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

Marco Altini [1,2], **Julien Penders** [1], **Ruud Vullers** [1] **and Oliver Amft** [2]

[1] imec The Netherlands, High Tech Campus 31, Eindhoven, NL
[2] Eindhoven University of Technology, Den Dolech 2, Eindhoven, NL

E-mail: altini.marco@gmail.com

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.

Keywords: Energy expenditure, heart rate, galvanic skin response, respiration, physical activity, personalization, normalization

1. Introduction

Currently, epidemiologists use accelerometers and Heart Rate (HR) monitors to objectively gather information about Energy Expenditure (EE) (Assah *et al* 2010, Ceesay *et al* 1989, Crouter *et al* 2006, Ekelund *et al* 2001). Different methods to EE estimation have been developed in the past, from *counts-based* estimation methods to activity-specific EE equations, developed using one or more wearable sensors. Activity-specific models consistently showed higher performance compared to single models (Altini *et al* 2012, Bonomi *et al* 2009, Ruch *et al* 2013, Rumo *et al* 2012, Tapia 2008).

For EE estimates, the inclusion of physiological signals such as HR, Galvanic Skin Response (GSR), respiration, skin temperature or humidity, in combination with accelerometers, consistently provided better results than accelerometers alone (Altini *et al* 2012, Brage *et al* 2007, Smolander *et al* 2008, Welk *et al* 2007). However, inter-individual differences in physiology, as well as the consequent need for individual calibration, limit accuracy and practical applicability of such systems (Altini *et al* 2013a, Brage *et al* 2007, Ceesay *et al* 1989). 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:

(i) 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.

(ii) 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.

2. Related Work

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 (Crouter *et al* 2006). 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 (Brage *et al* 2007). The highly correlated relation between HR and EE within one individual changes substantially between individuals (Altini *et al* 2013b).

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 (Freedson *et al* 2011, Rothney *et al* 2007). 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 (Bonomi *et al* 2012, Ruch *et al* 2013). Others extended the single model approach, performing activity recognition over a pre-defined set of activities, and then applying different methods to predict EE (Altini *et al* 2012, Bonomi *et al* 2009, Tapia *et al* 2008, Rumo *et al* 2012). 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 (Chen *et al* 2013). 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 Altini *et al* 2012, Bonomi *et al* 2009 and Ruch *et al* 2013, 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 (Altini *et al* 2013a, Brage *et al* 2007). Partitioning the EE estimation into activity-specific sub-problems is not sufficient to address the relation between physiological signals and EE in different individuals.

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 (Tulppo *et al* 2003), differences in respiration, skin temperature or GSR might be caused by different underlying processes or characteristics of the person (Saltin and Gagge 1971). We recently investigated the relation between multiple physiological signals (HR, respiration rate, GSR and skin humidity) and EE for activity-specific EE estimation models (Altini *et al* 2013a). 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 Altini *et al* 2013b, 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.

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 2.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 2.b). Figure 2.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 2.c). Figure 1 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.

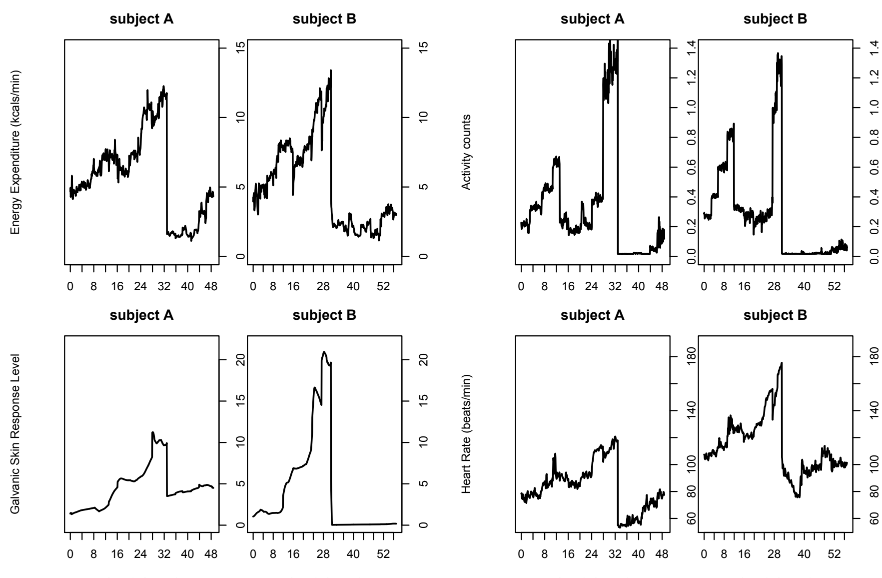

Figure 1. Reference EE, accelerometer and physiological data during a series of physical activities for two subjects with similar body size. While EE and accelerometer data show similar results and low inter-individual variability, big differences are found in both GSR and HR, highlighting the need for normalization of these parameters before their use for EE estimation.

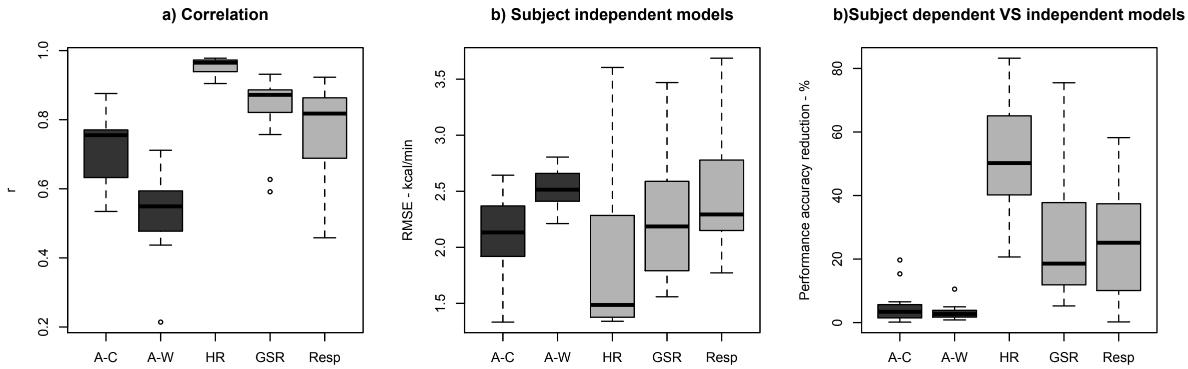

Figure 2. a) Correlation between accelerometer and physiological data with EE, b) Root Mean Square Error of subject independent EE models developed using accelerometer or physiological data, c) performance accuracy reduction when moving from subject dependent to subject independent models. A-C is accelerometer data at the chest, A-W is accelerometer data at the wrist, HR is heart rate, GSR is galvanic skin response, Resp is respiration.

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 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:

$$C = \{c_1, \ldots, c_n\}, \forall c_i \in C, \quad \exists \quad y_{act_i} = X_{act_i}\beta_{act_i} + \epsilon \tag{1}$$

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. 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 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:

$$X_{phy_n} = (X_{phy} - X_{phy_{base}})/X_{phy_{range}} \tag{2}$$

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:

$$X_{phy_{range}} = X_{phy_{high}} - X_{phy_{base}} \tag{3}$$

$X_{phy_{high}}$ 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 3, blocks *a,b*):

$$X_{phy_{high}} = X_{ADL}\beta_{ADL} + \epsilon \tag{4}$$

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.

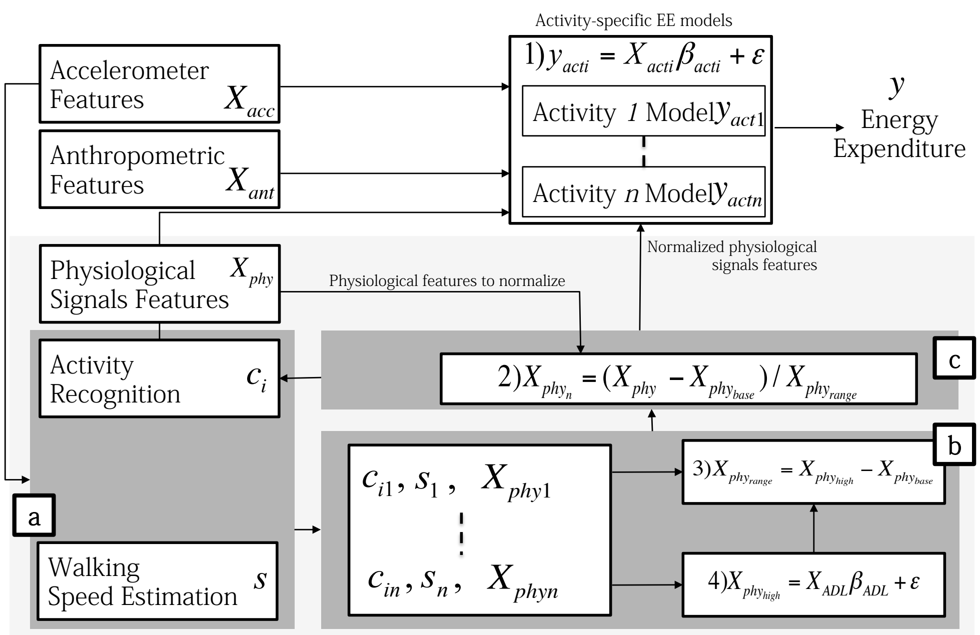

Figure 3. Overview of our method to normalize physiological signals and estimate EE. Normalized physiological signals are used for activity recognition and EE estimation models. a) components required for the recognition of type and intensity of ADLs, b) components for the estimation of normalization parameters and c) equation used to normalize physiological signals.

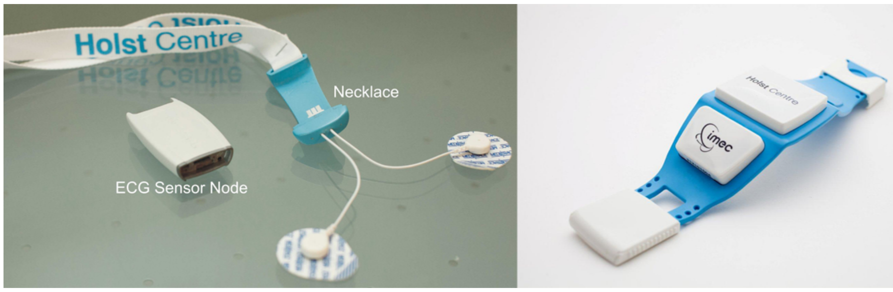

Figure 4. The two wearable sensors used in this experiment, ECG Necklace (left) and Wristband (right).

5. Measurement setup and data collection

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.

5.2. Instruments

Two wearable sensors were used for data collection, imec's ECG Neckalce and Wristband (see figure 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 $128Hz$ and accelerometer data at $32Hz$. Additionally, reference EE was collected using the *Cosmed* $K4b^2$ indirect calorimeter (McLaughlin *et al* 2001).

5.3. Experiment 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 (Basset *et al* 2012). 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.

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. 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. Methods implementation

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.

6.1.1. 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 (Altini *et al* 2012, Bonomi *et al* 2009).

Table 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

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 (Battiti *et al* 1994), 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 Altini *et al* 2012. 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.

6.2. 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.

6.3. 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 Browning and Kram 2005 and 5 ± 0.8 km/h in Minetti *et al* 2003). 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:

$$X_{ADL} = [X_{phyLying}, X_{phyWalking4}, X_{phyWalking5}, X_{phyWalking6}] \tag{5}$$

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 4, removing the *baseline* and dividing by the estimated *range*.

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.

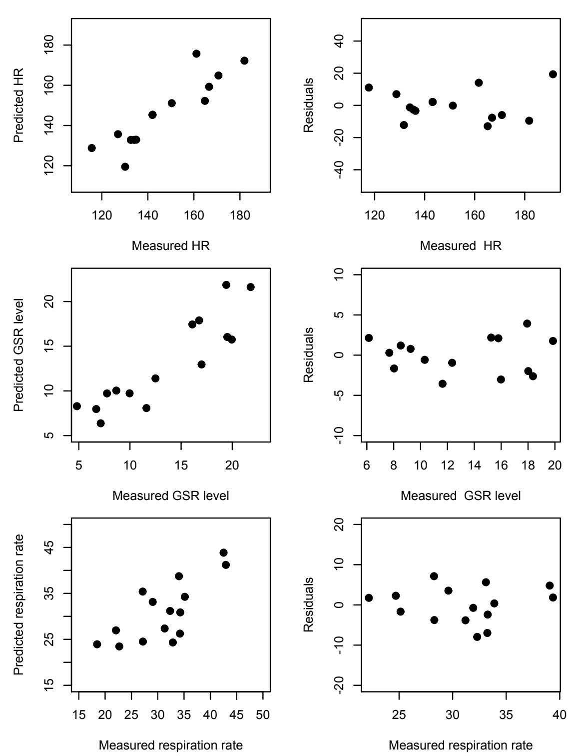

Figure 5. Scatterplot and residuals per study participant of measured (running on a treadmill) vs. predicted (from physiological signals during ADLs) physiological signals values during a *high intensity activity* ($X_{phy_{high}}$). $X_{phy_{high}}$ is used together with $X_{phy_{base}}$ to determine the *range* and normalize the signals.

7. Results

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 $\pm$ 0.09 km/h. Both models were previously reported in Altini *et al* 2013b. 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 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.

7.2. Personalized activity-specific EE estimation

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 ($p = 0.04 < \alpha$).

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 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

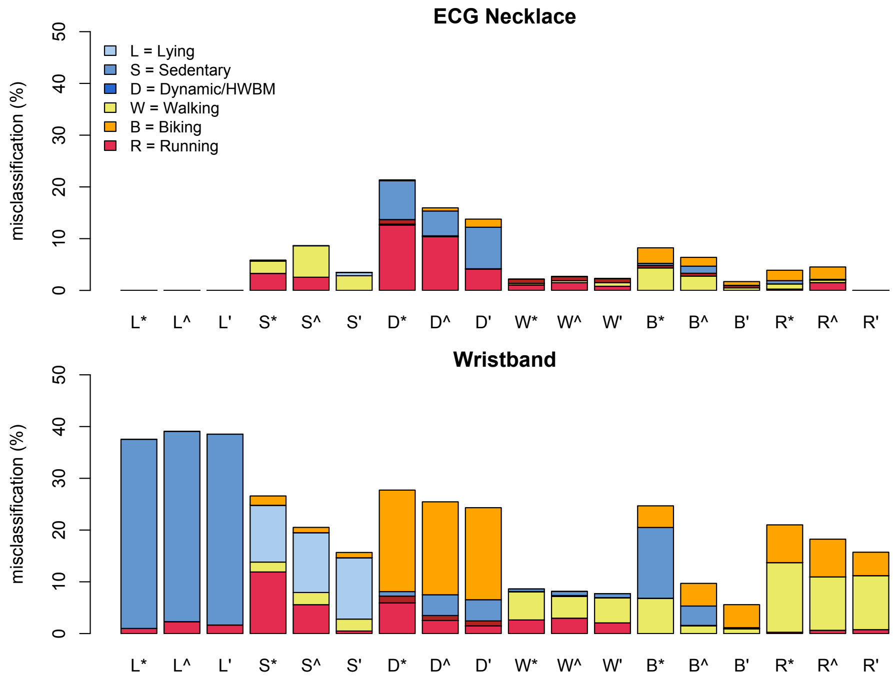

Figure 6. Misclassification of the activity recognition models per cluster of activities. *indicate accelerometer-only models, Λ indicate models combining accelerometer and physiological data, ' indicate models combining accelerometer and normalized physiological data.

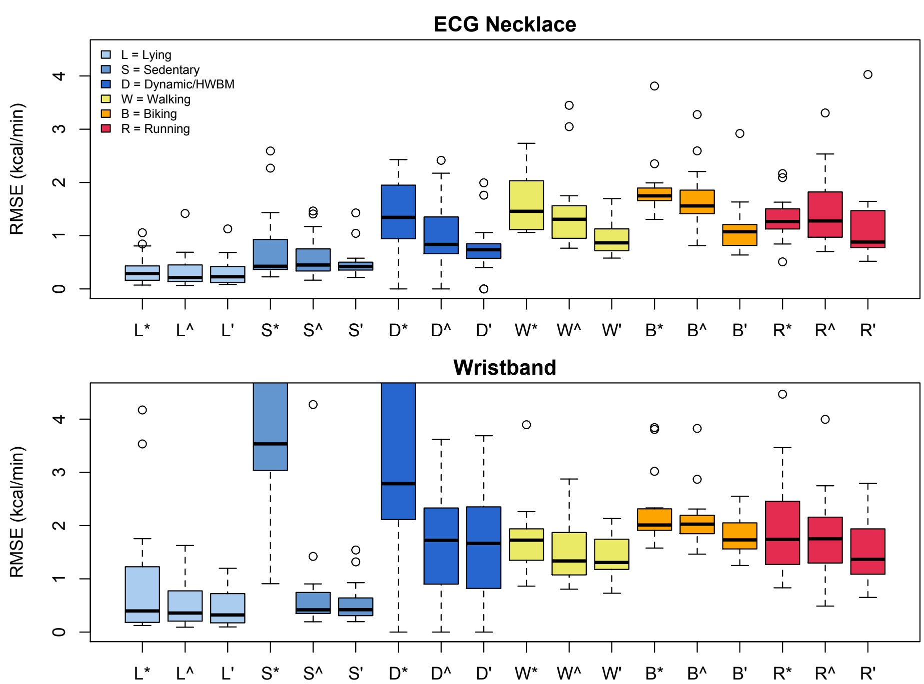

Figure 7. RMSE of activity-specific EE estimation models for the ECG Necklace and Wristband sensors, including misclassification effects. *indicate accelerometer-only models, Λ indicate models combining accelerometer and physiological data, ' indicate models combining accelerometer and normalized physiological data.

Table 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)**

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 2.

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 normalization 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.

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 different 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.

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, 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).

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.

References

Altini M, Penders J and Amft O 2012. Energy Expenditure estimation using wearable sensors: a new methodology for activity-specific models. *In Proceedings of Wireless Health 2012.*

Altini M, Penders J, Vullers R and Amft O 2013a. Combining wearable accelerometer and physiological data for activity and energy expenditure estimation. *In Proceedings of Wireless Health 2013.*

Altini M, Penders J, and Amft O 2013b. Personalizing energy expenditure estimation using a cardiorespiratory fitness predicate. *In Pervasive Health 13, in-press.*

Assah FK, Ekelund U, Brage S, Wright A, Mbanya JC and Wareham NJ 2010. 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. *Int J Epidemiol* 40(**1**): 112-120.

Bassett DR, Rowlands A, and Trost SG 2012 Calibration and validation of wearable monitors. *Med Sci Sports Exerc*, 44(1 Suppl 1):S32S38.

Battiti R 1994. Using mutual information for selecting features in supervised neural net learning. *Neural Networks, IEEE Transactions on*, 5(**4**):537550.

Bonomi AG, Plasqui G, Goris AHC and Westerterp KR 2009 Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. *Journal of Applied Physiology*, 107(**3**):655661.

Bonomi AG and Plasqui G 2012 Divide and conquer: assessing energy expenditure following physical activity type classification. *Journal of applied physiology*, 112(**5**):932.

Brage S, Ekelund U, Brage N, Hennings MA, Froberg K, Franks PW, and Wareham NJ 2007. Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *J Appl Physiol*, 103:68292.

Browning RC and Kram R 2005 Energetic cost and preferred speed of walking in obese vs. normal weight women. *Obes Res.* 13: 891899.

Ceesay S M, Prentice A M, and Day KC 1989 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**):175186, Mar. 1989.

Chen S, Altini M, Penders J, Lach J, and Amft O 2013. Unsupervised activity clustering to estimate energy expenditure with a single body sensor. *In BSN 13, in-press.*

Crouter S, Churilla J, and Bassett D 2006. Estimating energy expenditure using accelerometers. *European Journal of Applied Physiology*, 98(**6**):601612, Dec.

Ekelund U, Sjostrom M, Yngve A, Poortvliet E, Nilsson A, Froberg K, Wedderkopp N, Westerterp K 2001. Physical activity assessed by activity monitor and doubly labeled water in children. *Med Sci Sports Exerc* 33: 275281.

Freedson P, Lyden K, Kozey-Keadle S and Staudenmayer J 2011 Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: Validation on an independent sample. *Journal of Applied Physiology.* 111:1804-1812.

McLaughlin JE *et al* 2001. Validation of the COSMED K4 b2 portable metabolic system. *International journal of sports medicine*, 22(**4**):280284.

Minetti AE, Boldrini L, Brusamolin L, Zamparo P and McKee T 2003 A feedback controlled treadmill (treadmill-on-demand) and the spontaneous speed of walking and running in humans. *J Appl Physiol.* 95: 838843.

Rothney MP, Neumann M, Beziat A, and Chen KY 2007. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol*, 103(**4**):14191427.

Ruch N, Joss F, Jimmy G, Melzer K, Hnggi J, and Mder U 2013. Neural network versus activity-specific prediction equations for energy expenditure estimation in children. *Journal of applied physiology.*

Rumo M, Amft O, Troster G, and Mader U 2011 A stepwise validation of a wearable system for estimating energy expenditure in field-based research. *Physiological measurement*, 32(**12**):19832001.

Saltin B and Gagge AP 1971. Sweating and body temperatures during exercise. *Int J Biometeorol* 15: 189194.

Smolander J, Kinnunen ML, Laine K, Louhevaara V, Mannikko K and Rusko H 2008. 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**):325331.

Tapia EM 2008. Using machine learning for real-time activity recognition and estimation of energy expenditure. *In PhD thesis*, MIT.

Tulppo MP, Hautala AJ, Makikallio TH, Laukkanen RT, Nissila S and Hughson RL 2003. Effects of aerobic training on heart rate dynamics in sedentary subjects. *J Appl Physiol.* 95(**1**):364-72.

Weir J 1949. New methods for calculating metabolic rate with specific reference to protein metabolism. *J Physiol*, 109:19.

Welk GJ, McClain JJ, Eisenmann JC, and Wickel EE 2007. Field validation of the mti actigraph and bodymedia armband monitor using the ideea monitor. *Obesity*, 15(**4**):918928.