Energy Expenditure Estimation Using Wearable Sensors: A New Methodology for Activity-Specific Models

Marco Altini, Julien Penders
Holst Centre - imec The Netherlands
High Tech Campus 31
Eindhoven, The Netherlands
marco.altini@imec-nl.nl

Oliver Amft
ACTLab, Signal Processing Systems
Eindhoven University of Technology
Eindhoven, The Netherlands

ABSTRACT

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

Categories and Subject Descriptors

J.3 [**Computer Applications**]: Life & medical sciences—*Health*

General Terms

Algorithms, Experimentation

Keywords

Energy Expenditure, Activity Recognition

1. INTRODUCTION

Lack of physical activity is one of the major health problems in most of the western world. Even though our genome has not changed much over the last ten thousand years and more [11], activity patterns of our hunter-gatherers ancestors have been first modified by the agricultural and industrial revolution, and then completely disrupted by the shift towards computer-based work which took place over the past twenty years. As a result, two thirds of the world population is overweight and obesity affects a third of the population in the US at present. Other diseases, such as diabetes, are rapidly becoming widespread epidemics as well [26]. Accurate quantification and assessment of habitual physical activity in ambulatory settings is essential in order to find subtle but important links between not only sedentary time, but all the aspects of habitual physical activity, and health [15]. New technologies, seamlessly integrated in everyone′s life, able to monitor objectively and non-invasively our behavior, can provide unprecedented insights on these links.

Currently, epidemiologists use accelerometers [13] and HR monitors [10] to objectively gather information about physical activity. Traditionally, they make use of regression equations developed using data acquired over a certain protocol [5, 10, 13] to predict EE. For accelerometers, the rationale behind this approach is that body motion measured close to the body center of mass, is linearly related to EE. On the other hand, HR monitors exploit the linear relation between HR and oxygen uptake. Limitations of these approaches are the inability of single accelerometers worn close to the body center of mass to detect low and upper body motion [10, 21], the low accuracy of HR monitors during sedentary behavior and the need for individual calibration [8].

Recent work [4] showed that activity type can be reliably detected with wearable sensors, opening new opportunities for EE monitors. Over the last years, a few activity-specific algorithms have been reported [3, 7, 19, 21]. They first recognize the activity performed, and then apply a model developed for the specific activity, showing consistent improvements compared to previous methods. What is not clear at this stage, is the methodology to follow when developing such an algorithm. There is currently little agreement in literature, on which activities to detect. Some [3, 21] used multi-accelerometer systems and extensive protocols to detect a large amount of activities (26 and 52 respectively), exploiting the fact that frequent misclassification of the activities will most likely result in small EE errors, due to the similarity in the movement involved. Others, [7, 19] developed either multi-sensor or single accelerometer systems able to recognize a smaller set of activities with higher accuracy. Once the activity set has been selected, even less agreement is found on how to predict EE given an activity. Some works assign static *Metabolic equivalents* (METs, the ratio of metabolic rate during a specific activity to a

Wireless Health ′12, October 23-25, 2012, San Diego, USA

reference metabolic rate), combined with the subjects′ anthropometric parameters or fitness indicators such as the HR at rest [3]. Others applied a linear regression equation for each model [23, 21, 19]. At this stage, it is not clear whether combining METs values and regression models could provide better estimates [22], and whether each activity-specific regression model requires the same parameters. As a matter of fact, differences in protocols and evaluation measures make it impossible to compare the different approaches.

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

2. RELATED WORK

2.1 Epidemiological Studies

Accelerometers and HR monitors are the most commonly used single sensor devices in epidemiologic studies. Accelerometers use *activity counts*, a unit-less measure representative of whole body motion, as the independent variable in the linear regression model developed to predict EE [13]. Shortcomings of single regression models are; *a)* the accuracy of the monitor is highly dependent on the activities used to develop the model, *b)* a single model does not fit all the activities, since the slope and intercept of the regression model change based on the activity performed while data is collected [23, 21]. As a result, even when activity counts are representative of EE, the output can be misleading. Additionally, activity counts are defined differently by each sensor′s manufacturer (i.e. *Actigraph counts*, and the equations derived from them, are not directly comparable to *Actical* or *Actiheart counts* [25]). HR monitors suffer from different problems, the most common being the low accuracy during sedentary behavior [10], given that HR is affected by many other factors (e.g. stress and emotions), and the need for individual calibration [8]. Some of the issues have been tackled developing models that use more than one linear regression equation, such as Crouter′s 2-regression model [12] or Brage′s [9] branched equations. Even though these methods are promising, especially the ones combining HR and ACC data, they have shown limited improvements compared to ACC based simple linear regression models [5, 20].

2.2 Methods Based on Machine Learning

The latest monitors go towards two directions. Both strategies make use of pattern recognition and machine learning techniques. Some authors applied these methods to directly estimate EE from ACC features [14, 18]. Others, extended Crouter′s and Brage′s approach, performing activity recognition over a pre-defined set of activities, and then applying different methods to predict EE [3, 7, 23, 19, 21]. Given the significant amount of work adopting activity recognition as a first step to estimate EE, and the consistent improvements obtained [6], we believe this is the best methodology to follow when developing such algorithms. The principle behind activity recognition as a first step in EE estimation is that the slope and intercept of the regression models change based on the activity performed [23]. Tapia [21] developed a system composed of three accelerometers and applied a different regression equation for each activity classified. The regression models use ACC features as independent variables. The system can recognize about fifty activities with 50% accuracy in a subject-independent manner. *Root Mean Square Error* (RMSE) was reduced from 2.7 to 1.0 $METs$ compared to Crouter′s approach. Bonomi [7] proposed a similar approach, but with a single sensor device, mounted on the lower back. His system recognizes six clusters of activities and assigns a MET value to each one of them. *Total Energy Expenditure* (TEE) was validated against *Doubly Labeled Water* and showed accuracy up to $1MJoules/day$ when simple anthropometric parameters are used as independent variables together with the assigned METs. van Hees [23] also used a single sensor able to distinguish four activities, and then applied linear regression using a measure of motion intensity as the independent variable (similarly to Tapia). Albinali [3] developed a multi-sensor system, composed of three accelerometers able to distinguish twenty-two activities with 26% accuracy (subject-independent). He extended Bonomi′s approach, developing a custom MET table, which takes into account anthropometric variables, as well as the HR at rest, to predict EE more accurately at the individual level. This method showed improvements up to 15% compared to non-activity-based models. Rumo [19] also combined HR and ACC. His system consists of three sensors, two accelerometers and a HR belt, and can classify seven types of activities. Manual selection and the bootstrapping method were used to determine which independent variables to adopt for the activity-specific models. RMSE for the individual models ranged from 2.2 to 9.7 $KJoules$.

2.3 Towards Activity-Specific EE estimation

In this section we analyze shortcomings of state of the art activity-specific EE algorithms. More specifically, we believe the following limitations should be tackled; *a)* Activity-specific models that assign METs values to each activity classified assume that EE is constant within a cluster of activities [3, 7]. Nevertheless, most activities can be performed at different intensities, and including information about whole body motion or other features (ACC or HR) representative of variations in EE within an activity, would improve the estimate. *b)* Activity-specific regression models apply linear regression (e.g. using activity counts) even though there is no whole body motion involved [21], and therefore the motivation for applying linear regression, which is the linear relation between intensity of motion and EE, does not hold anymore. *c)* Activity-specific models that assign custom METs values [3] should carefully select the independent variables used. For example HR at rest, which is often used as an index of *cardiorespiratory fitness*, should not be used to predict EE at rest, since cardiorespiratory fitness is not related to *Basal Metabolic Rate* (BMR). Care should be taken when considering anthropometric variables as well. The energy cost of activities such as walking depends on body weight, while the assumption does not apply to biking or sitting.

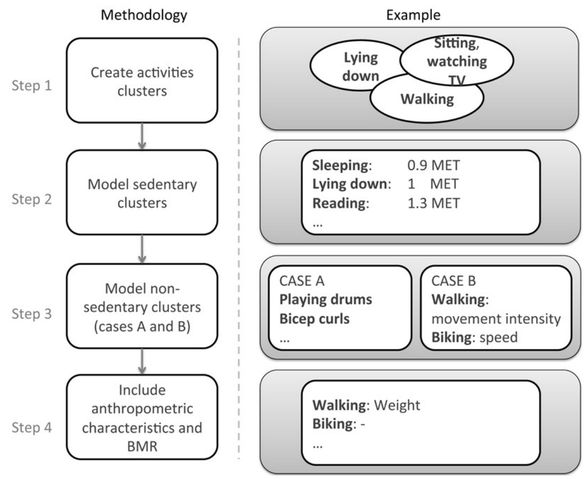

Figure 1: Block diagram of the proposed methodology, and example of an application.

3. A NEW METHODOLOGY

In this section we present our new approach to activity-specific EE modeling, and apply this methodology to the case of imec′s ECG necklace. Our approach to combine static METs with activity-specific regression equations takes four steps to derive an EE model. First, we categorize activities into clusters meaningful for EE estimates. Secondly, we separate sedentary and non-sedentary activities, and assign a static EE value to sedentary ones. Then, we examine the motion patterns of non-sedentary clusters to select the best independent variables for the prediction models. Finally, we include anthropometric characteristics to take into account differences in body size (see Figure 1).

Step 1- Create Activities Clusters

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

Step 2 - Model Sedentary Clusters

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

Step 3 - Model Non-Sedentary Clusters

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

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

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

Step 4 - Include Anthropometric Characteristics and Resting Metabolic Rate

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

4. USE CASE - A NECKLACE MONITOR

In this section we apply our proposed methodology to the use case of necklace which combines ACC and HR data in a single monitoring device. Details on the necklace and the experimental protocol can be found in Sec. 5.2.1 and 5.3.

4.1 Method Implementation

Step 1 - Create Activities Clusters

Different clusters of activities have been evaluated based on their impact on EE and the ability of a single device placed on the chest to detect them. By using a single monitoring device located at the chest it was not feasible to differentiate between sitting and standing. Therefore we grouped activities in the following clusters: *lying, low whole body motion (LWBM), high whole body motion (HWBM), walking, biking*

Table 1: Distribution of the activities into the six clusters used for activity recognition.

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

and *running* - see Table 1. LWBM and HWBM are clusters similar to the *sitting-standing* and *active standing* introduced in [7] to distinguish between sedentary and household activities with a single accelerometer, and are useful in isolating sedentary behavior even when sitting and standing cannot be distinguished.

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

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

Step 2 - Model Sedentary Clusters

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

Step 3 - Model Non-Sedentary Clusters

Using a single monitoring device sets limits regarding the number of distinguishable activities. Thus, the other activity clusters contain some variability in EE and have been modeled as non-sedentary clusters (Case B in Sec 3). What ACC features are representative of variations in EE within these cluster? LWBM and HWBM are clusters involving diverse and irregular motion patterns, that we captured using features representative of intensity and variability of motion over the three axes. Walking, running and biking involve repetitive patterns, that can be easily captured using measures of motion intensity or motion speed (see Table 4 for details). All models include the Heart Rate above Rest (HRaR) to complement acceleration features in capturing differences in EE, e.g. walking vs. walking carrying weights.

Step 4 - Include Anthropometric Characteristics and Resting Metabolic Rate

Body weight was included for all the clusters involving ambulation (HWBM, walking and running). No anthropometric variables were included for lying, LWBM and biking. RMR was in included in models not involving ambulation (lying, LWBM and biking). We computed RMR using simple anthropometric variables only (gender, age, weight and height). The Harris-Benedict formula estimates BMR, which is between 10 and 20% lower than RMR [5]. Therefore, we chose to increase BMR by 15%, to estimate RMR.

Table 2: Study participants′ characteristics

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

5. DATA COLLECTION AND ANALYSIS

5.1 Participants

Participants were 19 (14 male, 5 female) healthy imec-nl employees from diverse ethnic background - see Table 2. Imec′s internal Ethics Committee approved the study, and each participant signed an informed consent form.

5.2 Instruments

5.2.1 ECG Necklace

The ECG Necklace [17] is a low power wireless ECG platform. The system relies on an ultra-low-power ASIC for ECG read-out, and it is integrated in a necklace, providing ease-of-use and comfort while allowing flexibility in lead positioning and system functionality. It achieves up to 6 days autonomy on a 175 mAh Li-ion battery. For the current study, the ECG Necklace was configured to acquire one lead ECG data at 256 Hz, and ACC data from a three-axial accelerometer (ADXL330) at 64 Hz. The sensor was placed on the chest with an elastic belt. The x, y, and z axes of the accelerometer were oriented along the vertical, mediolateral, and antero-posterior directions of the body, respectively. Two gel electrodes were placed on the participant′s chest, in the lead II configuration. Data were recorded on the on-board SD card to ensure integrity. Data were also streamed in real-time to provide visual feedback of the system functionality to the experimenter.

5.2.2 Indirect Calorimeter

Breath-by-breath data were collected using the *Cosmed* $K4b^2$ indirect calorimeter. The Cosmed $K4b^2$ weights $1.5kg$, battery included, and showed to be a reliable measure of EE [16]. The system was manually calibrated before each experiment according to the manufacturer instructions. This process consists of allowing the system to warm-up, following a double calibration, first with ambient air and then with calibration gas values. A delay calibration was performed weekly to adjust for the lag time that occurs between the expiratory flow measurement and the gas analyzers.

5.3 Experiment Design

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

5.4 Study Design Choices

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

5.5 Pre-processing

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

5.5.1 ECG Necklace Data

Raw ECG and ACC data were downloaded from the SD card of the ECG Necklace using proprietary software developed by imec-nl. Raw data were exported into *csv* files containing time-stamped ECG and acceleration samples. A *Continuous Wavelet Transform* based beat detection algorithm was used to extract R-R intervals from ECG data, which output was examined to correct for missed beats.

5.5.2 Indirect Calorimeter Data

Breath-by-breath data acquired from the Comsed $K4b^2$ was resampled at $0.5\,Hz$. EE was calculated from O_2 consumption and CO_2 production using Weir′s equation [24]. The first 1 or 2 minutes of each activity were discarded to remove non-steady-state data.

5.6 Feature Extraction

Features extracted from the ECG necklace raw data were used to derive activity recognition and EE models. Activity recognition was performed on the six activity clusters introduced in Sec. 4.1. An activity-specific EE model was derived for each cluster. ACC data over the three axes were segmented in 4 second windows, band-pass (BP) filtered between 0.1 and 10 Hz, to isolate the dynamic component caused by body motion, and low-pass (LP) filtered at $1\,Hz$, to isolate the static component, due to gravity. Time and frequency features were extracted from each window over the three axes of the LP and BP signal. Time features included *mean, mean of the absolute signal, magnitude, mean distance between axes, skewness, kurtosis, variance, standard deviation, coefficient of variation, range, min, max, correlation, inter-quartiles range, median* and *zero crossing rate.* Frequency features included: *spectral energy, entropy, low frequency band signal power (*$0.1 - 0.75\,Hz$*), high frequency band signal power (*$0.75 - 10\,Hz$*), frequency and amplitude of the FFT coefficients.* These features were selected due to high accuracy showed in past research [4, 7, 21].

Three features were extracted from R-R intervals, computed over 15 seconds windows; *mean, variance and standard deviation.* Additionally, sleep HR was derived from the HR while lying down [8], and used to extract the HRaR. R-R intervals features were not included in the activity recognition model. Feature extraction was performed in MATLAB (MathWorks, Natick, MA).

5.7 Feature Selection

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

The final feature set was manually selected, taking into account the output from the automatic feature selection scheme when features showed high correlation. It includes: *mean of the absolute band-passed signal and inter-quartile range* — which capture the intensity of whole body motion, *mean distance between axes and median* — which capture posture information, *variance and standard deviation* — measures of the spread of the distribution, *zero crossing rate and main frequency peak* — which provide useful information on the repetitive pattern of certain activities, *low and high frequency band signal power.* We manually selected features for the EE models, according to the criteria illustrated in Sec. 4.1.

5.8 Statistics and Performance Measures

5.8.1 Activity Recognition

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

5.8.2 Energy Expenditure

Performance of the EE models were evaluated in a sub-

Table 3: Classification performance of the C4.5 classifier used to select the cluster model to predict EE.

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

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

5.8.3 Comparisons

Reported performance of EE models are highly dependent on the protocol used to validate the algorithms, which makes it impossible to compare different models from published results. We re-implemented six methods; two simple methods used in epidemiological studies, using ACC (*method ACC* [13]) or HR (*method HR* [10]) as independent variable of the regression model, and four activity-specific (AS) EE algorithms. The four models derive EE assigning static values to the detected activity (*method AS-static* [7, 3]), using a single linear regression model per activity and a measure of Motion Intensity (MI) as the independent variable (*method AS-MI* [21, 23]), combining ACC and HR features following automatic variables selection (*method AS-mixed* [19] - where HR is always used, and accelerometer features are part of one model only) or following the proposed methodology (*method AS-new*). To the best of our knowledge, this is the first comparison of state of the art activity specific models on the same dataset.

Table 4: Predictors and models used to estimate EE for each activities cluster. BW is body weight, MI is motion intensity, VAR is variance, STD is standard deviation, IQR is inter-quartile range.

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

6. RESULTS

6.1 Activity Clusters Classification

Subject independent classification accuracy of the classification tree used to select the cluster model to apply to estimate EE was 92.9%. Table 3 shows the performance of the classifier in terms of sensitivity and specificity for the six clusters. RMSE for walking, running and biking speed were 0.31 km/h, 0.77 km/h and 8.43 rpm. Biking speed errors can be reduced increasing the frequency resolution (i.e. using windows > 4 seconds). Utilizing 4 seconds window our system cannot detect speeds other than multiples of 0.25 Hz.

6.2 Activity Clusters Models

We derived six models (see Table 4), applying the proposed methodology. The total RMSE over the whole protocol, assuming a perfect classification of the activities, was 0.86 $kcal/min$. RMSE for lying, LWBM, HWBM, walking, running and biking were $0.24, 0.42, 0.63, 1, 27, 1.06$ and 1.29 $kcal/mim$ respectively. Misclassification lowers performance to $RMSE = 0.87\ kcal/min$. RMSE for the single clusters after classification were $0.24, 0.42, 0.61, 1, 27, 1.07$ and 1.44 $kcal/mim$. These results confirm that the classifier can be used to select activity cluster models.

6.3 EE Estimation Performance

Table 5 shows results in terms of RMSE averaged over all of the activities and per cluster. Simple methods used in epidemiological studies (methods ACC and HR) show the lowest performance and will not be further discussed.

Results of the *AS-static* method showed improvements compared to non-activity-specific models, but higher error compared to other activity-specific models, in all of the clusters. Recognizing an activity and assigning a static EE value works well on average but cannot capture the variability in EE within the cluster. Measures of motion intensity (*AS-MI*) seem to outperform HR for low to medium intensity activities (LWBM and HWBM), while activities where whole body motion is not representative of EE, such as biking (see Fig. 6), were better modeled by methods using HR as well (*AS-mixed* and *AS-new*).Walking patters were predicted accurately by methods using ACC only features (*AS-MI*) when differences in EE could be explained by motion patterns alone. The inability of these methods to detect the higher energy cost of carrying weights or walking uphill results in decrease of performance during these activities (see Fig. 5).

Overall, combining manually selected ACC and HR features, representative of variations in EE within a cluster, shows significant improvements compared to other methods. Estimates of compendium-based models (*AS-static*) were improved by 31 %. Regression based models that use a measure of motion intensity (*AS-MI*) or automatically selected variables (*AS-mixed*) as predictors, were improved by 18 and 19% respectively. Figures 2 to 6 show how combining features specifically selected for a cluster, based on motion patterns involved in the cluster, as well as physiological signals able to capture variations in EE when motion is constant, provides better estimates compared to other activity-specific methods, on almost all of the activities included in our protocol.

Table 5: Overall and per cluster performance (RMSE) of the methods implemented. Results are in kcal/min. AS is Activity Specific, MI is Motion Intensity. Refer to Sec. 5.8.3 for details on the methods.

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

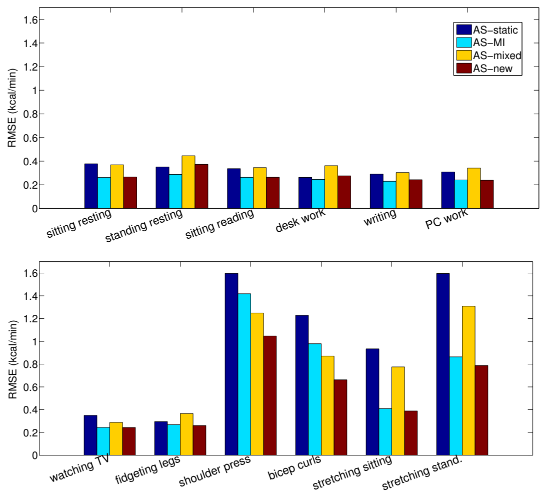

Figure 2: Comparisons of AS methods for the activities included in the *LWBM* cluster.

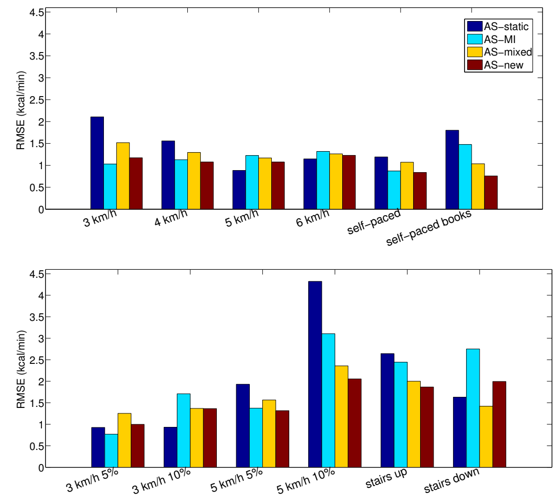

Figure 5: Comparisons of AS methods for the activities included in the *walking* cluster.

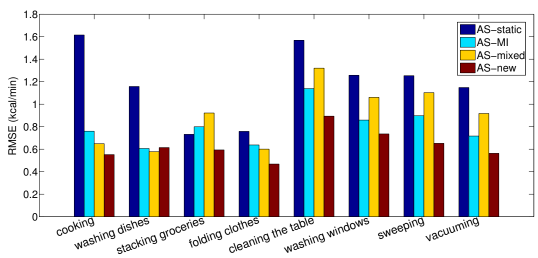

Figure 3: Comparisons of AS methods for the activities included in the *HWBM* cluster.

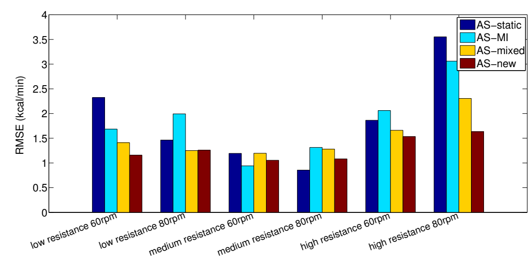

Figure 6: Comparisons of AS methods for the activities included in the *biking* cluster.

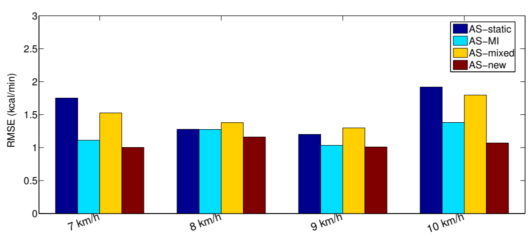

Figure 4: Comparisons of AS methods for the activities included in the *running* cluster.

7. CONCLUSIONS AND FUTURE WORK

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

An aspect of interest that was not further investigated during this study is the personalization of EE models that use physiological signals. Physiological signals (e.g. HR) differ greatly at the individual level, and require either individual calibration or normalization. We used the heart rate above rest as the only heart rate feature, to reduce between-subject differences in HR during different activities. We are currently investigating the possibility to include other factors able to explain between-subject differences in HR during different activities (e.g. cardiorespiratory fitness level), in order to further improve the activity-specific models.

8. ACKNOWLEDGMENTS

The authors would like to thank all of the participants for dedicating their time and effort to take part in this study.

9. REFERENCES

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

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

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

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

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

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

[7] A. G. Bonomi, G. Plasqui, A. H. C. Goris, and K. R. Westerterp. Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. *Journal of Applied Physiology*, 107(3):655–661, 2009.

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

[9] S. Brage, U. Ekelund, N. Brage, M. A. Hennings, K. Froberg, P. W. Franks, and N. J. Wareham. Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *J Appl Physiol*, 2007.

[10] S. M. Ceesay, A. M. Prentice, and K. C. Day. The use of heart rate monitoring in the estimation of energy expenditure: a validation study using indirect whole-body calorimetry. *The British journal of nutrition*, 61(2):175–186, Mar. 1989.

[11] L. Cordain, A. Gotshall, and S. Eaton. Evolutionary aspects of exercise. *World Rev Nutr Diet*, 81:49–60, 1997.

[12] S. Crouter, J. Churilla, and D. Bassett. Estimating energy expenditure using accelerometers. *European Journal of Applied Physiology*, 98(6):601–612, Dec. 2006.

[13] S. E. Crouter and D. R. Bassett. A refined 2-regression model for the actigraph accelerometer. *Medicine & Science in Sports & Exercise*, 42(5):1029–1037, 2010.

[14] P. Freedson. Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: Validation on an independent sample. *Journal of Applied Physiology*, 2011.

[15] S. J. Marshall and E. Ramirez. Reducing Sedentary Behavior : A New Paradigm in Physical Activity Promotion. *American Journal of Lifestyle Medicine*, 2011.

[16] J. E. McLaughlin, G. A. King, E. T. Howley, D. R. Bassett, and B. E. Ainsworth. Validation of the COSMED K4 b2 portable metabolic system. *International journal of sports medicine*, 22(4):280–284, May 2001.

[17] J. Penders. A low-power wireless ECG necklace for reliable cardiac activity monitoring on-the-move. In *EMBC '11*, 2011.

[18] M. P. Rothney, M. Neumann, A. Beziat, and K. Y. Chen. An artificial neural network model of energy expenditure using nonintegrated acceleration signals. *J Appl Physiol*, 103(4):1419–1427, Oct. 2007.

[19] M. Rumo, O. Amft, G. Tröster, and U. Mäder. A stepwise validation of a wearable system for estimating energy expenditure in field-based research. *Physiological measurement*, 32(12):1983–2001, 2011.

[20] D. K. Spierer, M. Hagins, A. Rundle, and E. Pappas. A comparison of energy expenditure estimates from the Actiheart and Actical physical activity monitors during low intensity activities, walking, and jogging. *European journal of applied physiology*, 111(4):659–67, Apr. 2011.

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

[22] V. T. van Hees and U. Ekelund. Novel daily energy expenditure estimation by using objective activity type classification : where do we go from here ? *Journal of Applied Physiology*, pages 639–640, 2009.

[23] V. T. van Hees, R. C. van Lummel, and K. R. Westerterp. Estimating Activity-related Energy Expenditure Under Sedentary Conditions Using a Tri-axial Seismic Accelerometer. *Obesity*, 17(6):1287–1292, Mar. 2009.

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

[25] G. J. Welk, J. McClain, and B. E. Ainsworth. Protocols for evaluating equivalency of accelerometry-based activity monitors. *Medicine and science in sports and exercise*, 44(1):S39–49, 2012.

[26] S. H. Wild, G. Roglic, A. Green, R. Sicree, and H. King. Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030. *Diabetes Care*, 27(10):2569, 2004.