have been developed to estimate VO_2max using data collected under supervised laboratory conditions or following strict protocols. However, to the best of our knowledge, this is the first work, which combines activity primitives and activity composites to include both low and high level contextual information when interpreting HR data in free living. We showed RMSE reductions of

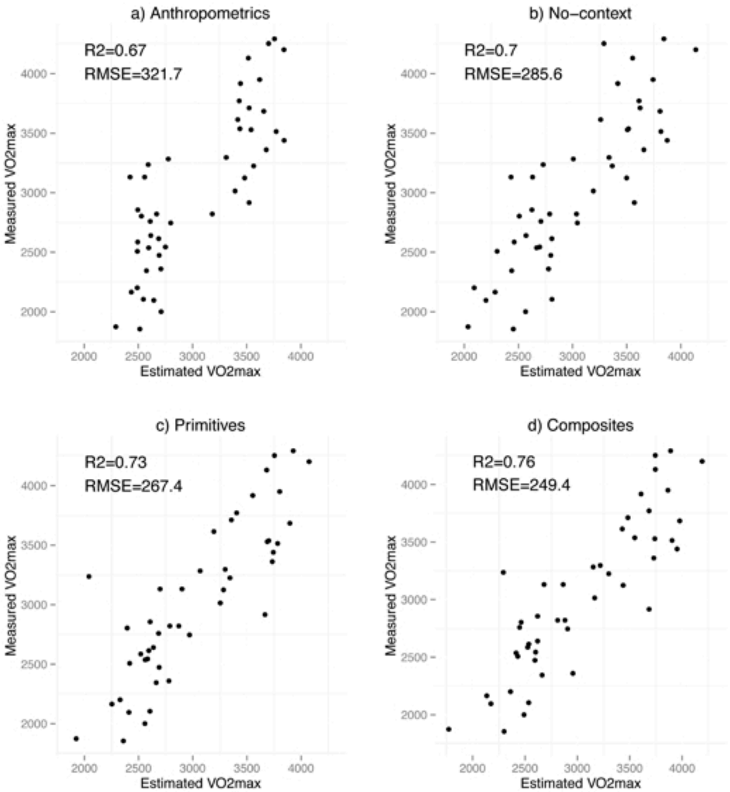

Figure 9.9: Estimated and measured VO_2max for the four conditions compared in this work. R^2 is increased and RMSE is reduced when adding more levels of contexts. The best results when VO_2max is estimated using HR contextualized by activity primitives and composites, as shown in d). HR data used as predictors was acquired during activity primitives and composites performed unsupervisedly in free living settings.

22.6% compared to estimates derived using anthropometric characteristics only, and RMSE reductions up to 10.3% compared to estimates derived using activity primitives.

We hypothesized that the presence of a multitude of stressors in free living required a novel approach over the prior estimation attempts used in laboratory settings. In particular, HR in free living is not only affected by low level activity primitives - as shown in the lab - but by both low level activity primitives and high level activity composites. Thus, incorporating knowledge of contextual information beyond activity primitives could potentially improve interpretation of HR in free living. Our results confirm the importance of activity composites in free living. RMSE was consistently reduced over a broad range of walking speeds, as shown in Fig. 9.8. We translated the need for high level contextual information into a hierarchical framework. In our previous work we introduced *relevant activity composites* for energy expenditure estimation [5]. We established *relevant activity composites* to relate discovered activity composites for which no supervised information exists, to behaviour-related HR.

In this work, discovered activity composites were ranked according to a correlation of selected features and HR. We could thus determine, which activity composites are better suited for CRF estimation. Relying on LDA or TMs in general provides the advantage of discovering activity composites which are individual for each participant, depending on their lifestyle. However, discovered activity composite do not provide semantics and comparison between participants is challenging. Typically, activity composite of interest are isolated and further classified using supervised methods [72, 112], thus requiring prior knowledge of the activity composites to discover, effectively limiting the unsupervised nature of the method. Ranking allowed for comparison of activity composite specific features across participants, thus making the approach unsupervised and generalizable to new participants.

We found a strong relation between the *relative time spent sedentary in each activity composite* and CRF. A possible explanation for the relation between HR contextualized by activity composites ranked by *relative time spent sedentary in each activity composite* and CRF is that activities composites in which people spend most of their time sedentary are typically representative of a stable physiological condition, which might be more representative of their CRF level. On the contrary, short or infrequent activities might involve more stressful situations as well as more intermittent HR, causing cardiovascular responses which are not as reliable for HR interpretation [101]. An example of an activity composite that maximizes the relative time spent sedentary is *working at the office*. While most of the time while working at the office an individual is probably sedentary, there can still be many periods of walking, that are therefore used to contextualize HR. In such periods, HR might be less affected by for example carrying loads, effects of previously performed intense exercise, walking hills, etc.) and therefore be more representative of CRF.

Sartor et al. [111] recently reviewed over ninety different VO_2max estimation tests. Most of these tests require intense activities and strict protocols, for example the most commonly used 2-mile run ($R^2 = 0.81$, [88]) or YMCA protocol ($R^2 = 0.56$, [110]). Our free living estimation falls on the high end of the reported R^2 for laboratory based submaximal tests, with accuracy close to what is reported for intense exercise protocols ($R^2 = 0.76$). However, the approach proposed in this work does not require intense activities, and is therefore suitable on a wider population. Additionally, the proposed approach does not require a specific test or laboratory infrastructure, and therefore VO_2max could be continuously assessed longitudinally over time, and not only re-assessed when the test is performed. Other studies investigated the relation between easily accessible measures such as HR or HR variability at rest and VO_2max [55]. However, these studies typically reported low levels of accuracy (R^2 = 0.29, [55]), showing that single measurements or spot measurements of physiological parameters and limited levels of context are insufficient for a reliable VO_2max estimate. A possible explanation for the better performance of the proposed approach compared to single spot checks and even some more intense laboratory protocols, is that by contextualizing HR over multiple days, our proposed approach is less prone to the day-to-day variability typical of physiological measurements as well as activity behavior.

We relied on the inverse relation between HR at a certain workload and VO_2max, as often reported for laboratory protocols. However, by using a non-nested hierarchical approach, where parameters varied based on the activities, we did not constrain the participant in performing specific activities or walking at predefined speeds. Instead, based on the participant's preferred walking speed in free living, the optimal parameters were used. The reason for using as predictor the HR while walking instead of the HR during other detected activities is that walking is the highest intensity activity that can be accurately quantified in free living, in terms of both activity type (i.e. *walking*) and intensity (i.e. *speed*). Additionally, *walking* is an activity basically everyone performs daily. On our free living dataset, participants spent more than an hour per day walking (71 ± 27 minutes), and about 10% of walking activities involved walking at 6 km/h (7 ± 5.4 minutes). Thus, *walking* confirmed to be a common activity of daily life, and a good candidate to contextualize HR for CRF estimation. Noteworthy, RMSE for VO_2max estimation was not consistently reduced by including in the models HR collected while walking at higher speeds. Thus, highlighting the additional complexity of analyzing HR data in free living.

Being able to accurately determine the user context in terms of activity type and intensity allows us to bring the principle used in laboratory based submaximal tests (i.e. the inverse relation between HR measured while performing an exercise at a certain intensity, such as biking at a fixed power on a cycle ergometer, and VO_2max) to free living settings. Contextualizing HR by means of low level activity primitives and speed improved correlation between free living HR and CRF. Combining activity primitives and activity composites further improved correlation. As a result, RMSE for CRF estimation against VO2max reference was reduced up to 22.6%.

9.8 Conclusion

In this paper, we estimate CRF in free living using data acquired with wearable sensors and mobile phones. We showed that RMSE for CRF estimation can be reduced up to 22.6% by including context. We obtained best results when including both low level activity primitives and high level activity composites, determined with a context recognition framework combining both supervised and unsupervised methods. CRF is a strong and independent predictor of all-cause and in particular cardiovascular mortality. The proposed CRF estimation model could be used to provide accurate information about an individual's health without the need for laboratory infrastructure or specific tests. New opportunities for applications targeting behavioral change by creating a feedback loop involving objectively measured PA level, as well as changes in CRF and associated reduced risk of disease, could be developed building up on the proposed approach.

Acknowledgment

The authors would like to thank G. Plasqui, G. Schiavone, G. ten Velde and S. Camps for support during data collection.

10

Cardiorespiratory fitness estimation using wearable sensors data: analysis of context-specific submaximal heart rates

M. Altini, P. Casale, J. Penders, G. ten Velde, G. Plasqui, O. Amft
Submitted to the Journal of Applied Physiology

Abstract

In this work, we propose to use pattern recognition methods to determine submaximal heart rate (HR) during specific contexts, such as walking at a certain speed, using wearable sensors in free-living. We then use context-specific HR to estimate cardiorespiratory fitness (CRF) from data acquired in free-living, without the need for laboratory protocols. CRF of 51 *participants (*24 *male,* 27 *female) was assessed by a maximal exertion test (*VO_2max*). Participants wore a combined accelerometer and HR monitor during a laboratory based simulation of activities of daily living and for two weeks in free-living. Accelerometer output was processed to determine participants' activities and walking speeds and these were used as specific contexts in which HR was analyzed. First, HR while lying down and walking at predefined speeds in laboratory settings was used together with anthropometric characteristics in a multiple regression model to estimate CRF. Explained variance (*R^2*) was* 0.64 *for anthropometrics only, and increased up to 0.74 for context-specific HR (*0.73 *to* 0.78 *when including fat-free mass). We then developed activity recognition and walking speed estimation algorithms to determine the same contexts (i.e. lying down and walking at different speeds) in free-living. Context-specific HR in free-living was highly correlated with measurements obtained during laboratory protocols (Pearson's r =* $0.71 - 0.75$*). HR while lying down and walking at predefined speeds in free-living, as detected by pattern recognition methods, was used together with anthropometric characteristics in a multiple regression model to estimate CRF.* R^2 *for CRF estimation was*

0.65 *when anthropometrics data only was used as predictors, and increased up to* 0.77 *when including free-living context-specific HR (i.e. HR while walking at* 5.5 *km/h).* R^2 *varied between* 0.73 *and* 0.80 *when including fat-free mass among the predictors. Subject independent evaluation of CRF estimation models using free-living data showed reduced RMSE between* 354.7 *ml/min (anthropometrics only) and* 281.0 *ml/min when including context-specific HR as predictors* (21% *error reduction). We conclude that pattern recognition techniques can be used to contextualize HR in free-living and estimated CRF with accuracy comparable to what can be obtained with simulated activities in laboratory settings.*

10.1 Introduction

Cardiorespiratory fitness (CRF) is among the most important determinants of health and wellbeing, being a diagnostic and prognostic health indicator for patients in clinical settings, as well as healthy individuals. CRF can be adopted as a proxy of cardiovascular and cardiorespiratory health [81, 111]. While recent developments in wearable sensor technologies improved the accuracy of physical activity monitoring devices in daily life, almost all solutions focus on behavioral aspects such as steps, activity type and energy expenditure (EE) [29]. Steps or EE are relevant markers of an individual's health, however they mainly reflect the individual's behavior, instead of the individual's health status. CRF estimation using wearable sensors could provide more insights on an individual's health status, non-invasively, and therefore help clinicians and individuals coaching or leading a more health lifestyle.

Currently, the gold standard for CRF measurement is performed by direct measurement of oxygen consumption during maximal exercise (i.e. VO_2max) [124]. However, VO_2max measurements require medical supervision and can be risky for individuals where exercise till maximal exertion is contra-indicated. Despite the indubitable importance of CRF in health, measurements of VO_2max are therefore rare [94] and less risky submaximal tests have been developed. Non-exercise CRF estimation models use easily accessible measures such as age, gender and a self-reported physical activity level [73, 93]. However, for individuals with similar anthropometric characteristics, CRF levels cannot be discriminated accurately. Alternatively, submaximal tests have been introduced to estimate VO_2max during specific protocols while monitoring HR at predefined workloads [18, 54]. The strict workload imposed by the protocol is used to exploit the inverse relation between HR in a specific context (e.g. while running or biking at a specific intensity) and VO_2max. However the need for laboratory equipment and the necessity to reperform the test to detect changes in CRF limit the practical applicability of such techniques. Ideally, we would like to estimate CRF in free-living during activities of daily living, thus without the need for specific laboratory tests or exercise protocols.

Miniaturized wearable sensors combining accelerometer and HR data provide a way to investigate the relation between physical activity, HR and VO_2max in free-living. Additionally, advances in signal processing and machine learning techniques, recently provided new methods to accurately recognize contexts in which HR can be analyzed, such as activity type, walking speed and EE [9, 29, 118], in free-living. Preliminary work explored the relation between physical activity as expressed by a step counter, and CRF [40]. While steps could provide useful insights, the relation between HR and VO_2 at a certain exercise intensity cannot be exploited using motion based sensors. Plasqui et al. [99] showed that a combination of average HR and physical activity over a period of 7 days correlates significantly with VO_2max. However, the relation between average HR and activity counts depends on the amount of activity performed, and therefore could also be affected by behavioral correlates of CRF. Tonis et al. [120] explored different parameters to estimate CRF from HR and accelerometer data during activities of daily living simulated in laboratory settings. However VO_2max reference and free-living data were not collected. When moving towards free-living settings, HR is of greater difficulty to interpret, since activities vary depending on the different lifestyles people adopt. However, previous studies exploring the relation between VO_2max and HR in free-living, showed positive results.

In this study, we aimed at investigating the relation between submaximal HR in specific contexts as recorded by wearable sensors in free-living, and CRF. We first isolated the same contexts in laboratory settings and free living. Then we analyzed the relation between context-specific HR during activities simulated in the lab and context-specific HR as detected by pattern recognition methods deployed in free-living, by using correlation and relative differences in HR for each context. Finally, we used context-specific HR in free-living to estimate CRF. Our results showed that VO_2max estimation using as predictors context-specific HR in free living provides accuracy comparable with laboratory derived models.

10.2 Methods

10.2.1 Participants

Participants were 51 (24 male, 27 female) healthy adults. Anthropometric characteristics and CRF level are reported in Table 10.1. Written informed consent was obtained by each participant. The study was approved by the medical ethics committee of Maastricht University.

10.2.2 ECG and accelerometer device

The sensor platform used was an ECG Necklace. The ECG Necklace [98] is a low power wireless ECG platform. The system relies on an ultra-low-power ASIC for ECG read-out, and it is integrated in a necklace, providing ease-of-use and comfort while allowing flexibility in lead positioning and system functionality. It achieves up to 6 days autonomy on a 175 mAh Li-ion battery. For the current study, the ECG Necklace was configured to acquire one lead ECG data at 256 Hz,

Table 10.1: Participants' characteristics.

Parameter	Mean ± SD
n	51 (24 male, 27 female)
Age (y)	25.1 ± 6.0
Body weight (kg)	68.4 ± 10.8
BMI (kg/m^2)	22.7 ± 2.5
Fat free mass (kg)	52.6 ± 9.2
VO_2max (ml/min)	3037.5 ± 671.6

and accelerometer data from a tri-axial accelerometer (ADXL330) at 64 Hz. The ADXL330 accelerometer provides a ś3g range and high sensitivity (300 mV/g), and was digitalized to 12 bits input by the ECG Necklace. The x, y, and z axes of the accelerometer were oriented along the vertical, mediolateral, and antero-posterior directions of the body, respectively. The ECG Necklace was not attached to the body, to improve user comfort during free-living. Two gel electrodes were placed on the participant's chest, in the lead II configuration. Data were recorded on the on-board SD card to ensure no data loss.

The ECG Necklace was previously validated as a reliable physical activity monitor able to quantify different physical activity parameters with high accuracy, such as activity type, walking speed and EE [9, 14]. A continuous wavelet transform based beat detection algorithm was used to extract RR intervals from ECG data. Segments of data identified as lying or sedentary (no or limited movement) as well as flat ECG signal or non-realistic HR were treated as monitor not worn. Non-realistic HR was identified as periods where consecutive RR intervals varied more than 20%, as typically performed in clinical practice for heart rate variability analysis.

10.2.3 Study design

The ECG Necklace was worn during laboratory protocols and free-living.

10.2.3.1 Laboratory protocols

Participants reported at the lab on three separate days and after refraining from drinking, eating and smoking in the two hours before the experiment. Two laboratory protocols were performed, while the third day was used for anthropometric measurements including the participant's body weight and height.

- The first protocol included simulated activities performed while connected to an indirect calorimeter (Omnical, Maastricht University, The Netherlands) for reference EE. Activities included: lying down, sitting, sit and write, standing, cleaning a table, sweeping the floor, walking (treadmill flat at 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6 km/h) and running (treadmill flat at 7, 8, 9, 10 km/h). Activities were carried out for a period of at least 4 minutes.

- The second protocol was a VO_2max test providing reference data for biking and CRF. VO_2max was determined during an incremental test on a cycle ergometer according to the protocol of Kuipers et al. [78]. After a 5-min warm-up at 100 W for men and 75 W for women, workload was increased by 50 W every 2.5 min. When the HR reached 35 bpm below the age-predicted maximal HR ($208 \times 0.7 \times age$) or the respiratory quotient exceeded 1, workload was increased by 25 W every 2.5 min until exhaustion. Expired air was continuously analyzed for O_2 consumption and CO_2 production using indirect calorimetry.

10.2.3.2 Free-living protocol

Participants wore the ECG necklace for 14 consecutive days in free-living while carrying out their normal activities of daily living. Participants were instructed to wear the ECG necklace during day and night, except during showering, water activities or charging of the ECG necklace. Charging was performed daily for 1 hour. Participants were also instructed to change electrodes daily or after physical exercise.

10.2.4 Data processing

Context-specific HR in laboratory settings was determined as the mean HR during scripted activities performed by the participant and combined with anthropometrics in a regression model to predict VO_2max. The regression model was analyzed to validate the assumption that submaximal context-specific HR can be used to estimate CRF level. Activity type recognition and walking speed models were built using data from laboratory settings, and used in free-living. For each participant, models were built using only data from other participants. Therefore, all models were non-individualized and no laboratory data from the participant to be validated was used for model building. The procedure used for model building and evaluation is shown in Figure 10.1. More details on the validation procedures are reported in the Statistics and performance measures Section. Context-specific HR in free-living was used in a multiple regression model to estimate VO_2max without the need for laboratory protocols and analyzed with respect to results obtained using submaximal context-specific HR acquired during activities of daily living simulated in laboratory settings.

10.2.4.1 Activity type and walking speed

The raw acceleration signal was downloaded and processed for two purposes. The first purpose was to develop an activity recognition algorithm using data acquired during simulated activities of daily living in the laboratory protocols. The activity recognition algorithm was then used to detect the activity types performed during the free-living protocol. Secondly, the raw acceleration signal was processed to determine walking speed for activities recognized as walking. The acceleration signal was segmented in non-overlapping intervals of 5 seconds. This

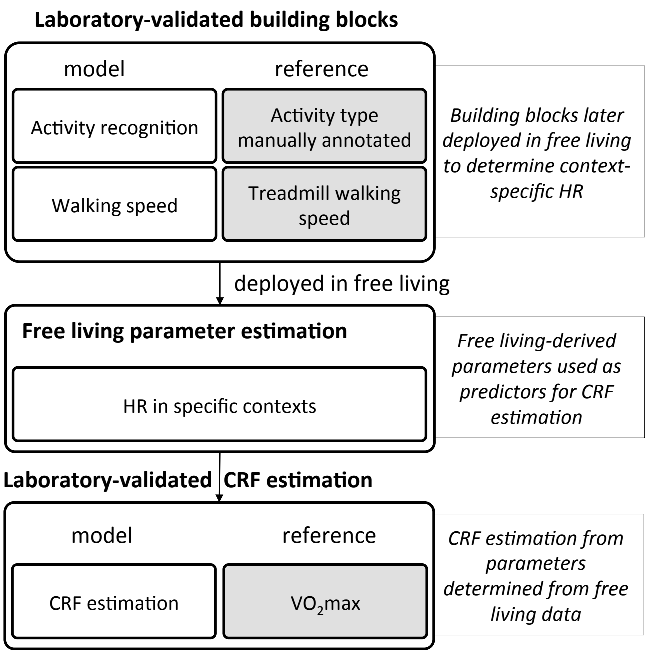

Figure 10.1: Block diagram of the proposed approach and validation procedure. Activity recognition and walking speed estimation models are built and validated using supervised laboratory recordings. Then, models are deployed in free-living. Activity recognition and walking speed estimation are used to determine HR in specific contexts in free-living. Finally, HR in specific contexts (e.g. HR while lying down or walking at a certain speed) are used as predictors for VO_2max estimation, effectively estimating CRF level from free-living data. All models are validated using leave one subject out cross validation, i.e. no data used for model validation was used for model building, as described in the Statistics section. An example of activity recognition and walking speed estimation models output is shown in Figure 10.3.

segment length was selected based on previous studies [118]. Segmented data was separately filtered by two filters to create different feature sets. One feature set included accelerometer data band-pass filtered between 0.1 and 10 Hz, to isolate the dynamic component due to body motion, while the second feature set included accelerometer data low-pass filtered at 1 Hz, to isolate the static component, due to gravity. Figure. 10.2 shows an example of raw data, low-passed data and band-passed data for one participant during one of the laboratory protocols. Features used for activity recognition were: mean of the absolute signal, inter-quartile range, median, variance, main frequency peak and low frequency band signal power. All accelerometer features but the median, were derived from band-pass filtered data. These features were derived and selected based on our previous work [9], using a different dataset. HR was extracted from RR intervals, and averaged over 15 seconds windows. Laboratory activities were grouped into six clusters to be used for activity classification. The six clusters were lying (lying down), sedentary (sitting, sit and write, standing), dynamic (cleaning the table, sweeping the floor), walking, biking and running. Activities were derived using a Support Vector Machine (SVM), a classifier showing good results in our previous research [9]. Features for the multiple linear regression model used to estimate walking speed were: mean of the absolute signal, inter-quartile range, variance, main frequency peak, high frequency band signal power and height of the participant, and were also based on our previous work [10]. All accelerometer features used for the walking speed models were derived from band-pass filtered data.

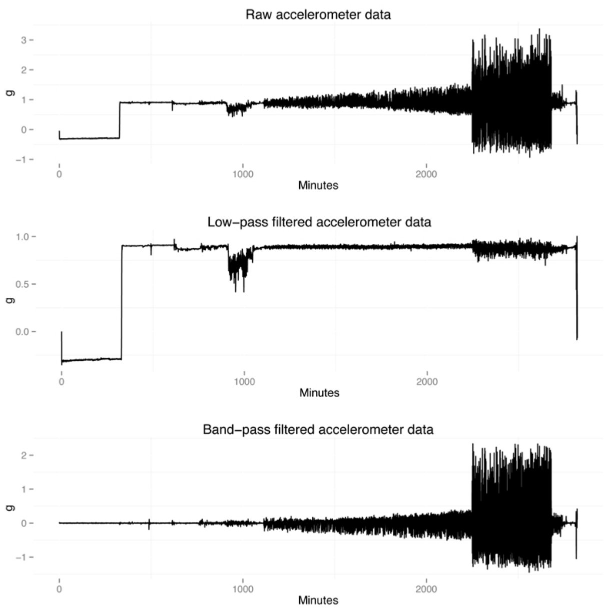

Figure 10.2: Raw accelerometer data (top), low-pass filtered data (center) and band-pass filtered data (bottom). The gravity component is isolated when using low-pass filtered data, as shown in the center plot. This information is particularly useful to distinguish postures. Band-pass filtered data isolates the accelerometer component due to body motion, showing increased values for higher intensity motions. Band-pass filtered data is particularly useful to distinguish ambulatory activities and walking speeds. Data were downsampled for visualization purposes.

10.2.4.2 CRF estimation

CRF was estimated using multiple linear regression models. First, we investigated the relation between HR in specific contexts as acquired during activities of daily living simulated in laboratory settings, and VO_2max. We predicted VO_2max by combining anthropometric characteristics and HR while lying down and while walking at 3.5 and 5.5 km/h. We chose lying down and walking at 3.5 and 5.5 km/h as specific contexts since lying down and walking are activities of daily living commonly performed by healthy individuals in most environments. Additionally, the average walking speeds in healthy individuals was reported in previous studies between 5 and 6 km/h (5.3 km/h in [36] and 5 ± 0.8 km/h in [89]).

Then, we analyzed the relation between context-specific HR during activities of daily living simulated in laboratory settings, and context-specific HR during the same activities as detected by our activity recognition and walking speed models, in free-living. The analysis of the relation between context-specific HR in laboratory settings and free-living consisted of computing the correlation coefficient and relative differences between HR in laboratory settings and free living. This analysis is merely to provide some perspective on context-specific HR with respect to laboratory measurements. However, free-living regression models are built and evaluated using free-living data only.

Finally, we predicted VO_2max by combining anthropometric characteristics and HR while lying down and while walking at 3.5 and 5.5 km/h as determined from free-living data, to evaluate the ability of the context-specific HR detected using pattern recognition methods to estimate CRF.

10.2.4.3 Statistics

Activity recognition and walking speed estimation models were derived using laboratory data and evaluated using leave-one-participant-out cross validation. The same training set, consisting of data from all participants but one, was used to build feature selection, activity recognition and walking speed estimation and CRF estimation models. The remaining data was used for validation. The procedure was repeated for each participant and results were averaged. Performance of the activity recognition models was evaluated using the class-normalized accuracy, using laboratory recordings. Results for walking speed estimation were reported in terms of Root-mean-square error (RMSE), where the outcome variable was speed in km/h. The relation between HR and CRF were reported using Pearson's correlation coefficient (r) for both activities simulated in laboratory settings and free-living data. The relation between context-specific HR during activities of daily living simulated in laboratory settings and in free-living as detected by pattern recognition methods was reported using Pearson's correlation coefficient (r) and the mean and standard deviation of the difference between context-specific HR in laboratory settings and in free-living. Results for CRF estimation models were reported in terms of explained variance (R^2). The Bland-Altman plot was used to determine the agreement between measured and predicted CRF. Finally, subject-independent evaluation for CRF estimation models was also performed, using leave one participant out cross-validation. We reported results for subject independent CRF estimation in terms of RMSE, where the outcome variable was VO_2max in ml/min as measured in laboratory conditions. Paired t-tests were used to compare results. Significance was set at $\alpha < 0.05$.

10.3 Results

10.3.1 Descriptive statistics

The dataset considered for this work contained 491 days of data collected from 51 participants in free-living, thus about 10 days per participant, including accelerometer and ECG data. Eighty-three hours of laboratory recordings including reference VO_2, VCO_2, acceleration, ECG and VO_2max were collected for model building and evaluation. Laboratory measurements were discarded for two participants where we observed measurement errors. Anthropometric characteristics and CRF level for the participants are reported in Table 10.1. Fig. 10.3 shows an exemplary output of the walking speed and activity recognition models for one participant during 24 hours of free-living recordings. Context-specific HR as identified using activity recognition and walking speed models in free-living is also shown in Fig. 10.3.

10.3.2 CRF estimation from context-specific submaximal HR during simulated activities of daily living

HR during activities of daily living simulated in laboratory settings was 66.2 $\pm$ 12.3 bpm for lying, 91.0 $\pm$ 15.3 bpm for walking at 3.5 km/h and 107.8 $\pm$ 17.7 bpm for walking at 5.5 km/h. Pearson's correlation between context-specific submaximal HR as measured during activities of daily living simulated in laboratory settings and CRF was -0.43 for lying down, -0.47 for walking at 3.5 km/h and -0.51 for walking at 5.5 km/h. Thus, confirming the hypothesis that submaximal HR is inversely related to CRF. Explained variance (adjusted R^2) for multiple regression models including sex, body weight and age as predictors of CRF, was 0.64. Adjusted R^2 increased when including context-specific HR, and was 0.69 for lying, 0.72 for walking at 3.5 km/h and 0.74 for walking at 5.5 km/h. Thus, confirming that activities of higher submaximal intensities explain more of the variance in the model. Results are reported in Table 10.2 while Fig. 10.4 shows scatterplots of reference against fitted values as well as Bland-Altman plots. When including more advanced anthropometrics, such as fat free mass instead of body weight, R^2 was 0.73 when no HR was used among the predictors, 0.74 for lying, 0.76 for walking at 3.5 km/h and 0.78 for walking at 5.5 km/h.

10.3.3 Context recognition; activity type and walking speed

Laboratory recordings with reference activity type were used to determine accuracy of the models used in free-living. Accuracy of the SVM activity recognition

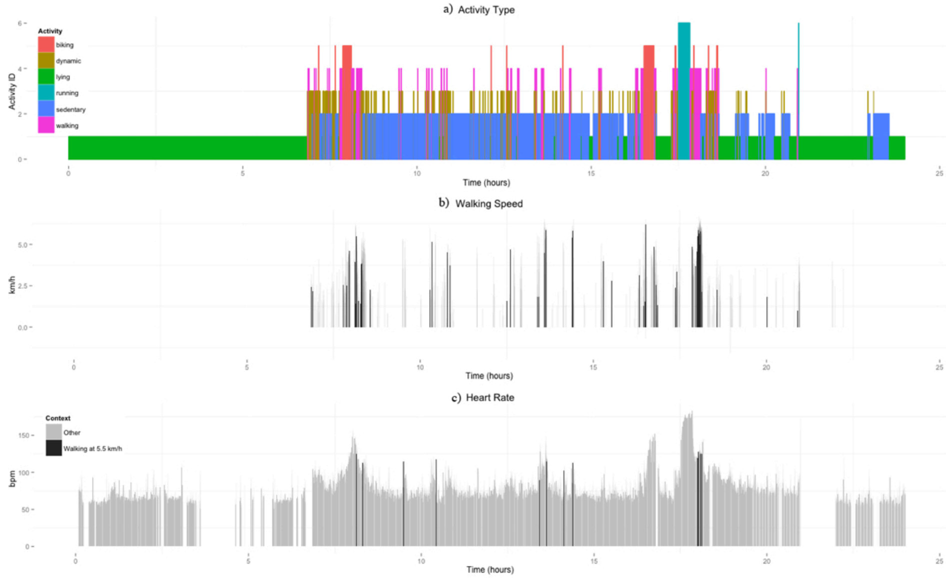

Figure 10.3: Exemplary output of the models used to contextualize HR in free-living in this work, for one participant. a) Recognized activity types. Commuting by bike, training (running), sleep and a mostly sedentary job during waking hours can be easily identified from this plot. b) Estimated walking speeds when the activity type algorithm identifies the walking activity. c) HR and contextualized HR. Contextualized HR, i.e. in this example the HR while walking at 5.5 km/h, is highlighted in black.

classifier was 94.1%. More specifically, the accuracy was 96.4% for lying, 95.6% for sedentary activities, 83.3% for dynamic, 98.2% for walking, 91.4% for biking and 99.7% for running. The explained variance for the walking speed model was 0.85 (R^2). Walking speed estimation RMSE for subject independent analysis was 0.37 km/h across all speeds. Activities in free-living over the complete dataset were recognized as follows: 44.4% lying, 36.4% sedentary, 9.5% dynamic, 5.4% walking, 3.8% biking and 0.4% running. Average walking speed was 3.6 $\pm$ 1.5 km/h. Participants spent on average 77.7 minutes per day walking, 11.9 minutes of which were at 3.5 km/h and 11.6 minutes of which were at 5.5 km/h.

10.3.4 Relation between context-specific submaximal HR during activities of daily living simulated in laboratory settings and in free-living

Pearson's correlation between context-specific submaximal HR measured during activities of daily living simulated in laboratory settings and in free-living as detected by pattern recognition methods was 0.71 for lying down, 0.71 for walking

Table 10.2: Multiple linear regression models for VO_2max estimation from activities of daily living simulated in laboratory settings.

Model description	**Predictors**	R^2
Anthropometric characteristics only	Body weight, age, sex	0.64
Context-specific HR	HR while lying down in laboratory settings, body weight, age, sex	0.69
	HR while walking at 3.5 km/h in laboratory settings, body weight, age, sex	0.72
	HR while walking at 5.5 km/h in laboratory settings, body weight, age, sex	0.74

at 3.5 km/h and 0.75 for walking at 5.5 km/h. Mean difference between context-specific HR in laboratory settings and free-living was 2.9 ± 8.7 for lying (mean HR while lying down was 63.2 bpm in free-living and 66.2 bpm in laboratory settings), 8.7 ± 11.2 for walking at 3.5 km/h (mean HR while walking at 3.5 km/h was 99.9 bpm in free-living and 91.0 bpm in laboratory settings) and -2.7 ± 11.5 for walking at 5.5 km/h (mean HR while walking at 5.5 km/h was 106.3 bpm in free-living and 107.8 bpm in laboratory settings). Thus, all differences were below 10%. Histograms of the differences and scatterplots of context-specific HR in laboratory settings and free-living are shown in Fig. 10.5.

10.3.5 CRF estimation from context-specific submaximal HR in free-living

HR during specific contexts in free-living was 63.2 ± 9.3 bpm for lying, 99.9 ś±11.6 bpm for walking at 3.5 km/h and 106.3 ± 11.8 bpm for walking at 5.5 km/h. Pearson's correlation between context-specific submaximal HR as measured in free-living and CRF was -0.54 for lying down, -0.52 for walking at 3.5 km/h and -0.60 for walking at 5.5 km/h. Thus, confirming the hypothesis that submaximal HR is inversely related to CRF. Adjusted R^2 increased from the case where no HR was included (R^2 = 0.65), when including context-specific HR. More specifically R^2 was 0.73 for lying, 0.74 for walking at 3.5 km/h and 0.77 for walking at 5.5 km/h. Thus, confirming that activities of higher submaximal intensities explain more of the variance in the model, even when carried out in free-living. Results for all models are reported in Table 10.3 and Bland-Altman plots for all models are shown in Fig. 10.6. When including more advanced anthropometrics, such as fat free mass instead of body weight, R^2 was 0.73 when no HR was used among the predictors, 0.77 for lying and 0.80 for walking at 3.5 km/h and 5.5 km/h.

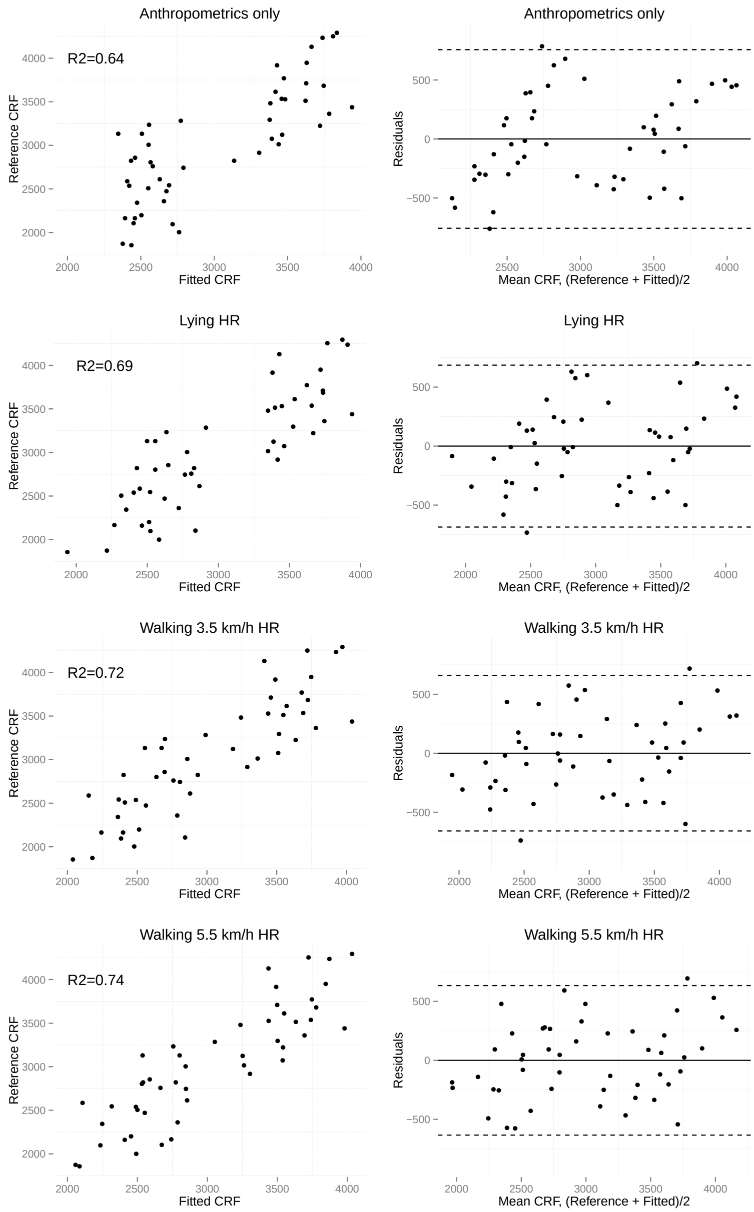

Figure 10.4: Accuracy of the prediction models for CRF estimation. Regression plots and Bland-Altman plots are shown for models using as predictors anthropometrics and context-specific HR during activities of daily living simulated in laboratory conditions. R^2 is also reported.

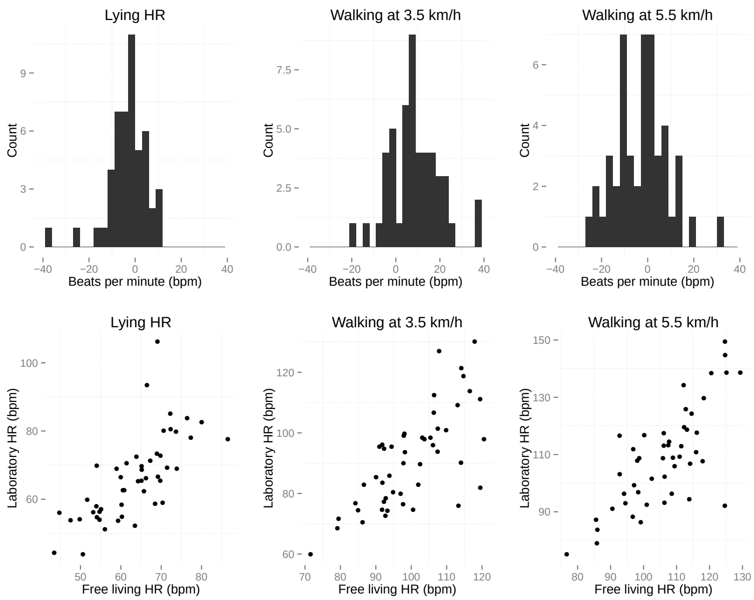

Figure 10.5: Top row: histograms of differences between context-specific HR in laboratory settings and free-living. Bottom row: scatterplots showing the relation between context-specific HR in laboratory settings and free-living.

10.3.6 Cross-validation of VO_2max estimates

VO_2max estimation models derived from free-living data were cross-validated using the leave-one-out technique. Results are shown in Fig. 7 and 8 and reported in Table 10.4 and 10.5. Cross-validation of VO_2max estimates using as predictors context-specific HR as measured during activities of daily living simulated in laboratory settings: RMSE for the model including anthropometric characteristics only as predictors was 358.3 ml/min (R^2 was 0.66). RMSE was reduced when including HR in specific contexts among the predictors, with RMSE = 314.3 ml/min (R^2 = 0.73) for lying down, RMSE = 310.0 ml/min (R^2 = 0.75) for walking at 3.5 km/h, and RMSE = 284.7 ml/min (R^2 = 0.78) for walking at 5.5 km/h as specific contexts. Thus, RMSE was reduced up to 21% when including context-specific HR among the predictors. Cross-validation of VO_2max estimates using as predictors context-specific HR as derived by pattern recognition methods in free-living: RMSE for the model including anthropometric characteristics only as predictors was 354.7 ml/min (R^2 was 0.67). RMSE was reduced when including HR in specific contexts among the predictors, with RMSE = 309.4 ml/min (R^2 = 0.75) for lying down, RMSE = 305.91 ml/min (R^2 = 0.76) for walking at 3.5 km/h, and RMSE = 281.0 ml/min (R^2 = 0.79) for walking at 5.5 km/h as specific free-living contexts. Thus, RMSE was also reduced up to 21% when including context-specific HR as

Table 10.3: Multiple linear regression models for VO_2max estimation from free-living data.

Model description	**Predictors**	R^2
Anthropometric characteristics only	Body weight, age, sex	0.65
Context-specific HR	HR while lying down in free-living, body weight, age, sex	0.73
	HR while walking at 3.5 km/h in free-living, body weight, age, sex	0.74
	HR while walking at 5.5 km/h in free-living, body weight, age, sex	0.77

Table 10.4: Cross validation of multiple linear regression models for VO_2max estimation using as predictors context-specific HR as measured during activities of daily living simulated in laboratory settings.

Model description	**Predictors**	**RMSE ml/min**	R^2
Anthropometric characteristics only	Body weight, age, sex	358.3	0.66
Context-specific HR	HR while lying down in laboratory settings, body weight, age, sex	314.3	0.73
	HR while walking at 3.5 km/h in laboratory settings, body weight, age, sex	310.0	0.75
	HR while walking at 5.5 km/h in laboratory settings, body weight, age, sex	284.7	0.78

determined from pattern recognition methods, among the predictors.

10.4 Discussion

In this work, we proposed a method to estimate VO_2max in free-living, without the need for laboratory tests or specific protocols. While many methods have been developed to estimate VO_2max using data collected under supervised laboratory conditions or following strict protocols, limited work tried to estimate CRF using wearable sensors and data collected under unsupervised settings in free-living [40, 99]. We adopted pattern recognition techniques to determine specific contexts, e.g. low intensity activities of daily living such as lying down and walking at predefined speeds, to contextualize submaximal HR without the need for a

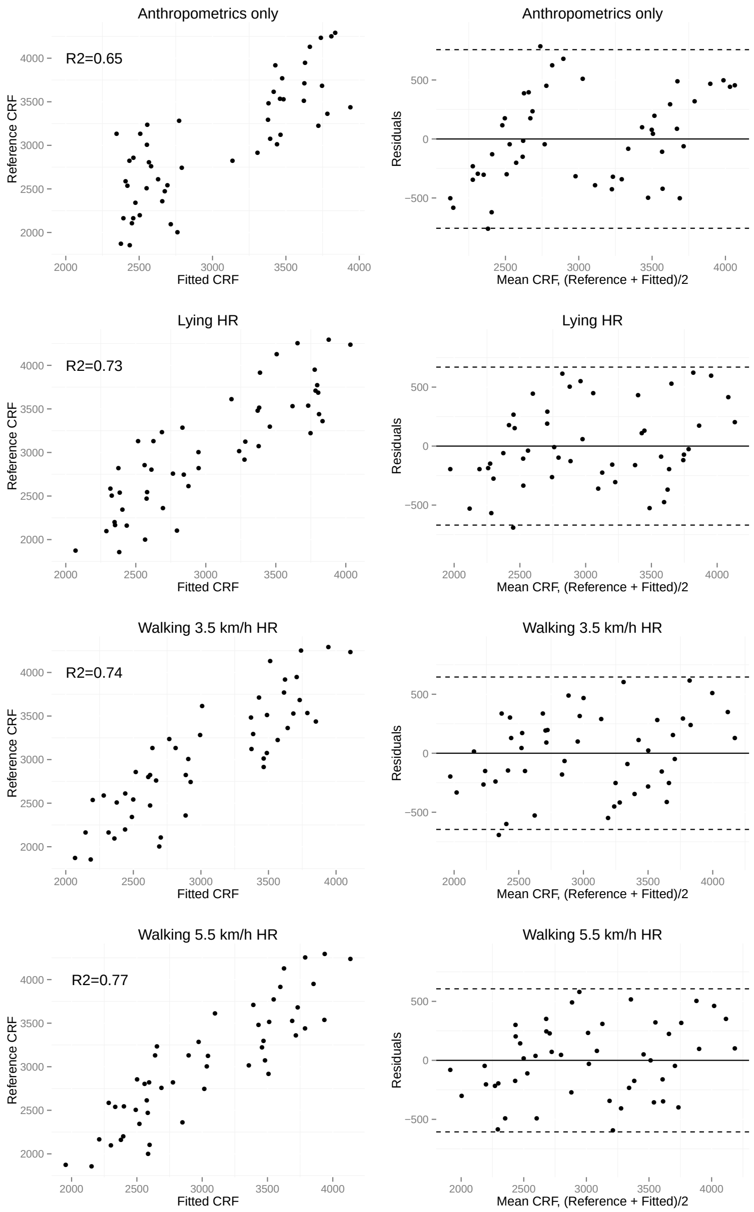

Figure 10.6: Accuracy of the prediction models for CRF estimation. Regression plots and Bland-Altman plots are shown for models using as predictors anthropometrics and context-specific HR in free-living. R^2 is also reported.

Table 10.5: Cross validation of multiple linear regression models for VO_2max estimation using as predictors context-specific HR as detected by pattern recognition methods in free-living.

Model description	**Predictors**	**RMSE ml/min**	R^2
Anthropometric characteristics only	Body weight, age, sex	354.7	0.67
Context-specific HR	HR while lying down in free-living, body weight, age, sex	309.4	0.75
	HR while walking at 3.5 km/h in free-living, body weight, age, sex	305.9	0.76
	HR while walking at 5.5 km/h in free-living, body weight, age, sex	281.0	0.79

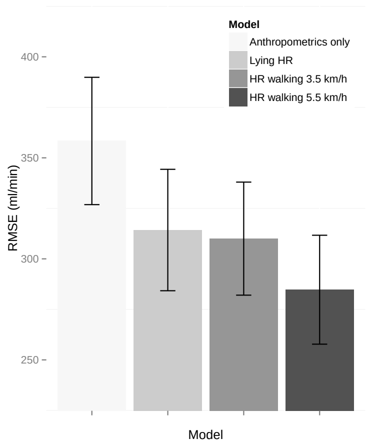

Figure 10.7: RMSE for subject independent cross validation of the CRF estimation models using context-specific HR as measured during activities of daily living simulated in laboratory settings. Error bars represent standard error. RMSE is reduced including context-specific HR next to anthropometrics, with lower error shown for higher intensity activities.

strict exercise protocol. We first validated the effectiveness of submaximal context-specific HR as a predictor of VO_2max during activities of daily living simulated in laboratory settings. Then we analyzed the correlation and relative differences between context-specific HR during activities simulated in the lab and context-specific HR as detected by pattern recognition methods deployed in free-living. Finally, we used context-specific HR in free-living to estimate CRF. Our results showed that VO_2max estimation using as predictors context-specific HR in free

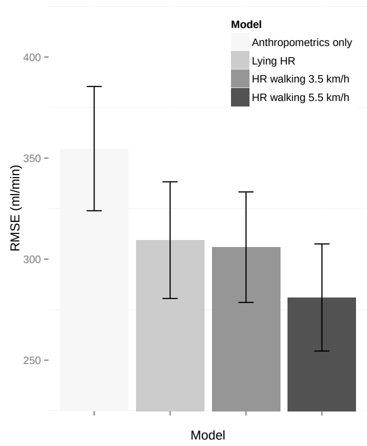

Figure 10.8: RMSE for subject independent cross validation of the CRF estimation models using context-specific HR in free-living. Error bars represent standard error. RMSE is reduced including context-specific HR next to anthropometrics, with lower error shown for higher intensity activities.

living provides accuracy comparable with laboratory derived models. Our results confirm our assumptions, showing that RMSE for VO_2max estimation could be reduced up to 21% compared to anthropometric characteristics only, by using as predictors HR in specific contexts as determined by pattern recognition methods in free-living.

10.4.1 Context-specific HR during activities of daily living simulated in laboratory settings

The main assumption behind this study was that submaximal HR is inversely related to VO_2max, and that the correlation is higher during submaximal activities of higher intensity. Our laboratory recordings confirm this assumption. Pearson's correlation between context-specific HR and VO_2max went from -0.43 to -0.51 for lying and walking activities. Multiple regression models showed higher explained variance (R^2 between 0.64 and 0.74) when including context-specific HR. Increasing activity intensity, i.e. from lying to slow walking (3.5 km/h) to faster walking (5.5 km/h) further improved R^2. These results are in agreement with a significant body of literature relying on submaximal HR for VO_2max estimation during more intense activities, such as biking or running, compared to the low intensity activities used in this study [132].

10.4.2 Context recognition in free-living

We deployed activity recognition and walking speed estimation algorithms in free-living, in order to contextualized submaximal HR without the need for strict exercise protocols or laboratory tests. Our activity recognition model showed high accuracy in detecting lying and walking activities (96.4-98.2%), given the peculiar accelerometer fingerprints of such activities, characterized either by different accelerometer orientation with respect to other activities or very specific repetitive movements. The activities chosen as free-living contexts were lying down and walking, for the following reasons. First, those are common activities performed by healthy individuals in most environments. Secondly, the inverse relation between HR at rest or sleeping HR and CRF was already shown in previous research, highlighting how this parameter can be valuable for VO_2max estimation. Finally, walking activities can be discriminated in intensity, by detecting walking speed, using simply an accelerometer. This is an important factor when trying to detect specific context in free-living, since detecting only activity type, if the activity can be carried out at different intensities, would not be sufficient to determine the same context for each individual. However, walking is an activity that can be accurately quantified in terms of both type (i.e. walking) and intensity (i.e. speed). The proposed activities are low intensity and were performed daily by the participants involved in our study, as shown by the analysis of free-living data. Our study population spent on average 44.4% of the free-living time lying down and 5.4% of the free-living time walking. Of the time spent walking, 11.9 minutes daily were spent at 3.5 km/h, while 11.6 minutes daily were spent at 5.5 km/h, the two speeds used by our models to contextualize HR. Considering that many fitness tests require protocols shorter than 11 minutes (e.g. the common 6-minutes walking test), we believe a total of 10 minutes daily is a sufficient amount of data for prediction of VO_2max, at least in the population of healthy adults considered in this study. We could evaluate activity recognition and walking speed models only under laboratory conditions, where reference was present. The dynamic activity cluster was recognized with accuracy below average. We interpret that activities with high variability in movement and execution between participants and using a single chest-worn sensor resulted in higher classifier confusions. However, the high accuracy of walking speed estimation models and activity recognition for walking provide confidence for the free-living detection of activities used to contextualize HR. Additionally, from the cross-validation analysis results we can see how subject independent models built using activities of daily living simulated in laboratory settings (RMSE were 314.3 ml/min, 310.0 ml/min and 284.7 ml/min for lying, walking at 3.5 km/h and walking at 5.5 km/h were respectively) are similar to RMSE results obtained contextualizing HR using pattern recognition methods in free-living (309.4 ml/min, 305.9 ml/min and 281.0 ml/min for lying, walking at 3.5 km/h and walking at 5.5 km/h respectively). These results can serve as indirect validation of the accuracy of activity recognition and walking speed estimation in properly detecting the relevant contexts in free-living.

10.4.3 Context-specific HR in free-living

Context-specific HR in free-living showed relations with VO_2max similar to what we reported in laboratory settings. The inverse relation between HR at a certain workload and VO_2max is the key principle behind laboratory based submaximal CRF tests and this relation showed to be valid not only in laboratory settings but also in free-living as well. The correlation between HR while lying down in free-living and VO_2max was -0.54 and it was increased up to -0.60 when the HR while walking at 5.5 km/h in free-living was used. Explained variance also increased, between 0.65 when anthropometrics characteristics only were used to estimate VO_2max, and 0.77 when using context-specific HR. We also analyzed the relation between HR during the same activities carried out in laboratory settings and free-living. We expected differences in HR due to the different settings, e.g. walking in free-living might include carrying weights, walking on inclined surfaces, or other factors that might raise HR, while lying down in laboratory settings might be more stressful than sleeping, therefore lowering HR with respect to laboratory conditions. Additionally, a single laboratory measurement might be affected by factors such as the previous days physical activity, while free-living recordings averaged over multiple days might provide more stable representations of a participant's physiology. However, analyzing the relation between laboratory and free-living HR in the same contexts can be useful to determine to what extent laboratory recordings can be reproduced in free-living as well as the ability of pattern recognition methods to detect differences between contexts such as lying down or walking at different speeds, in unsupervised free-living conditions. The relatively high correlation between laboratory and free-living HR (0.71-0.75), as well as similar mean values and consistent differences between conditions (i.e. higher HR for walking at higher speed, or higher intensity, in our case HR for laboratory activities and free-living was 66.2 bpm and 63.2 bpm for lying, 91.0 and 99.9 for walking at 3.5 km/h and 107.8 and 106.3 for walking at 5.5 km/h) are all promising results that free-living data can be used as a reliable substitute of laboratory recordings for context-specific submaximal HR.

10.4.4 Fat free mass

Analysis of VO_2max estimation including fat free mass instead of body weight among the predictors resulted in higher accuracy, as expected and previously shown in literature [99]. In particular, R^2 was increased between 0.74 and 0.78 for laboratory based measurements and between 0.77 and 0.80 for context-specific HR determined in free-living. However, since the aim of our work is to provide VO_2max estimation outside of the laboratory environment, we focus on simple anthropometrics only (i.e. body weight, age and sex) in the remaining of our discussion.

10.4.5 Cross-validation of VO_2max estimates

We also performed cross validation using subject independent models for VO_2max estimation as our aim was to validate the proposed methods using state of the art techniques able to validate the model on unseen data. Results for cross validation were consistent with what was shown before. Our results confirm that when estimating CRF, the individual's anthropometric characteristics are not sufficient to provide an accurate estimate. Differences in CRF among participants with similar body size (e.g. similar body weight and height) are not distinguishable if no physiological data is used in the models. Thus, the lower RMSE showed by VO_2max estimation models including HR as predictor shows the ability of submaximal context-specific HR to discriminate between such participants with similar anthropometric characteristics and further reduce VO_2max estimation error. As expected, contextualizing HR using more intense activities, such as walking at 5.5 km/h instead of lying, provides better results. It is interesting to note that subject independent analysis RMSE was reduced consistently between models using anthropometrics only and context-specific HR (for any activity), both in laboratory settings and free-living. However, increasing the intensity of the specific context analyzed, e.g. from lying down to walking at 3.5 km/h to walking at 5.5 km/h did not consistently reduce RMSE. RMSE for models including HR while lying down and slow walking (i.e. walking at 3.5 km/h) were similar, highlighting that the physiological responses to exercise we are interested in monitoring, might require a certain level of intensity for the model to benefit beyond what can be already achieved using lying HR as predictor. These findings are valid both in laboratory settings using HR during simulated activities of daily living and in free-living using HR as detected by pattern recognition methods.

10.4.6 Comparison with prior work

Little work was reported in literature on protocol-free VO_2max estimation. Previous studies aiming at estimating VO_2max in free-living conditions were either limited to using physical activity-related parameters, such as steps, as proposed by Cao et al. [40], HR normalized by activity intensity, as proposed by Plasqui et al. [99], or requiring intense exercise such as running [132]. Results for VO_2max estimation reported in terms of R^2 or RMSE cannot be easily compared between studies, due to the dependency of these parameters on the study's participants characteristics, for example body weight and VO_2max levels. However, we report in this section R^2 results as typically reported by other studies to put ours in perspective with current state of the art in VO_2max estimation. For some studies, e.g. [99], participants had similar characteristics to our study, and therefore comparisons can be meaningful. We reported R^2 of 0.79 for our subject independent analysis. Results reported by Plasqui et al. on a cross-validation sample for his method showed that using as predictor HR divided by activity counts, a measure of motion intensity, VO_2max could be predicted with $R^2 = 0.72$. The populations in the two studies are comparable, and therefore further contextualizing HR in free-living (i.e. using as predictor HR while walking at a certain speed) seems beneficial. Other protocols involving more intense activities, such as running, did not provide better results. For example, by combining the ratio of inverse foot-ground contact time and HR during steady state running, Weyand et al. [132] reported R^2 = 0.74 in the experimental group and R^2 = 0.67 in the cross-validation group.

By using context-specific HR in free-living as predictor, we obtained results comparable to or better than previous free-living studies and are also comparable to what was reported using similar metrics in laboratory settings or while performing strict protocols [111]. For example, ninety-two different VO_2max protocols were reviewed in a recent analysis by Sartor et al. [111]. Additionally to the free-living studies here discussed, the authors suggested that many other submaximal tests could be performed in free-living, without laboratory infrastructure. However, most of these tests require intense activities and strict protocols, for example the most commonly used 2-mile run (Mello et al. [88], R^2 = 0.81), Canadian aerobic fitness test (Jette et al. [75], $R^2 = 0.82$), or YMCA (Santo et al. [110], $R^2 = 0.56$). The accuracy of the best performing tests is comparable to our free-living estimation. However, the approach proposed in this work does not require intense activities, and is therefore suitable on a wider population. Additionally, the proposed approach does not require a specific test, and therefore VO_2max could be continuously assessed longitudinally over time, and not only re-assessed when the test is performed. The effectiveness of context-specific HR as derived in free-living with respect to laboratory based protocols was also validated in our own analysis, showing comparable RMSE and R^2 when including laboratory derived HR or free-living HR.

Other studies investigate the relation between easily accessible measures such as HR or HR variability at rest and VO_2max [55]. However, these studies typically reported low levels of accuracy (Esco et al. [55], $R^2 = 0.29$), showing that single measurements or spot measurements of physiological parameters and limited levels of context are insufficient for a reliable VO_2max estimate. A possible explanation for the better performance of the proposed approach compared to both single spot checks and more intense protocols that can be carried out in free-living, is that by contextualizing HR over multiple days, our proposed approach is less prone to the day-to-day variability typical of physiological measurements.

The clear advantage of the current approach is the ability to provide estimates during normal activities of daily living, as carried out by individuals. We validated our models independently on the participant, using cross-validation and the leave-one-out technique. Additionally, for all our models, we used as predictor body weight instead of fat-free mass to provide estimates from easily accessible measures that can be acquired without complex and expensive laboratory infrastructure. Thus, our results are extendable to new participants without the need of re-training the models or other laboratory protocols. The current implementation could be directly deployed to new studies in free-living conditions.

10.4.7 Limitations and future work

A limitation of this study is the validation on healthy adults only, with similar lifestyles in a Dutch setting. Future work should investigate if the proposed CRF estimation model is suitable for other groups such as the obese and persons affected by chronic disease, and if the proposed activity recognition system or other activity recognition systems trained to recognize only the relevant activities to contextualize HR (e.g. lying and walking) can be suitable for these populations. In non-healthy populations changes in CRF could provide an additional marker of disease progression. Additionally, future work should address the ability of the proposed method not only to estimate CRF for an individual, but to track changes in CRF over time, e.g. by means of a physical activity intervention. In this study, we assumed VO_2max to remain constant over a period of two weeks, since participants were not implementing changes to their lifestyle, and typical interventions to modify VO_2max are of much longer duration (e.g. 3 months to 1 year).

CRF is a strong and independent predictor of all-cause and cardiovascular mortality. When evaluating the suitability and practical applicability of a new test, many parameters should be accounted for. The cost, convenience and infrastructure required are current barriers to widespread VO_2max measurements, despite the well-known relevance in healthcare. The proposed CRF estimation model is applicable to a wide population, since it does not require intense physical exercise, and requires accelerometer and HR data only. Such measures, are becoming more and more widespread due to mainstream availability of wearable technology, including combined accelerometer and HR monitors. Similarly, the processing capabilities of modern mobile phones are sufficient for practical deployment of machine learning methods.

10.4.8 Conclusions

In conclusion, this work showed that contextualized HR in free-living can be used to provide VO_2max estimation with accuracy comparable to other methods relying on submaximal HR measured in laboratory settings. To the best of our knowledge, this is the first study using context-specific HR determined automatically using machine learning techniques in free-living to estimate VO_2max. The proposed approach could be used to provide more information about an individual's health without the need for laboratory infrastructure or specific tests. Building up on the proposed approach, new opportunities for applications targeted at inducing behavioral change could be developed. For example, by creating a feedback loop between objectively measured physical activity, and changes in CRF and associated reduced risk of disease.

Acknowledgments

The authors would like to thank Giuseppina Schiavone and Stefan Camps for their support during data collection.

11
Conclusions and future directions

Physical activity is key in maintaining a healthy lifestyle in the context of energy balance, obesity prevention and management as well as in a broader sense in the context of many other diseases resulting by lack of physical activity, such as cardiovascular disease. Being able to accurately quantify physical activity is important for epidemiological research so that relations between physical activity, health status, environmental factors, and so on, can be determined. Similarly, accurate quantification of physical activity can be key in deploying just in time interventions and promote behavioral change by providing individuals with an objective assessment of their physical activity behavior.

When analyzing the importance of physical activity in health, another important aspect to consider is how performed physical activity reflects into changes in physical fitness, and health status. Changes in physical fitness and heath status are typically measured in terms of CRF, a useful diagnostic and prognostic health indicator for patients in clinical settings, as well as healthy individuals.

The ability to properly measure and quantify both physical activity in terms of EE and physical fitness in terms of CRF could be key for epidemiologists to understand the relation between EE, CRF and health status. Additionally, new applications could be developed to promote behavioral change and personalized coaching. For example, by providing tailored feedback between physical activity behavior (or EE) and health markers (i.e. estimated CRF level), individuals could be helped in maintaining a healthy lifestyle.

To this aim, technological solutions able to unobtrusively measure both physical activity and fitness in free-living conditions are necessary. Thus, in this thesis we introduced new methods and models to provide accurate EE estimation at the individual level without requiring individual calibration and to estimate VO_2max, in both laboratory and unsupervised free-living conditions, using wearable sensors data.

More specifically, we investigated four goals related to 1) selection of methods, sensor number and positioning for EE estimation, 2) physiological data normalization to reduce EE estimation error without requiring individual calibration, 3) VO_2max estimation using wearable sensor data, without the need for laboratory protocols, and 4) personalized EE estimation and VO_2max estimation in free-living conditions.

11.1 Selection of methods, sensor number and positioning

Current solutions for EE estimation are affected by many limitations and up to now have been developed without systematically analyzing a series of aspects. For example, the number of sensors used and positioning on the body impact EE estimation accuracy at different levels. The accuracy of the activity recognition system employed as a first stage in activity-specific EE models and the way activities are misclassified has to be analyzed in the context of the resulting EE estimation error. Additionally, proposed methods using either static MET tables, accelerometer, physiological data (e.g. HR) or both have not been systematically analyzed to highlight what features are ideal depending on the activity performed. At the beginning of our work, a multitude of approaches had already been proposed, however, no clear methodology had been established.

In this thesis we proposed a new methodology to develop activity-specific EE estimation algorithms. Our methodology relied on classifying clusters of activities and then estimating EE using either static MET values for sedentary activities, or a combination of accelerometer and HR features for moderate to vigorous activities. The proposed method outperformed other techniques previously reported in literature and was then employed in a multi-sensor system to analyze the impact of sensor number and positioning on EE estimation. Results showed that one single sensor is sufficient for accurate EE estimation, provided that the sensor is close to the body's center of mass. While activity recognition accuracy drops when only one sensor is used, the misclassification of activities for sensors close to the body's center of mass is typically between activities with similar EE level, therefore minimizing EE estimation error. Additionally, by using static MET values instead of accelerometer or HR features for sedentary activity clusters, we prevent the estimate from being affected by physiological changes not due to physical activity, for example changes in HR at rest due to stress.

Summary of the findings:

- One single sensor is sufficient for accurate EE estimation, showing no loss in performance compared to multi-sensor systems including up to 5 sensors. However, the sensor needs to be placed close to the body's center of mass and activity-specific models need to be used.

- Static MET values for sedentary activities and a combination of accelerometer and physiological features for moderate to intense activities provide the most accurate EE estimates. RMSE for EE estimation was reduced by 88% compared to simple linear regression models and by 23% compared to activity-specific methods using METs lookup for active clusters.

11.2 Physiological data normalization

In parallel to the problem of defining the optimal methodology in terms of sensor number and positioning as well as feature selection in activity-specific EE models, we faced as major limitation the need for individual calibration. While models using physiological signals such as HR have consistently shown higher accuracy in EE estimation, the need for individual calibration limited practical applicability. Individual calibration is needed since the relation between e.g. HR and EE is specific for one individual, and cannot be generalized to a wider population. Especially with the commercialization of many physical activity monitors, solutions aiming at automatically normalizing HR are necessary to fully exploit the potential of using physiological data for EE estimation.

Thus, once we defined an optimal methodology for activity-specific modeling of EE estimates, the main body of this thesis focused on the major issue affecting EE estimation based on physiological data, i.e. the need for individual calibration. To automatically normalize physiological data, we proposed two methods. The first method relied on contextualizing HR during low intensity activities, i.e. determining HR while walking at a specific speed. We contextualized HR by combining an activity recognition classifier and a regression model for walking speed estimation. Then, contextualized HR was used to predict a normalization parameter, thus avoiding the need for individual calibration. A representation of this procedure is shown in Fig. 11.1. The proposed method using contextualized HR data as predictor for normalization parameters was then extended to other physiological data (galvanic skin response and respiration rate), as a generic methodology. We showed reduced EE estimation (RMSE) for models relying on HR, galvanic skin response and respiration rate between 15% and 33% compared non-normalized models. Then, as a second method we proposed a modeling technique relying on a hierarchical approach using Bayesian modeling. We first predicted CRF, i.e. the main underlying cause of individual differences in HR during moderate to intense physical activities between individuals. Then, we included CRF as group level predictor in EE estimation models. By including CRF level at the second level of a hierarchical Bayesian model, we avoided the need for individual calibration or explicit HR normalization since CRF accounted for the different relation between HR and EE in different individuals. Adopting this method we showed reduced EE estimation error by 18.2% on average.

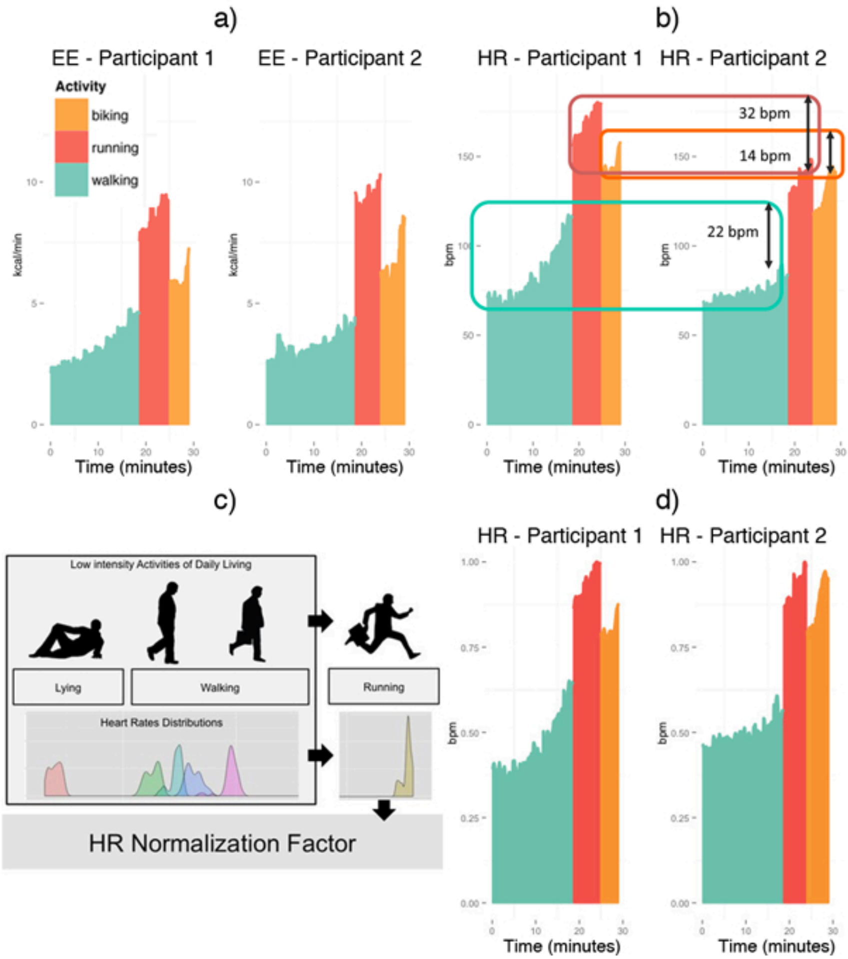

Figure 11.1: a) EE is mainly dependent on body size, therefore two participants with similar body weight and height consume approximately the same energy across activities. b) The HR of the same participants might be different during the same activities, with differences as big as $20-30\%$, depending on fitness level. As a result, estimating EE from HR would cause under and overestimations, since HR needs to be first normalized across participants. c) Block diagram of the automatic physiological data normalization method proposed in this thesis, for the case of HR. By determining HR in specific contexts (e.g. at rest or while walking at different speeds), the HR while running is predicted, without the need for the user to actually perform any intense activity. The predicted HR is used as HR normalization factor, to normalize HR at runtime d) Normalized HR, for the same participants shown in a and b, showing how individual differences in physiological responses due to e.g. CRF have now being removed, since the HR across activities are now comparable, similarly to what is shown in plot a for EE.

Summary of the findings:

- Physiological data during low intensity activities of daily living can be used to estimate the physiological data value during intense exercise, without the need for performing such intense exercise. In particular, in this thesis normalization parameters for HR, GSR and respiration rate were estimated by regression models explaining 90%, 88% and 72% of the variance respectively. Therefore, determining normalization parameters without the need for laboratory protocols and intense activities is possible.

- Estimated physiological data normalization parameters for different signals (HR, GSR, respiration) can be used to normalize such physiological signals and therefore estimate EE more accurately at the individual level, without the need for individual calibration. RMSE for EE estimation was reduced between 15 and 33% when using normalized physiological signals.

- EE estimation error due to activity misclassification in activity-specific EE estimation models is highly reduced when physiological data or normalized physiological data are used for activity recognition as well. Misclassification effect (i.e. increased RMSE due to the application of the wrong EE model due to an error of the activity recognition system) when no physiological signals were used was 20% for the ECG Necklace and 125% for the Wristband (due to the high confusion between active and inactive clusters). Including normalized physiological signals reduced the misclassification effect to 4% for the ECG Necklace and 19% for the Wristband.

- The different relation between HR and EE for individuals of different CRF level can be automatically accounted for by using a hierarchical approach to EE modeling, without the need for explicit HR normalization. This method uses CRF at the second level of a hierarchical model, and shows that EE estimation RMSE can be reduced by 18% when the source of between-individual variability (i.e. CRF) is known or estimated.

11.3 VO_2max estimation using wearable sensor data

Another major contribution of this thesis was the shift from focusing on quantifying physical activity behavior, e.g. what individuals do (activity types and intensities performed), to quantifying markers of health status influenced by physical activity behavior, such as levels of CRF. Unsupervised CRF estimation in free-living conditions was an area barely touched by previous research, mainly focusing on laboratory based protocols and strict exercises. In particular, we developed the first method to estimate VO_2max from contextualized HR data collected during activities of daily living, simulated in laboratory settings. The main assumption behind this work was that differences in physiology for individuals of different fitness levels, as typically captured during high intensity exercise, such as subaxmial fitness tests, would be present already during low intensity activities of daily living. This hypothesis derives by our personalization methods for EE estimation, where we were able to account for individual differences in fitness by estimating normalization parameters using low intensity activities of daily living. Thus, a hierarchical Bayesian regression approach was used, with model coefficients that varied depending on the performed activity. Results showed that VO_2max estimation RMSE could be reduced up to 27% compared to models including anthropometric characteristics but no contextualized HR as predictors.

Summary of the findings:

- CRF can be estimated from HR contextualized during low intensity activities of daily living such as walking at slow speeds. In particular, CRF estimation RMSE could be reduced up to 27% compared to models using anthropometrics only as predictors.

- CRF estimation is of sufficient accuracy to allow for personalized EE estimation without the need for HR normalization, when including estimated CRF as group level predictor in a hierarchical Bayesian model for EE estimation. Adopting this method we showed reduced EE estimation error by 18.2% on average.

11.4 Personalized EE estimation and VO_2max estimation in free-living

Finally, in the last part of the thesis we focused on bringing the methods proposed for personalized EE estimation and CRF estimation to free-living conditions. We relied on pattern recognition methods to automatically recognize specific low intensity activities of daily living that are used to contextualize physiological data, estimate HR normalization parameters and CRF. Before, such detected low intensity activities were typically performed in supervised laboratory settings. Most research up to now, for both EE estimation and CRF estimation focused on laboratory based protocols for model development and evaluation. Therefore new methods and models able to account for differences in individual behavior during unsupervised free-living activities are needed for practical applicability of personalized EE estimation and CRF estimation methods.

In free-living conditions, contextualizing and interpreting physiological data is challenging, due to the effect of both low-level activity primitives (e.g. lying down, walking, etc.) and high-level activity composites (e.g. commuting, working, socializing, etc.) on physiological data. Thus, a method was developed to combine low-level activity primitives and high-level activity composites using topic models. Using the proposed method, physiological data was analyzed not only in the context of low level activities, as it can be done under supervised laboratory conditions, but also depending on higher level activity composites. An example of the proposed context recognition framework is shown in Fig. 11.2.

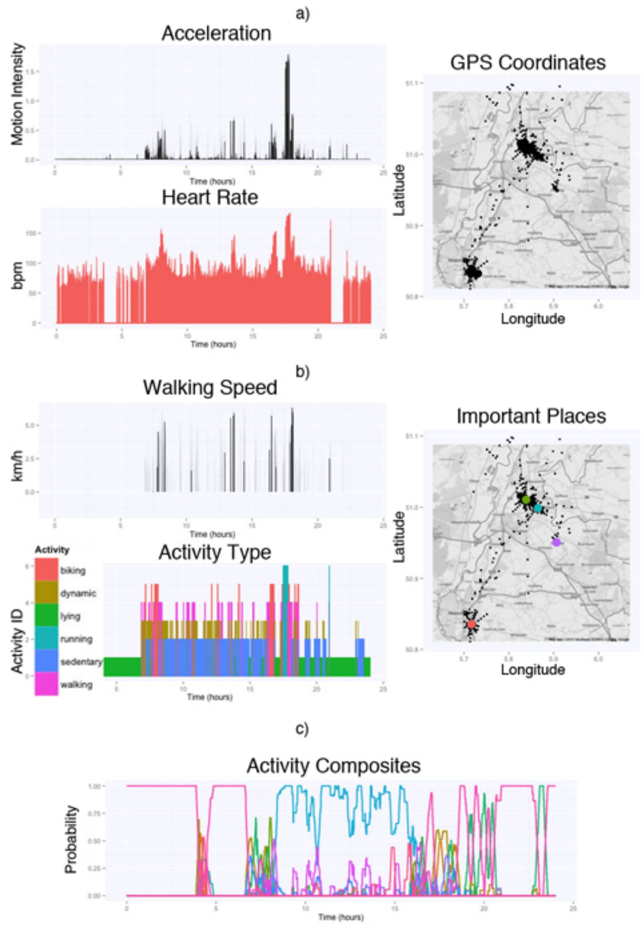

Figure 11.2: Context recognition framework proposed for the free-living works included in this thesis a) Example of raw data streams for accelerometer, HR and GPS data. b) Activity types (i.e. low level activity primitives such as walking or lying down) are derived from raw data using supervised methods. Walking speed is derived from raw data using regression models. Important places are derived by clustering GPS coordinates. c) High level activity composites, representative e.g. of activities such as sleeping, working or commuting, are derived unsupervisedly using topic models and activity primitives as building blocks. Walking speeds, low level activity types and activity composites were then used to contextualize HR in free-living conditions, for both EE estimation and CRF estimation applications.

Optimal contexts (i.e. combinations of activity primitives and composites) for analyzing HR for EE estimation and CRF estimation were determined without supervision. We showed that the proposed method can be used to estimate HR normalization parameters in free-living conditions, and therefore personalize laboratory derived EE models. Then, we also applied the proposed method to VO_2max estimation, showing reduced estimation error compared to existing methods. Finally, we analyzed context-specific HR as derived in both laboratory and free-living settings, showing that VO_2max estimation as obtained using only data acquired in free-living provides accuracy comparable or superior to laboratory based models. We validated the proposed methods on a dataset including 50 participants wearing a combined accelerometer and HR monitor for two weeks in unsupervised free-living conditions. EE RMSE was reduced by 10.7% while VO_2max estimation RMSE was reduced by up to 22.6% compared to alternative methods.

Summary of the findings:

- By ranking activity composites based on extracted features (e.g. distribution of low level activity primitives within the activity composite), activity composites can be analyzed across participants even if they are determined unsupervisedly. Therefore, the best activity composites for a given application could be selected, so that HR can be analyzed in the context of both low level activity primitives and high level activities composites.
- Contextualizing HR using both low level activity primitives and high level activity composites can provide more accurate EE estimates and CRF estimates with respect to no context or HR contextualized using low intensity activities only. In particular RMSE for EE estimation was reduced by 29.4% compared to models using non-normalized HR and by 19.8% compared to models using HR normalization parameters estimated using low level activity primitives only, and no activity composites. Additionally, VO_2max RMSE was reduced by up to 22.6% compared to alternative methods.
- VO_2max estimation error obtained using as predictor context-specific HR acquired in free-living conditions is smaller or comparable to VO_2max estimation error obtained using as predictor context-specific HR acquired in laboratory conditions.

11.5 Limitations

One of the limitations of this thesis' work, is that EE estimates were validated always using indirect calorimetry. Even when developing free-living models, using contextualized HR, the effectiveness of the estimated HR normalization parameters in reducing EE estimation error was validated in laboratory settings. While double labelled water (DLW) is the only recognized method to obtain reference EE in free-living, DLW reports only total EE after a period of one or two weeks. Thus, DLW is not informative in terms of minute-by-minute EE estimation accuracy. An EE estimation model that would consistently overestimate light activities and consistently underestimate intense activities could perform optimally according to DLW, due to an averaging of multiple errors. Thus, we validated our approach using laboratory data and reference indirect calorimetry, since only under these conditions we can acquire minute-by-minute EE reference for different activities, and evaluate the models' accuracy. Similarly, when developing our models we could evaluate activity recognition and walking speed accuracy only under laboratory conditions, where reference was present. Another limitation of this thesis is the validation on young healthy adults only, with similar lifestyles in a Dutch setting. While our goal was to provide accurate EE estimation and CRF estimation for healthy individuals, additional work is required to investigate if the proposed estimation models are suitable for other groups such as different age groups (e.g. the elderly), the obese and persons affected by chronic disease, and if the proposed activity recognition system can be suitable for these populations. Finally, future work is needed to determine the ability of the proposed method not only to estimate CRF for an individual, but to track changes in CRF over time, e.g. by means of a physical activity intervention, since in this thesis we assumed VO_2max to remain constant for the studies duration.

11.6 Future directions

In conclusion, the research included in this thesis showed that machine learning techniques can be used to normalize and contextualize physiological data in either laboratory or unsupervised free-living conditions. By determining multiple levels of contexts we showed that personalized EE estimation and VO_2max estimation using contextualized physiological data are both possible in unsupervised free-living settings. Therefore, higher accuracy could be obtained in EE estimation and VO_2max estimation, compared to previous efforts, with consistent error reductions ranging between 10% and 33%.

We see the following as main future directions that could be built upon the methods included in this thesis:

- Physiological deviations from a person's baseline: The proposed context recognition framework, including low and high level activity recognition, could be used to first establish a person's physiological signals baseline in different contexts, and then detect physiological signals deviations from a person's baseline. Basically, by determining the distribution of physiological data in different contexts (e.g. the distribution of HR data of an individual while doing a certain activity in a certain place), deviations from a persons's normal values could be detected. Example applications could be psychological stress detection, by determining when physiological data (e.g. higher HR or lower HRV) is consistently far from a person's baseline values. As a result, personalized alarms or coaching could be delivered, since a person's values would not be compared to population-based parameters but only to the person's history. Another application relying on the same framework could be early detection of health issues that often translate in physiological changes from a person's baseline (or normal) values, such as for example heart failure patients who's physiological signals change well before exacerbations.

- Hierarchical modeling for implicit physiological signal normalization: In this thesis we showed how using a hierarchical Bayesian approach where the factor influencing individual differences (i.e. CRF) between the outcome variable (i.e EE) and the predictor variable (i.e. HR) was placed at the second level of the hierarchical structure, therefore providing more accurate estimates. The reason being that the different relation between predictor (HR) and outcome variable (EE) due to CRF was accounted for by letting the HR coefficient depend on the second level predictor, CRF (i.e. the relation between HR and EE is controlled by level of CRF). The same approach could possibly be used to reduce error in other applications with consistent differences between individuals. For example, for physiological data, another application could be psychological stress detection, which also can rely on HR. However the relation between HR and psychological stress is different in different individuals, and using a hierarchical model with as group level predictors variables controlling the relation between HR and stress, could improve stress estimation/detection at the individual level.

- Context recognition: As context recognition used to contextualize physiological data we relied on supervised methods for low level activity detection and topic models for high level unsupervised activity discovery. The proposed method could be extended by exploring other ways to determine high level activities, unsupervisedly. For example, by using convolutional neural networks, which could be fed with low level raw data and would automatically build high level abstractions of the user's behavior, similarly to topic models. Another alternative could be to use methods requiring user's interaction, such as active learning, to further expand and personalize the detected context.

- Healthy living: From a clinical or consumer application perspective, the link between VO_2max and reduced risk of disease could be exploited to develop new applications built on the thesis results. For example, targeting behavioral change by closing the feedback loop between activity behavior (or EE) and estimated fitness (CRF). Providing users or patients with insights on how their activity behavior and lifestyle influences their health status, as estimated by CRF. To this aim, the VO_2max estimation should first be validated longitudinally, to determine if the models proposed in this work are not only able to provide accurate VO_2max estimation at the cross-sectional level, but also longitudinally within one individual with a varying physical activity behavior (e.g. taking up a more active lifestyle).

Acknowledgments

First and foremost, I would like to thank my PhD supervisor, prof. Oliver Amft, for his guidance during these years. I am very grateful for everything I have learned from him. Thank you, Oliver, for patiently coaching me during my doctorate work and for giving me the freedom to explore my own ideas. I would also like to thank prof. Jan Bergmans, for giving me the possibility to continue my work at TU/e when bureaucracy got in the way, letting me finalize my research towards a PhD degree. Additionally, I would like to thank all members of my doctoral Committee, whose comments were very valuable in further improving the thesis.

Most of the work presented in this thesis has been conducted at imec, the Netherlands. For this reason I would like to thank everyone at imec and Holst Centre for their support and help during these years. A special thanks goes to all the colleagues that participated in my studies and helped me collecting the datasets necessary for the development of my work. I am particularly grateful to Marij Heil and Bert Gyselinckx, who allowed me to continue with my research even when I eventually decided to leave imec.

Many thanks go to my first mentor at imec, Julien Penders, who has supported me during the first part of my doctorate research. Since the very beginning you gave me the opportunity to shape the project according to my research interests, I thank you for that and for the many inspiring meetings in the early mornings.

The latest experimental data of this thesis have been recorded at the Human Biology department at Maastricht University, and I would like to thank everyone supporting my work there, in particular Guy Plasqui and Gabrielle ten Velde.

I would also like to thank my former advisors back in Italy during my Master's, Luca Benini and Elisabetta Farella. Thank you for your guidance, inspiration and for putting me in contact with Julien in the early days of my research. None of this would have happened without your support.

Besides my colleagues at imec, TU/e and now Bloom Technologies, I would like to thank everyone else that supported me during these years, all my friends in Italy, my new friends in the Netherlands, my family and Alessandra, who's always been supportive, even when my choices brought us ten thousand kilometers apart.

Resume

Marco Altini is a PhD candidate at the Signal Processing Systems group at Eindhoven University of Technology, NL. His PhD research at TU/e focuses on applying machine learning techniques to develop methods and models to derive personalized assessment of physical activity and health markers, using data obtained with wearable sensors. During this period, he published more than 20 peer reviewed papers for international conferences and journals. Some of his papers were shortlisted for best paper awards and selected as research highlights on the journal's homepages. Additionally, he published three patents, which have been licensed to multiple customers in the high-end fitness industry.

Marco is also leading data science activities at Bloom Technologies, San Francisco, CA, a digital health startup focusing on helping expecting mothers have a healthy pregnancy. At Bloom, Marco is working on combining data collected in clinical settings as well as consumer generated data to help shedding light on many poorly understood links between physiological changes naturally occurring during pregnancy, behavior and pregnancy outcomes.

Before his PhD, he obtained his M.Sc. degree summa cum laude in Computer Science Engineering in 2010 from the University of Bologna. Between 2009 and 2014 he worked at imec as part of the Human++ program, developing hardware, firmware, software and algorithms for Body Area Networks applications.

Since 2012, he has also been developing mobile apps using the phone's sensors to provide unique insights on the user's physiological status (such as heart rate and heart rate variability) without the need for external hardware. Some of his work was top-charting, reaching the 5th place among the most downloaded paid apps in the Netherlands. He is now also working as freelance and curating a blog HRV4Training.com, covering technology, sport, physiology and data.

Publications by the Author

Journal Papers

[JP-1] M. Altini, J. Penders, R. Vullers, and O. Amft. Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning. IEEE Journal of Biomedical and Health Informatics, 19(1):219-226, 2015.
See Chapter 3.

[JP-2] M. Altini, J. Penders, R. Vullers, and O. Amft. Automatic heart rate normalization for accurate energy expenditure estimation. Methods Inf Med, 53(5):382-388, 2014.
See Chapter 5.

[JP-3] M.Altini, J.Penders, R.Vullers, and O.Amft. Personalizing energy expenditure estimation using physiological signals normalization during activities of daily living. Physiol Meas, 35(9):1797, September 2014.
See Chapter 6.

[JP-4] M. Altini, P. Casale, J. Penders, and O. Amft. Personalized cardiorespiratory fitness and energy expenditure estimation using hierarchical Bayesian models. Journal of Biomedical Informatics, 56:195-204, 2015.
See Chapter 7.

[JP-5] M. Altini, P. Casale, J. Penders, and O. Amft. Personalization of energy expenditure estimation in free living using topic models. IEEE Journal of Biomedical and Health Informatics, 19(5):1577-1586, 2015.
See Chapter 8.

[JP-6] M. Altini, P. Casale, J. Penders, and O. Amft. Cardiorespiratory fitness estimation in free living using wearable sensors. Submitted to Artificial Intelligence in Medicine.
See Chapter 9.

[JP-7] M. Altini, P. Casale, J. Penders, G. ten Velde, G. Plasqui, and O. Amft. Cardiorespiratory fitness estimation using wearable sensors data: analysis of

context-specific submaximal heart rates. Submitted to the Journal of Applied Physiology.
See Chapter 10.

[JP-8] M. Altini, J. Penders, and O. Amft, Estimating oxygen uptake during non-steady-state activities and transitions using wearable sensors. Accepted for publication in IEEE Journal of Biomedical and Health Informatics.

[JP-9] P. Casale, M. Altini, and O. Amft, Transfer learning in body sensor networks using ensembles of randomised trees. IEEE Internet of Things Journal, 2(1):33-40, 2015.

[JP-10] J. Penders, M. Altini, C. van Hoof, and E. Dy, Wearable sensors for healthier pregnancies. Proceedings of the IEEE. 103(2):179-191, 2015.

[JP-11] D. Buxi, S. Kim, N. van Helleputte, M. Altini, J. Wijsman, F. Yazicioglu, J. Penders, and C. van Hoof, Correlation between electrode-tissue impedance and motion artifact in biopotential recordings. IEEE Sensors Journal, 12(12):3373-3383, 2012.

Conference Papers

[CP-1] M. Altini, J. Penders, and O. Amft. Energy expenditure estimation using wearable sensors: A new methodology for activity-specific models. In Proceedings of the Conference on Wireless Health, WH '12, pages 1:1-1:8, New York, NY, USA, 2012. ACM.
See Chapter 2.

[CP-2] M. Altini, J. Penders, and O. Amft. Personalizing energy expenditure estimation using a cardiorespiratory fitness predicate. In Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on, pages 65-72. IEEE, 2013.
See Chapter 5.

[CP-3] M. Altini, J. Penders, R. Vullers, and O. Amft. Combining wearable accelerometer and physiological data for activity and energy expenditure estimation. page 1. ACM, 2013.

[CP-4] M. Altini, J. Penders, and O. Amft, Body weight-normalized energy expenditure estimation using combined activity and allometric scaling clustering, in Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pages 6752-6755, 2013.

[CP-5] M. Altini, J. Penders, R. Vullers, and O. Amft. Personalized physical activity monitoring on the move. In Proceedings of the 4th Conference on Wireless Health, page 8. ACM, 2013.

[CP-6] M. Altini, M. van Dort, R. Vullers, C. van Hoof, and O. Amft. Self-calibration of walking speed estimations using smartphone sensors. In IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), pages 10-18, 2014.

[CP-7] M. Altini, S. Polito, J. Penders, H. Kim, N. Van Helleputte, S. Kim and F. Yazicioglu. An ECG patch combining a customized ultra-low-power ECG SoC with bluetooth low energy for long term ambulatory monitoring. In Proceedings of the 2nd Conference on Wireless Health, page 15. ACM, 2011.

[CP-8] M. Altini, S. Del Din, S. Patel, S. Schachter, J. Penders, and P. Bonato. A low-power multi-modal body sensor network with application to epileptic seizure monitoring. In Engineering in Medicine and Biology Society, EMBC, Annual International Conference of the IEEE, pages 1806-1809, 2011.

[CP-9] M. Altini, E. Farella, M. Pirini, and L. Benini, A Cost-effective Indoor Vibrotactile Navigation System for the Blind, in Proc. HEALTHINF, pages 477-481, 2011.

[CP-10] M. Altini, J. Penders, and H.W. Roebbers: An Android-based body area network gateway for mobile health applications. Wireless Health pages 188-189, 2010.

[CP-11] M. Altini, D. Brunelli, E. Farella and L. Benini, Bluetooth indoor localization with multiple neural networks, In Wireless Pervasive Computing (ISWPC), 5th IEEE International Symposium on, pages 295-300, 2011.

[CP-12] S. Chen, M. Altini, J. Penders, O. Amft, and J. Lach. Unsupervised activity clustering to estimate energy expenditure with a single body sensor. In Body Sensor Networks (BSN), IEEE International Conference on, pages 1-6, 2013.

[CP-13] J. Penders, M. Altini, J. van de Molengraft, F. Yazicioglu, and C. Van Hoof. A low-power wireless ECG necklace for reliable cardiac activity monitoring on-the-move. In Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society. Invited paper. 2011.

[CP-14] P. Casale, M. Altini, and O. Amft. Transfer learning in body sensor networks using ensemble of randomised trees. In BSN, pages 39-44, 2014.

[CP-15] J. Penders, M. Altini , J. Wijsman, R. Vullers, and C.V. Hoof. Low-power wearable sensing for preventive health(care). In EMBC 2013.

[CP-16] S. Shrestha, T. Torfs, H. Kim, I. Romero, D. Buxi, T. Berset, M. Altini, and F. Yazicioglu. Optimized R peak detection algorithm for ultra low power ECG system. In Biomedical Circuits and Systems Conference (BioCAS), IEEE, pages 225-228. 2011.

[CP-17] R. Vullers, J. Penders, S. Brongersma, J. Wijsman, M. Altini, and C.V. Hoof. Wearable smart systems for personal healthcare and lifestyle, SSI 2013.

Patents

[P-1] M. Altini. Device and method for calculating cardiorespiratory fitness level and energy expenditure of a human being. Application submitted. 2012.

[P-2] M. Altini. Device and method for automatically normalizing the physiological signals of a living being. Application submitted. 2013.

[P-3] M. Altini, J. Penders, T. Berset, E. Dy. System and method for detecting and quantifying deviations from the physiological signals normality of a subject. Application submitted. 2014.

[P-4] J. Penders, M. Altini, T. Berset, E. Dy. A method and system for biometric and context based messaging. Application submitted. 2014.

Bibliography

[1] M. Abitbol. Effect of posture and locomotion on energy expenditure. *American Journal of Physical Anthropology*, 77(2):191–199, 1988.

[2] B. E. Ainsworth, W. L. Haskell, M. C. Whitt, M. L. Irwin, A. M. Swartz, S. J. Strath, W. L. O'Brien, D. R. Bassett, K. H. Schmitz, P. O. Emplaincourt, and et al. Compendium of physical activities: an update of activity codes and met intensities. *Medicine & Science in Sports & Exercise*, 32(9 Suppl):S498–504, 2000.

[3] F. Albinali, S. S. Intille, W. Haskell, and M. Rosenberger. Using wearable activity type detection to improve physical activity energy expenditure estimation. In *UbiComp*, pages 311–320, 2010.

[4] M. Altini, P. Casale, J. Penders, and O. Amft. Cardiorespiratory fitness estimation in free living using wearable sensors. *submitted to Artificial Intelligence in Medicine*, 2015.

[5] M. Altini, P. Casale, J. Penders, and O. Amft. Personalization of energy expenditure estimation in free living using topic models. *IEEE Journal of Biomedical and Health Informatics*, 19(5):1577–1586, 2015.

[6] M. Altini, P. Casale, J. Penders, and O. Amft. Personalized cardiorespiratory fitness and energy expenditure estimation using hierarchical bayesian models. *Journal of Biomedical Informatics*, 56:195–204, 2015.

[7] M. Altini, S. Del Din, S. Patel, S. Schachter, J. Penders, and P. Bonato. A low-power multi-modal body sensor network with application to epileptic seizure monitoring. In *EMBC, 2011 Annual International Conference of the IEEE*, pages 1806–1809. IEEE, 2011.

[8] M. Altini, S. Del Din, S. Patel, S. Schachter, J. Penders, and P. Bonato. A low-power multi-modal body sensor network with application to epileptic seizure monitoring. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 1806–1809. IEEE, 2011.

[9] M. Altini, J. Penders, and O. Amft. Energy expenditure estimation using wearable sensors: A new methodology for activity-specific models. In *Proceedings of the Conference on Wireless Health*, WH '12, pages 1:1–1:8, New York, NY, USA, 2012. ACM.

[10] M. Altini, J. Penders, and O. Amft. Personalizing energy expenditure estimation using a cardiorespiratory fitness predicate. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on*, pages 65–72. IEEE, 2013.

[11] M. Altini, J. Penders, R. Vullers, and O. Amft. Combining wearable accelerometer and physiological data for activity and energy expenditure estimation. page 1. ACM, 2013.

[12] M. Altini, J. Penders, R. Vullers, and O. Amft. Personalized physical activity monitoring on the move. In *Proceedings of the 4th Conference on Wireless Health*, page 8. ACM, 2013.

[13] M. Altini, J. Penders, R. Vullers, and O. Amft. Personalizing energy expenditure estimation using physiological signals normalization during activities of daily living. *Physiol Meas*, 35(9):1797, September 2014.

[14] M. Altini, J. Penders, R. Vullers, and O. Amft. Estimating energy expenditure using body-worn accelerometers: a comparison of methods, sensors number and positioning. *IEEE Journal of Biomedical and Health Informatics*, 19(1):219–226, 2015.

[15] M. Altini, J. Penders, R. Vullers, O. Amft, et al. Automatic heart rate normalization for accurate energy expenditure estimation. *Methods Inf Med*, 53(5):382–388, 2014.

[16] O. Amft, C. Lombriser, T. Stiefmeier, and G. Tröster. Recognition of user activity sequences using distributed event detection. In *Smart Sensing and Context*, pages 126–141. Springer, 2007.

[17] F. K. Assah, U. Ekelund, S. Brage, A. Wright, J. C. Mbanya, and N. J. Wareham. Accuracy and validity of a combined heart rate and motion sensor for the measurement of free-living physical activity energy expenditure in adults in cameroon. *International journal of epidemiology*, page dyq098, 2010.

[18] P. O. Åstrand and I. Ryhming. A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. *Journal of Applied Physiology*, 7(2):218–221, 1954.

[19] L. Atallah, B. Lo, R. King, and G.-Z. Yang. Sensor positioning for activity recognition using wearable accelerometers. *IEEE Transactions on Biomedical Circuits and Systems*, 5(4):320–329, 2011.

[20] L. Bao and S. Intille. Activity recognition from user-annotated acceleration data. In *Pervasive '04*, volume 3001, pages 1–17. 2004.

[21] D. R. Bassett, A. Rowlands, and S. G. Trost. Calibration and validation of wearable monitors. *Med Sci Sports Exerc*, 44(1 Suppl 1):S32–S38, Jan. 2012.

[22] R. Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on Neural Networks*, 5(4):537–550, 1994.

[23] G. G. Berntson. Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology*, 34:623–648, 1997.

[24] S. N. Blair and S. Brodney. Effects of physical inactivity and obesity on morbidity and mortality: current evidence and research issues. *Medicine & Science in Sports & Exercise*, (31):S646–62, 1999.

[25] S. N. Blair, J. B. Kampert, H. W. Kohl, C. E. Barlow, C. A. Macera, R. S. Paffenbarger, and L. W. Gibbons. Influences of cardiorespiratory fitness and other precursors on cardiovascular disease and all-cause mortality in men and women. *Jama*, 276(3):205–210, 1996.

[26] S. N. Blair, H. W. Kohl, R. S. Paffenbarger, D. G. Clark, K. H. Cooper, and L. W. Gibbons. Physical fitness and all-cause mortality: A prospective study of healthy men and women. *Jama*, 262(17):2395–2401, 1989.

[27] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.

[28] D. S. Bocalini, L. d. Santos, and A. J. Serra. Physical exercise improves the functional capacity and quality of life in patients with heart failure. *Clinics*, 63(4):437–442, 2008.

[29] A. G. Bonomi. Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. *Journal of Applied Physiology*, 107(3):655–661, 2009.

[30] A. G. Bonomi and G. Plasqui. "Divide and conquer": assessing energy expenditure following physical activity type classification. *Journal of Applied Physiology*, 112(5):932, Mar. 2012.

[31] L. Bouarfa, L. Atallah, R. M. Kwasnicki, C. Pettitt, G. Frost, and G.-Z. Yang. Predicting free-living energy expenditure using a miniaturized ear-worn sensor: An evaluation against doubly labeled water. *Biomedical Engineering, IEEE Transactions on*, 61(2):566–575, 2014.

[32] S. H. Boutcher and P. Stein. Association between heart rate variability and training response in sedentary middle-aged men. *European journal of applied physiology and occupational physiology*, 70(1):75–80, 1995.

[33] C. V. Bouten, K. T. Koekkoek, M. Verduin, R. Kodde, and J. D. Janssen. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity. *IEEE Transactions on Biomedical Engineering*, 44(3):136–147, 1997.

[34] S. Brage. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. *Journal of Applied Physiology*, 96(1):343–351, Aug. 2003.

[35] S. Brage. Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *J Appl Physiol*, 2007.

[36] R. C. Browning and R. Kram. Energetic cost and preferred speed of walking in obese vs. normal weight women. *Obesity Research*, 13(5):891–899, 2005.

[37] M. Buchheit, C. Simon, A. Charloux, S. Doutreleau, F. Piquard, and G. Brandenberger. Heart rate variability and intensity of habitual physical activity in middle-aged persons. *Medicine and science in sports and exercise*, 37(9):1530, 2005.

[38] J. Bussmann, W. Martens, J. Tulen, F. Schasfoort, H. Van Den Berg-Emons, and H. Stam. Measuring daily behavior using ambulatory accelerometry: the activity monitor. *Behavior Research Methods, Instruments, & Computers*, 33(3):349–356, 2001.

[39] F. Buttussi and L. Chittaro. Mopet: A context-aware and user-adaptive wearable system for fitness training. *Artificial Intelligence in Medicine*, 42(2):153–163, 2008.

[40] Z.-B. Cao, N. Miyatake, M. Higuchi, K. Ishikawa-Takata, M. Miyachi, and I. Tabata. Prediction of VO2max with daily step counts for japanese adult women. *European journal of applied physiology*, 105(2):289–296, 2009.

[41] M. R. Carnethon, M. Gulati, and P. Greenland. Prevalence and cardiovascular disease correlates of low cardiorespiratory fitness in adolescents and adults. *Jama*, 294(23):2981–2988, 2005.

[42] C. J. Caspersen, K. E. Powell, and G. M. Christenson. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public health reports*, 100(2):126, 1985.

[43] S. M. Ceesay, A. M. Prentice, and K. C. Day. The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. *The British Journal of Nutrition*, 61(2):175–186, Mar. 1989.

[44] M. Chan, D. Estève, J.-Y. Fourniols, C. Escriba, and E. Campo. Smart wearable systems: Current status and future challenges. *Artificial intelligence in medicine*, 56(3):137–156, 2012.

[45] S. Chen, J. Lach, O. Amft, M. Altini, and J. Penders. Unsupervised activity clustering to estimate energy expenditure with a single body sensor. In *Body Sensor Networks (BSN), 2013 IEEE International Conference on*, pages 1–6. IEEE, 2013.

[46] T. S. Church, C. P. Earnest, J. S. Skinner, and S. N. Blair. Effects of different doses of physical activity on cardiorespiratory fitness among sedentary, overweight or obese postmenopausal women with elevated blood pressure: a randomized controlled trial. *Jama*, 297(19):2081–2091, 2007.

[47] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes, S. McClean, and D. Finlay. Optimal placement of accelerometers for the detection of everyday activities. *Sensors*, 13(7):9183–9200, 2013.

[48] K. H. Cooper. A means of assessing maximal oxygen intake: correlation between field and treadmill testing. *Jama*, 203(3):201–204, 1968.

[49] L. Cordain, A. Gotshall, and S. Eaton. Evolutionary aspects of exercise. *World Rev Nutr Diet*, 81:49–60, 1997.

[50] L. Cordain, R. Gotshall, S. B. Eaton, and S. Eaton 3rd. Physical activity, energy expenditure and fitness: an evolutionary perspective. *International journal of sports medicine*, 19(5):328–335, 1998.

[51] S. Crouter, J. Churilla, and D. Bassett. Estimating energy expenditure using accelerometers. *European Journal of Applied Physiology*, 98(6):601–612, Dec. 2006.

[52] S. E. Crouter and D. R. Bassett. A refined 2-regression model for the actigraph accelerometer. *Medicine & Science in Sports & Exercise*, 42(5):1029–1037, 2010.

[53] S. Crumpton, H. N. Williford, S. O'Mailia, M. S. Olson, and L. E. Woolen. Validity of the polar m52 heart rate monitor in predicting VO2max. *Medicine & Science in Sports & Exercise*, 35(5):S193, 2003.

[54] C. B. Ebbeling, A. Ward, E. M. Puleo, J. Widrick, and J. M. Rippe. Development of a single-stage submaximal treadmill walking test. *Medicine and science in sports and exercise*, 23(8):966–973, 1991.

[55] M. R. Esco, E. M. Mugu, H. N. Williford, A. N. McHugh, and B. E. Bloomquist. Cross-validation of the polar fitness test tm via the polar f11 heart rate monitor in predicting VO2 max. *Age (yrs)*, 24:5–1, 2011.

[56] M. R. Esco, M. S. Olson, H. N. Williford, D. L. Blessing, D. Shannon, and P. Grandjean. The relationship between resting heart rate variability and heart rate recovery. *Clinical Autonomic Research*, 20(1):33–38, 2010.

[57] J. Fang, J. Wylie-Rosett, H. W. Cohen, R. C. Kaplan, and M. H. Alderman. Exercise, body mass index, caloric intake, and cardiovascular mortality. *American journal of preventive medicine*, 25(4):283–289, 2003.

[58] A. Ferro-Luzzi, N. Norgan, and G. Pastore. Energy balance methodology. In *Nutritional Status Assessment*, pages 113–130. Springer, 1991.

[59] D. Frankenfield, L. Roth-Yousey, and C. Compher. Comparison of Predictive Equations for Resting Metabolic Rate in Healthy Nonobese and Obese Adults: A Systematic Review. 105(5):775–789, May 2005.

[60] P. Freedson. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: Validation on an independent sample. *Journal of Applied Physiology*, 2011.

[61] A. Gelman and J. Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, 2006.

[62] J. D. George, P. R. Vehrs, P. E. Allsen, G. W. Fellingham, and A. G. Fisher. Development of a submaximal treadmill jogging test for fit college-aged individuals. *Medicine and science in sports and exercise*, 25(5):643–647, 1993.

[63] M. Gietzelt, K. Wolf, M. Kohlmann, M. Marschollek, R. Haux, et al. Measurement of accelerometry-based gait parameters in people with and without dementia in the field. *Methods Inf Med*, 52(4):319–325, 2013.

[64] F. Gómez-Gallego, J. R. Ruiz, A. Buxens, S. Altmäe, M. Artieda, C. Santiago, M. González-Freire, Z. Verde, D. Arteta, A. Martínez, et al. Are elite endurance athletes genetically predisposed to lower disease risk? *Physiological genomics*, 41(1):82–90, 2010.

[65] P. Gordon-Larsen, M. C. Nelson, and B. M. Popkin. Longitudinal physical activity and sedentary behavior trends: adolescence to adulthood. *American journal of preventive medicine*, 27(4):277–283, 2004.

[66] C. C. Grant, J. R. Clark, D. C. J. van Rensburg, and M. Viljoen. Relationship between exercise capacity and heart rate variability: Supine and in response to an orthostatic stressor. *Autonomic Neuroscience*, 151(2):186–188, 2009.

[67] R. Guidoux, M. Duclos, G. Fleury, P. Lacomme, N. Lamaudière, P.-H. Manenq, L. Paris, L. Ren, and S. Rousset. A smartphone-driven methodology for estimating physical activities and energy expenditure in free living conditions. *Journal of biomedical informatics*, 2014.

[68] W. L. Haskell, I.-M. Lee, R. R. Pate, K. E. Powell, S. N. Blair, B. A. Franklin, C. A. Macera, G. W. Heath, P. D. Thompson, and A. Bauman. Physical activity and public health: updated recommendation for adults from the american college of sports medicine and the american heart association. *Circulation*, 116(9):1081, 2007.

[69] D. P. Heil, P. S. Freedson, L. E. Ahlquist, J. Price, and J. M. Rippe. Nonexercise regression models to estimate peak oxygen consumption. *Medicine and science in sports and exercise*, 27(4):599–606, 1995.

[70] F. B. Hu, T. Y. Li, G. A. Colditz, W. C. Willett, and J. E. Manson. Television watching and other sedentary behaviors in relation to risk of obesity and type 2 diabetes mellitus in women. *Jama*, 289(14):1785–1791, 2003.

[71] G. Hu, Q. Qiao, K. Silventoinen, J. G. Eriksson, P. Jousilahti, J. Lindström, T. T. Valle, A. Nissinen, and J. Tuomilehto. Occupational, commuting, and leisure-time physical activity in relation to risk for type 2 diabetes in middle-aged finnish men and women. *Diabetologia*, 46(3):322–329, 2003.

[72] T. Huynh, M. Fritz, and B. Schiele. Discovery of activity patterns using topic models. In *Proceedings of the 10th international conference on Ubiquitous computing*, pages 10–19. ACM, 2008.

[73] A. S. Jackson, S. N. Blair, M. T. Mahar, L. T. Wier, R. M. Ross, and J. E. Stuteville. Prediction of functional aerobic capacity without exercise testing. *Medicine and Science in Sports and exercise*, 22(6):863–870, 1990.

[74] E. Jequier. Long-term measurement of energy expenditure in man: direct or indirect calorimetry. *Recent advances in obesity research*, 3:130–135, 1981.

[75] M. Jetté, J. Campbell, J. Mongeon, and R. Routhier. The canadian home fitness test as a predictor of aerobic capacity. *Canadian Medical Association Journal*, 114(8):680, 1976.

[76] P. T. Katzmarzyk. Physical activity, sedentary behavior, and health: paradigm paralysis or paradigm shift? *Diabetes*, 59(11):2717–2725, 2010.

[77] A. M. Kriska, A. Saremi, R. L. Hanson, P. H. Bennett, S. Kobes, D. E. Williams, and W. C. Knowler. Physical activity, obesity, and the incidence of type 2 diabetes in a high-risk population. *American Journal of Epidemiology*, 158(7):669–675, 2003.

[78] H. Kuipers, F. Verstappen, H. Keizer, P. Geurten, and G. Van Kranenburg. Variability of aerobic performance in the laboratory and its physiologic correlates. *International journal of sports medicine*, 6(04):197–201, 1985.

[79] S. Kumanyika, E. Obarzanek, N. Stettler, R. Bell, A. Field, S. Fortmann, and Y. Hong. The need for comprehensive promotion of healthful eating, physical activity, and energy balance. a scientific statement from american heart association council on epidemiology, prevention, interdisciplinary committee for prevention. *Circulation, Journal of the American Heart Association*, 2008.

[80] C. M. Lee and A. Mendoza. Dissociation of heart rate variability and heart rate recovery in well-trained athletes. *European journal of applied physiology*, 112(7):2757–2766, 2012.

[81] D.-c. Lee, E. G. Artero, X. Sui, and S. N. Blair. Review: Mortality trends in the general population: the importance of cardiorespiratory fitness. *Journal of Psychopharmacology*, 24(4 suppl):27–35, 2010.

[82] J. A. Levine, M. W. Vander Weg, J. O. Hill, and R. C. Klesges. Non-exercise activity thermogenesis the crouching tiger hidden dragon of societal weight gain. *Arteriosclerosis, thrombosis, and vascular biology*, 26(4):729–736, 2006.

[83] A. Loimaala, H. Huikuri, P. Oja, M. Pasanen, and I. Vuori. Controlled 5-mo aerobic training improves heart rate but not heart rate variability or baroreflex sensitivity. *Journal of Applied Physiology*, 89(5):1825–1829, 2000.

[84] S. J. Marshall and E. Ramirez. Reducing Sedentary Behavior : A New Paradigm in Physical Activity Promotion. *American Journal of Lifestyle Medicine*, 2011.

[85] J. E. McLaughlin. Validation of the COSMED K4 b2 portable metabolic system. *International Journal of Sports Medicine*, 22(4):280–284, May 2001.

[86] E. P. Meijer, K. R. Westerterp, and F. T. Verstappen. Effect of exercise training on total daily physical activity in elderly humans. *European journal of applied physiology and occupational physiology*, 80(1):16–21, 1999.

[87] E. L. Melanson and P. S. Freedson. The effect of endurance training on resting heart rate variability in sedentary adult males. *European journal of applied physiology*, 85(5):442–449, 2001.

[88] R. P. Mello, M. M. Murphy, and J. A. Vogel. Relationship between a two mile run for time and maximal oxygen uptake. *The Journal of Strength & Conditioning Research*, 2(1):9–12, 1988.

[89] A. E. Minetti, L. Boldrini, L. Brusamolin, P. Zamparo, and T. McKee. A feedback-controlled treadmill (treadmill-on-demand) and the spontaneous speed of walking and running in humans. *Journal of Applied Physiology*, 95(2):838–843, 2003.

[90] J. N. Morris and M. D. Crawford. Coronary heart disease and physical activity of work. *British Medical Journal*, 2(5111):1485, 1958.

[91] S. J. Morris and J. A. Paradiso. Shoe-integrated sensor system for wireless gait analysis and real-time feedback. In *Engineering in Medicine and Biology, EMBS 2002*, volume 3, pages 2468–2469. IEEE, 2002.

[92] A. Must, J. Spadano, E. H. Coakley, A. E. Field, G. Colditz, and W. H. Dietz. The disease burden associated with overweight and obesity. *Jama*, 282(16):1523–1529, 1999.

[93] B. M. Nes, I. Janszky, L. J. Vatten, T. Nilsen, S. T. Aspenes, and U. Wisløff. Estimating VO2peak from a nonexercise prediction model: the hunt study, norway. *Med Sci Sports Exerc*, 43(11):2024–30, 2011.

[94] V. Noonan and E. Dean. Submaximal exercise testing: clinical application and interpretation. *Physical Therapy*, 80(8):782–807, 2000.

[95] D. O. Olguın and A. S. Pentland. Human activity recognition: Accuracy across common locations for wearable sensors. 2006.

[96] J. Parkka, M. Ermes, and M. van Gils. Automatic feature selection and classification of physical and mental load using data from wearable sensors. In *Information Technology and Applications in Biomedicine (ITAB), 2010 10th IEEE International Conference on*, pages 1–5. IEEE, 2010.

[97] R. Patte, M. Pratt, S. Blair, W. Haskell, C. Macera, C. Bouchard, et al. A recommendation from the centers for disease control and prevention and the american college of sports medicine. *JAMA*, 273(5):402–7, 1995.

[98] J. Penders, J. van de Molengraft, M. Altini, F. Yazicioglu, and C. Van Hoof. A low-power wireless ecg necklace for reliable cardiac activity monitoring on-the-move. In *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011.

[99] G. Plasqui and K. R. Westerterp. Accelerometry and heart rate as a measure of physical fitness: proof of concept. *Medicine and science in sports and exercise*, 37(5):872–876, 2005.

[100] G. Plasqui and K. R. Westerterp. Accelerometry and heart rate as a measure of physical fitness: cross-validation. *Medicine and science in sports and exercise*, 38(8):1510–1514, 2006.

[101] E. Redding, M. Wyon, J. Shearman, and L. Doggart. Validity of using heart rate as a predictor of oxygen consumption in dance. *Journal of Dance Medicine & Science*, 8(3):69–72, 2004.

[102] I. Romero, B. Grundlehner, and J. Penders. Robust beat detector for ambulatory cardiac monitoring. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 950–953. IEEE, 2009.

[103] M. P. Rothney, M. Neumann, A. Beziat, and K. Y. Chen. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol*, 103(4):1419–1427, Oct. 2007.

[104] L. B. Rowell, J. A. Murray, G. L. Brengelmann, and K. K. Kraning. Human cardiovascular adjustments to rapid changes in skin temperature during exercise. *Circulation Research*, 24(5):711–724, 1969.

[105] N. Ruch, F. Joss, G. Jimmy, K. Melzer, J. Hänggi, and U. Mäder. Neural network versus activity-specific prediction equations for energy expenditure estimation in children. *Journal of Applied Physiology*, 115(9):1229–1236, 2013.

[106] J. R. Ruiz, J. Ramirez-Lechuga, F. B. Ortega, J. Castro-Pinero, J. M. Benitez, A. Arauzo-Azofra, C. Sanchez, M. Sjöström, M. J. Castillo, A. Gutierrez, et al. Artificial neural network-based equation for estimating vo 2max from the 20m shuttle run test in adolescents. *Artificial intelligence in medicine*, 44(3):233–245, 2008.

[107] M. Rumo, O. Amft, G. Tröster, and U. Mäder. A stepwise validation of a wearable system for estimating energy expenditure in field-based research. *Physiological Measurement*, 32(12):1983–2001, 2011.

[108] A. Salarian, H. Russmann, F. J. Vingerhoets, P. R. Burkhard, and K. Aminian. Ambulatory monitoring of physical activities in patients with parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 54(12):2296–2299, 2007.

[109] B. Saltin and A. Gagge. Sweating and body temperatures during exercise. *International journal of biometeorology*, 15(2):189–194, 1971.

[110] A. S. Santo and L. A. Golding. Predicting maximum oxygen uptake from a modified 3-minute step test. *Research quarterly for exercise and sport*, 74(1):110–115, 2003.

[111] F. Sartor, G. Vernillo, H. M. de Morree, A. G. Bonomi, A. La Torre, H.-P. Kubis, and A. Veicsteinas. Estimation of maximal oxygen uptake via submaximal exercise testing in sports, clinical, and home settings. *Sports medicine*, 43(9):865–873, 2013.

[112] J. Seiter, O. Amft, M. Rossi, and G. Tröster. Discovery of activity composites using topic models: An analysis of unsupervised methods. *Pervasive and Mobile Computing*, 2014.

[113] G. D. Smith and J. Morris. Assessment of physical activity, and physical fitness, in population surveys. *Journal of epidemiology and community health*, 46(2):89, 1992.

[114] J. Smolander, T. Juuti, M.-L. Kinnunen, K. Laine, V. Louhevaara, K. Männikkö, and H. Rusko. A new heart rate variability-based method for the estimation of oxygen consumption without individual laboratory calibration: application example on postal workers. *Applied Ergonomics*, 39(3):325–331, 2008.

[115] D. K. Spierer, M. Hagins, A. Rundle, and E. Pappas. A comparison of energy expenditure estimates from the Actiheart and Actical physical activity monitors during low intensity activities, walking, and jogging. *European Journal of Applied Physiology*, 111(4):659–67, Apr. 2011.

[116] S. J. Strath, D. R. Bassett Jr, A. Swartz, and D. L. Thompson. Simultaneous heart rate-motion sensor technique to estimate energy expenditure. *Medicine and science in sports and exercise*, 33(12):2118–2123, 2001.

[117] R. A. Tanhoffer, A. I. Tanhoffer, J. Raymond, N. A. Johnson, A. P. Hills, and G. M. Davis. Energy expenditure in individuals with spinal cord injury quantified by doubly-labelled water and a multi-sensor armband. *Journal of physical activity & health*, 2014.

[118] E. Tapia. Using machine learning for real-time activity recognition and estimation of energy expenditure. In *PhD thesis, MIT*, 2008.

[119] L. Thorsen, E. Skovlund, S. B. Strømme, K. Hornslien, A. A. Dahl, and S. D. Fosså. Effectiveness of physical activity on cardiorespiratory fitness and health-related quality of life in young and middle-aged cancer patients shortly after chemotherapy. *Journal of Clinical Oncology*, 23(10):2378–2388, 2005.

[120] T. Tönis, K. Gorter, M. Vollenbroek-Hutten, and H. Hermens. Comparing VO2max determined by using the relation between heart rate and accelerometry with submaximal estimated VO2max. *The Journal of sports medicine and physical fitness*, 52(4):337–343, 2012.

[121] M. P. Tulppo, T. H. Mäkikallio, T. Seppänen, R. T. Laukkanen, and H. V. Huikuri. Vagal modulation of heart rate during exercise: effects of age and physical fitness. *American Journal of Physiology-Heart and Circulatory Physiology*, 274(2):H424–H429, 1998.

[122] V. T. van Hees and U. Ekelund. Novel daily energy expenditure estimation by using objective activity type classification : where do we go from here ? *Journal of Applied Physiology*, pages 639–640, 2009.

[123] V. T. van Hees, R. C. van Lummel, and K. R. Westerterp. Estimating Activity-related Energy Expenditure Under Sedentary Conditions Using a Tri-axial Seismic Accelerometer. *Obesity*, 17(6):1287–1292, Mar. 2009.

[124] L. Vanhees, J. Lefevre, R. Philippaerts, M. Martens, W. Huygens, T. Troosters, and G. Beunen. How to assess physical activity? how to assess physical fitness? *European Journal of Cardiovascular Prevention & Rehabilitation*, 12(2):102–114, 2005.

[125] H. Vathsangam, A. Emken, E. T. Schroeder, D. Spruijt-Metz, and G. S. Sukhatme. Determining energy expenditure from treadmill walking using hip-worn inertial sensors: An experimental study. *IEEE Transactions on Biomedical Engineering*, 58(10):2804–2815, 2011.

[126] H. Vathsangam, B. A. Emken, E. T. Schroeder, D. Spruijt-Metz, and G. S. Sukhatme. Hierarchical linear models for energy prediction using inertial sensors: a comparative study for treadmill walking. *Journal of ambient intelligence and humanized computing*, 4(6):747–758, 2013.

[127] M. Wei, J. B. Kampert, C. E. Barlow, M. Z. Nichaman, L. W. Gibbons, R. S. Paffenbarger Jr, and S. N. Blair. Relationship between low cardiorespiratory fitness and mortality in normal-weight, overweight, and obese men. *Jama*, 282(16):1547–1553, 1999.

[128] J. Weir. New methods for calculating metabolic rate with specific reference to protein metabolism. *J Physiol*, 109:1–9, 1949.

[129] G. J. Welk, J. McClain, and B. E. Ainsworth. Protocols for evaluating equivalency of accelerometry-based activity monitors. *Medicine and Science in Sports and Exercise*, 44(1):S39–49, 2012.

[130] G. J. Welk, J. J. McClain, J. C. Eisenmann, and E. E. Wickel. Field validation of the mti actigraph and bodymedia armband monitor using the ideea monitor. *Obesity*, 15(4):918–928, 2007.

[131] K. Westerterp, L. Wouters, and L. W. van Marken. The maastricht protocol for the measurement of body composition and energy expenditure with labeled water. *Obesity research*, 3:49–57, 1995.

[132] P. G. Weyand, M. Kelly, T. Blackadar, J. C. Darley, S. R. Oliver, N. E. Ohlenbusch, S. W. Joffe, and R. W. Hoyt. Ambulatory estimates of maximal aerobic power from foot-ground contact times and heart rates in running humans. *Journal of Applied Physiology*, 91(1):451–458, 2001.

[133] S. H. Wild, G. Roglic, A. Green, R. Sicree, and H. King. Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030. *Diabetes Care*, 27(10):2569, 2004.

[134] R. F. Yazicioglu, P. Merken, R. Puers, and C. V. Hoof. A 60 w 60 nv/hz readout front-end for portable biopotential acquisition systems. *Solid-State Circuits, IEEE Journal of*, 42(5):1100–1110, 2007.

[135] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800. ACM, 2009.