Personalization of Energy Expenditure Estimation in Free Living Using Topic Models

Marco Altini[1], Pierluigi Casale[2], Julien Penders[2] and Oliver Amft[3]

***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.**

***Index Terms*—Context, Energy Expenditure, Heart Rate, Topic Models**

I. Introduction

Wearable technology can provide novel insights on the relation of physical activity (PA) and health [1]. Energy expenditure (EE) is the most common parameter used to quantify PA [2], and is typically estimated using acceleration and heart rate (HR) sensors [3], [4]. 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 [5], [6], [7]. Using HR in activity-specific regression equations showed consistent improvements in EE estimation compared to using acceleration only [8], [9]. However, HR during an activity is specific to a person since it depends on the individual's cardiorespiratory fitness (CRF) level [10]. 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 [11]. 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 [8]. 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 normalization. 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.

[1]M. Altini is with Bloom Technologies, Diepenbeek, BE, and Eindhoven University of Technology, NL (e-mail: altini.marco@gmail.com)

[2]P. Casale and J. Penders are with imec-NL, Eindhoven, NL

[3]O. Amft is with University of Passau, DE, and Eindhoven University of Technology, NL

II. Related Work

Accelerometer and HR monitors are the most commonly used devices for EE estimation [3], [4]. 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 [5], [6], [7]. Including HR data in the activity-specific linear models showed consistent improvements in EE estimation accuracy compared to algorithms using accelerometer only data [8], [9]. 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 [8]. Fig. 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 [10]. Thus, estimating EE based on HR results in under and overestimations [8], [12].

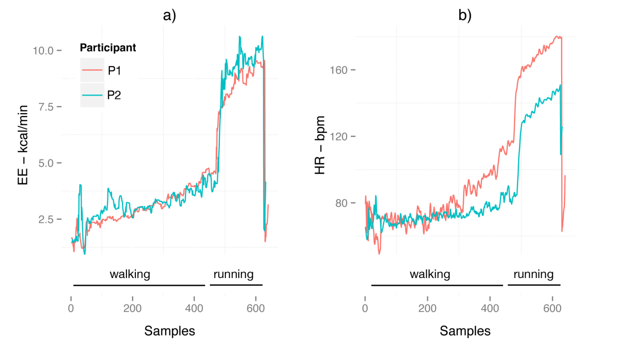


Fig. 1. Relation between EE and HR in two participants during walking and running activity primitives. a) Absolute EE levels are similar due to similar body weight. b) HR differs between participants due to different CRF level (VO_2max participant 1 is 2104 ml/min, VO_2max participant 2 is 3130 ml/min).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

A. Personalized EE Estimation

HR showed higher correlation with EE compared to accelerometer data [12]. However, subject-independent models including HR performed sub-optimally, confirming the need for individual calibration [12]. 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 [8], 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.

B. 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 could potentially improve interpretation of HR or other physiological data in free living. Fig. 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 [13]. Activity primitives are typically considered as a set of atomic activities that can be determined on a short time window [6], 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 [5], [7]. 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 [14]. For activity recognition, the same concept was applied to discover activity composites from activity primitives [13]. Recent work investigated the impact of multiple latent Dirichlet allocation (LDA) parameters for activity composites discovery, showing promising results [15]. 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*.

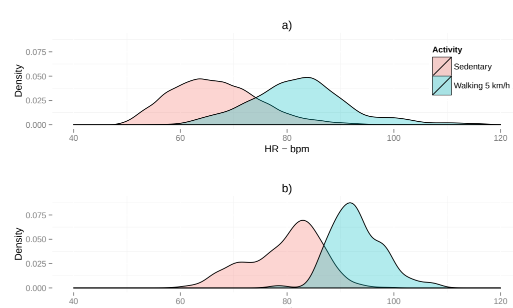


Fig. 2. HR during activity primitives and activity composites performed in free living by one participant. Activity composites were manually annotated by the participant. HR during the same activity primitives changes substantially depending on the activity composites: a) work, b) social interactions.

III. 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. III-B. Following a top down approach, EE was estimated by activity-specific models (see Fig. 3). For each activity primitive c_i, a regression model is defined:

$$\begin{aligned} C = \{c_1, \ldots, c_{cn}\}, \quad \forall c_i \in C, \\ \exists \quad y_{act_i} = X_{act_i}\beta_{act_i} + \epsilon \\ X_{act_i} = \{X_{acc_i}, X_{ant_i}, X_{hr_i}\} \end{aligned} \tag{1}$$

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. 3.

A. HR Normalization Parameter Estimation

Normalized HR was obtained as shown in Fig. 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:

$$X_{hr} = \frac{HR}{HR_{np}} \tag{2}$$

$$HR_{np} = \overline{HR}_{ctx*}\beta_{np} + \epsilon \tag{3}$$

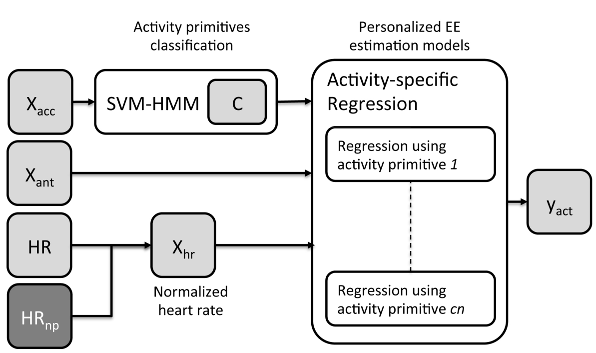


Fig. 3. Proposed approach to personalized EE estimation. HR data HR were normalized by the HR normalization parameter HR_{np}, resulting in the normalized HR X_{hr}, before being used in activity-specific EE models.

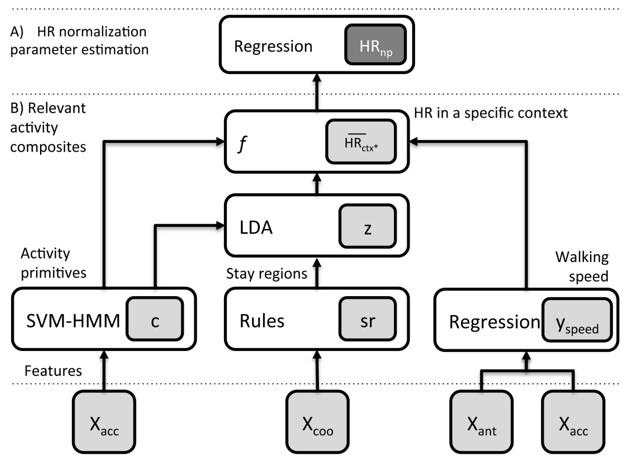


Fig. 4. Proposed approach to determine the HR in a specific context $\overline{HR}_{ctx*}$, i.e. HR while walking at a certain speed during *relevant activity composites*, and estimate the HR normalization parameter HR_{np}. Activity primitives c and stay regions sr are determined from accelerometer 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, $\overline{HR}_{ctx*}$.

$\overline{HR}_{ctx*}$ 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. V). 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:

$$\theta_s \sim \mathrm{Dir}(\alpha) \quad 1 \leq s \leq S \tag{4}$$

$$z_n \sim \mathrm{Mult}(\theta_s) \quad 1 \leq s \leq S, \quad 1 \leq n \leq N \tag{5}$$

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, $p(y_n|z_n)$:

$$y_n \sim \mathrm{Mult}(\beta_{z_n}) \quad 1 \leq n \leq N \tag{6}$$

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

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

$$\prod_{s=1}^{S} \int p(\theta_s, \alpha) \prod_{n=1}^{N} \sum_{z=1}^{K} p(z_{sn|\theta_s}) p(y_{sn}|z_{sn}, \beta) d\theta_s \tag{7}$$

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.

B. 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. 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. 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. 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. 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:

$$r_T = \{r_{rank_{T1}}, \ldots, r_{rank_{TN}}\}, \tag{8}$$

$$r_{rank_i} = r(\overline{HR}_{ctx_{par=\{1,\ldots,npar\},i}}, HR_{np_{par=\{1,\ldots,npar\}}}) \tag{9}$$

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. 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.

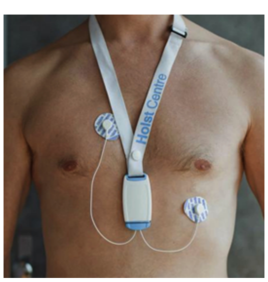


Fig. 5. ECG Necklace, the wearable sensor used to collect accelerometer and ECG data in this study. The ECG Necklace was worn during laboratory protocols and free living recordings close to the body's center of mass

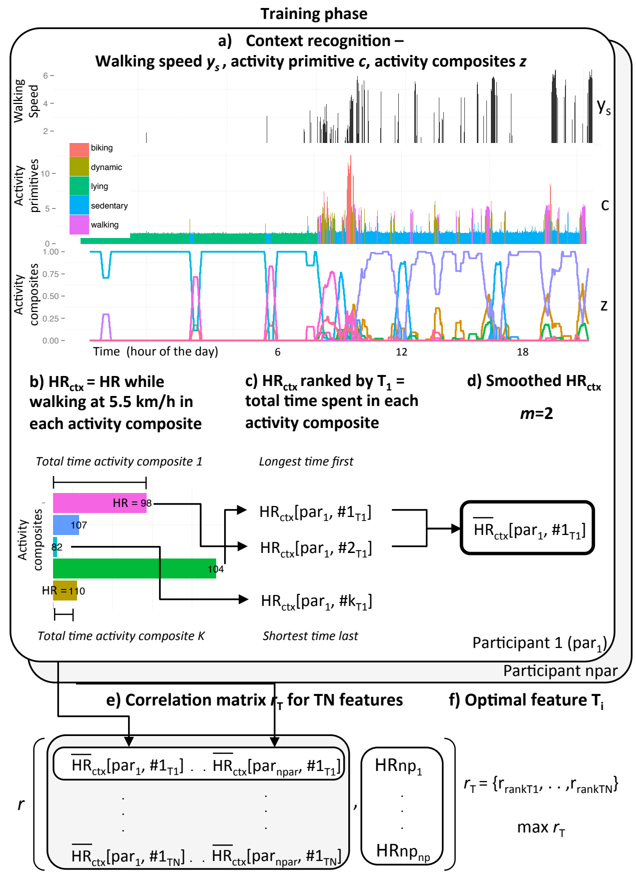


Fig. 6. Exemplary diagram of our approach to discover *relevant activity composites* for the case of walking at 5.5 km/h. $a-b$) Walking speed y_s, activity primitives c and activity composites z are used to determine HR in specific contexts, HR_{ctx}. c) HR_{ctx} are ranked by activity composite feature 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*.

IV. Evaluation Study

A. Participants and Data Acquisition

Participants were 34 (14 male, 20 female), 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. 5). 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 [7]. The ECG Necklace was worn during laboratory protocols and free living recordings. 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 [16]. VO_2max was determined during an incremental test on a cycle ergometer [17]. 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.

B. Experiment 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.

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 [18].

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.

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.

V. Implementation

A. 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 [19], 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.

B. 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.

C. 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

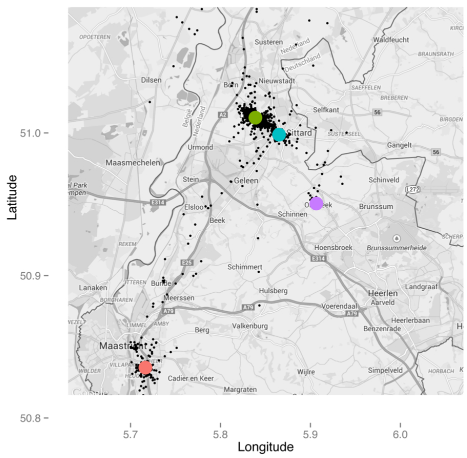


Fig. 7. Exemplary stay regions detection from noisy GPS data for one participant. Small dots represent all recorded GPS data points, while bigger dots represent detected stay regions.

selection [8]. Free living walking speeds used to contextualize HR were 4.5 km/h (4 to 5 km/h 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 [20] and $5{\pm}0.8$ km/h in [21]).

D. 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 [22]. 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. 7.

E. 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 [15]. Parameters were optimized using an implementation of the variational expectation-maximization algorithm proposed in [14]. The function f (see Fig. 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}.

F. 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 [8] as well as others [6], 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. IV-B3 as benchmarks for the proposed approach (referred to as *combined*).

G. Personalized EE Estimation

EE was estimated by first classifying the activity performed among the ones listed in Sec. V-B 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. IV-B3 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*).

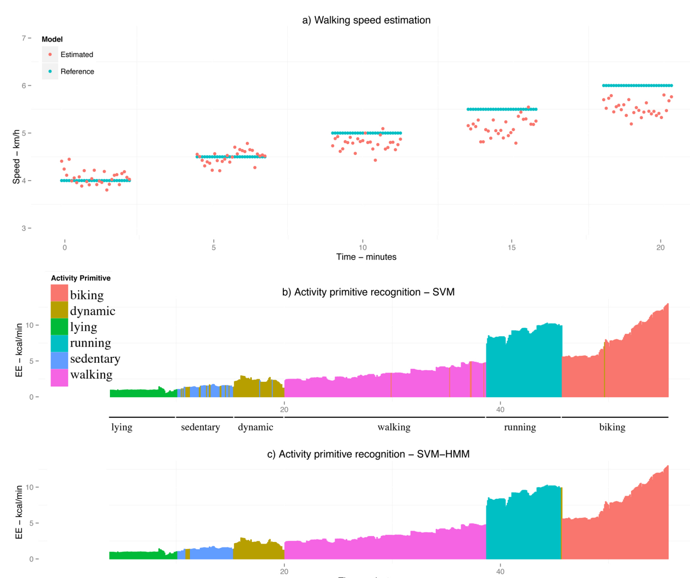


Fig. 8. Exemplary walking speed estimation and activity primitives recognition for one participant. Activities were manually annotated and performed sequentially. Improvements in activity primitives recognition using a combined SVM-HMM compared to a single SVM are shown in plots *b* and *c*.

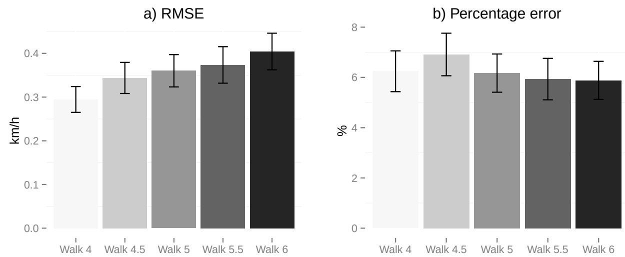


Fig. 9. a) RMSE for walking speed models across the speed range used to contextualize HR in free living. b) Percentage error across the same speeds.

VI. Results

A. 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. 9, while an exemplary output of the activity primitives recognition classifier and walking speed estimation model is shown in Fig. 8.

B. HR normalization parameter

An example of LDA-derived activity composites is shown in Fig. 6. Activities composites were ranked according to the features listed in Sec. V-E. 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. 10 shows the relation between measured and predicted HR_{np} for the different cases considered in this work.

C. EE estimation

EE estimation results are shown in Fig. 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.

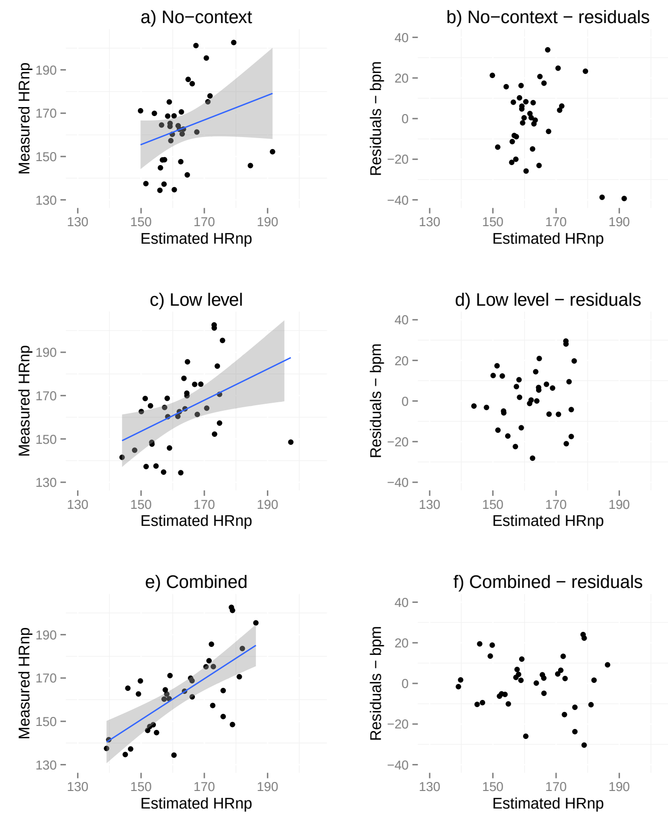


Fig. 10. a,c,e) Relation between measured and estimated HR normalization parameters for the three conditions compared in this work: a) no-context, c) low level, e) combined. b,d,f) Residuals plots for the three conditions compared in this work: b) no-context, d) low level, f) combined. For low level and combined, only data while walking at 5.5 km/h was used, as it provided the optimal results (see Sec. VI-B).

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 [12]. 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.

VII. 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 [8], [12], [6]. 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

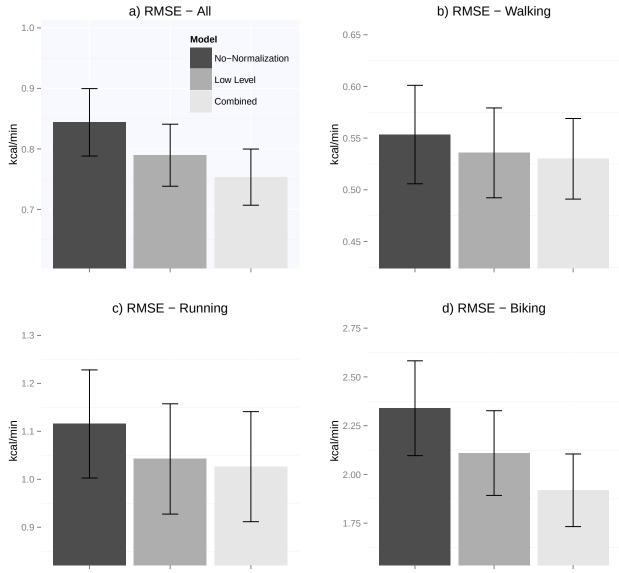


Fig. 11. EE estimation RMSE and standard errors for a) all activities averaged, b) walking c) running and d) biking. Three models are compared, *No-normalization*, *Low Level*: HR was normalized using a normalization factor predicted from HR while walking at 5.5 km/h, and *Combined*, i.e. the proposed approach, normalizing HR using a normalization factor predicted from HR while walking at 5.5 km/h during *relevant activity composites*.

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 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 [13], [15], 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 [23]. 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 [24], [25]. 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 [26]. 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 [20] , [21] 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.

References

[1] 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*, vol. 54, no. 12, pp. 2296–2299, 2007.

[2] 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*, vol. 58, no. 10, pp. 2804–2815, 2011.

[3] 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*, vol. 61, no. 2, pp. 175–186, Mar. 1989.

[4] S. E. Crouter and D. R. Bassett, "A refined 2-regression model for the actigraph accelerometer," *Medicine & Science in Sports & Exercise*, vol. 42, no. 5, pp. 1029–1037, 2010.

[5] A. G. Bonomi, "Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer." *Journal of Applied Physiology*, vol. 107, no. 3, pp. 655–661, 2009. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/19556460

[6] E. Tapia, "Using machine learning for real-time activity recognition and estimation of energy expenditure," in *PhD thesis, MIT*, 2008.

[7] 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*, no. 99, p. 1, 2014.

[8] 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*. IEEE, 2013, pp. 65–72.

[9] 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*, vol. 96, no. 1, pp. 343–351, Aug. 2003. [Online]. Available: http://dx.doi.org/10.1152/japplphysiol.00703.2003

[10] 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*, vol. 24, no. 5, pp. 711–724, 1969.

[11] 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*, vol. 7, no. 2, pp. 218–221, 1954.

[12] M. Altini, J. Penders, R. Vullers, and O. Amft, "Personalizing energy expenditure estimation using physiological signals normalization during activities of daily living," *Physiol Meas*, vol. 35, no. 9, p. 1797, September 2014.

[13] T. Huynh, M. Fritz, and B. Schiele, "Discovery of activity patterns using topic models," in *Proceedings of the 10th international conference on Ubiquitous computing*. ACM, 2008, pp. 10–19.

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

[15] 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.

[16] J. Weir, "New methods for calculating metabolic rate with specific reference to protein metabolism," *J Physiol*, vol. 109, pp. 1–9, 1949.

[17] 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*, vol. 6, no. 04, pp. 197–201, 1985.

[18] 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*, vol. 3, pp. 49–57, 1995.

[19] 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*, ser. WH '12. New York, NY, USA: ACM, 2012, pp. 1:1–1:8. [Online]. Available: http://doi.acm.org/10.1145/2448096.2448097

[20] R. C. Browning and R. Kram, "Energetic cost and preferred speed of walking in obese vs. normal weight women," *Obesity Research*, vol. 13, no. 5, pp. 891–899, 2005.

[21] 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*, vol. 95, no. 2, pp. 838–843, 2003.

[22] 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*. ACM, 2009, pp. 791–800.

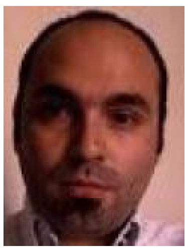
[23] E. Redding, M. Wyon, J. Shearman, and L. Doggart, "Validity of using heart rate as a predictor of oxygen consumption in dance," *Journal of Dance Medicine & Science*, vol. 8, no. 3, pp. 69–72, 2004.

[24] 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*, vol. 61, no. 2, pp. 566–575, 2014.

[25] 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.

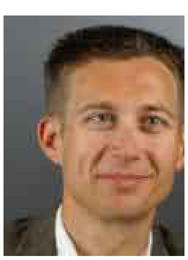
[26] M. Altini, J. Penders, R. Vullers, and O. Amft, "Personalized physical activity monitoring on the move," in *Proceedings of the 4th Conference on Wireless Health*, ser. WH '13. New York, NY, USA: ACM, 2013, pp. 8:1–8:2.

Marco Altini received the M.Sc. degree cum laude in engineering and computer science from the University of Bologna in 2010. He is currently pursuing the Ph.D. degree with the Technical University of Eindhoven. His current research interests include development and implementation of machine learning techniques for health and wellbeing applications, combining multiple data sources.



Pierluigi Casale is researcher at Holst Center-IMEC, and former post-doc at TU Eindhoven. He received the M.Sc. in Electronic Engineering from the University of Rome in 2007 and the Ph.D. in Applied Mathematics from the University of Barcelona in 2011. His main research interests are Machine Learning and Pattern Recognition methods for Wearable Sensors with applications in human behavior analysis, healthcare and assisted living.

Julien Penders received the M.Sc. degree in systems engineering from the University of Liege, Liege, Belgium, and the M.Sc. degree in biomedical engineering from Boston University, Boston, MA, in 2004 and 2006, respectively. He is currently the Program Manager with the Holst Centre/IMEC, Eindhoven, The Netherlands. He has authored or co-authored over 50 papers in journals and conference proceedings on body area networks.



Oliver Amft received the M.Sc. from TU Chemnitz in 1999 and the Ph.D. from ETH Zurich in 2008, both in Electrical Engineering and Information Technology. Oliver is a Full Professor (W3) and Chair of Sensor Technology at University of Passau. He is also affiliated with the Wearable Computing Lab, ETH Zurich (CH) and the Signal Processing Systems section at TU Eindhoven (NL). Oliver has co-authored more than 90 publications in this field.