Estimating Energy Expenditure Using Body-Worn Accelerometers: A Comparison of Methods, Sensors Number and Positioning

Marco Altini, Julien Penders, Ruud Vullers, and Oliver Amft

Abstract—**Several methods to estimate energy expenditure (EE) using body-worn sensors exist; however, quantifications of the differences in estimation error are missing. In this paper, we compare three prevalent EE estimation methods and five body locations to provide a basis for selecting among methods, sensors number, and positioning. We considered 1) counts-based estimation methods, 2) activity-specific estimation methods using METs lookup, and 3) activity-specific estimation methods using accelerometer features. The latter two estimation methods utilize subsequent activity classification and EE estimation steps. Furthermore, we analyzed accelerometer sensors number and on-body positioning to derive optimal EE estimation results during various daily activities. To evaluate our approach, we implemented a study with 15 participants that wore five accelerometer sensors while performing a wide range of sedentary, household, lifestyle, and gym activities at different intensities. Indirect calorimetry was used in parallel to obtain EE reference data. Results show that activity-specific estimation methods using accelerometer features can outperform counts-based methods by 88% and activity-specific methods using METs lookup for active clusters by 23%. No differences were found between activity-specific methods using METs lookup and using accelerometer features for sedentary clusters. For activity-specific estimation methods using accelerometer features, differences in EE estimation error between the best combinations of each number of sensors (1 to 5), analyzed with repeated measures ANOVA, were not significant. Thus, we conclude that choosing the best performing single sensor does not reduce EE estimation accuracy compared to a five sensors system and can reliably be used. However, EE estimation errors can increase up to 80% if a nonoptimal sensor location is chosen.**

Index Terms—**Accelerometers, energy expenditure (EE), physical activity (PA), wearable sensors.**

I. Introduction

PHYSICAL activity (PA) and exercise capacity are among the most important determinants of health and wellbeing. Ubiquitous sensing technologies, able to monitor objectively and noninvasively human behavior, started providing unprecedented insights into the relation between PA and health [1].

Energy expenditure (EE) is the most commonly used single metric to quantify PA. Different methods to estimate EE have been developed in the past, from counts-based estimation methods to activity-specific EE equations, developed using one or more accelerometers. Counts-based estimation methods are developed by fitting a single regression line to all the data, independently of the activity performed. On the other hand, in activity-specific estimation methods, the estimation process is split into two steps. First, activities are classified into clusters that group them according to a certain criteria (e.g., EE level [2], motion patterns [3], etc.). Second, an activity-specific model is applied to estimate EE. Activity-specific EE models [3]–[5] showed higher performance compared to single models [6], [7]. However, little agreement is found in the literature regarding number of accelerometers, location on the body, and the role of accelerometer features [e.g., used for activity recognition only (activity-specific models using METs lookup), or for both activity recognition and activity-specific EE models (activity-specific using accelerometer features)] [5], [8]. Even though the use of a single sensor is more practical, recent advances in sensor technology and the ease of integrating small accelerometers into shoes [9], watches or mobile phones, reduced obtrusiveness of wearable sensors, allowing researchers to deploy multisensor systems.

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

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

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

Manuscript received July 26, 2013; revised January 30, 2014; accepted March 17, 2014. Date of publication March 20, 2014; date of current version December 30, 2014.

M. Altini is with the Eindhoven University of Technology, 5612 Eindhoven, The Netherlands (e-mail: marco.altini@imec-nl.nl).

J. Penders and R. Vullers are with Holst Centre/IMEC, 5656 AE Eindhoven, The Netherlands (e-mail: julien.penders@imec-nl.nl; Ruud.Vullers@imec-nl.nl).

O. Amft is with the University of Passau, 94032 Passau, Germany, and also with the Eindhoven University of Technology, 5612 Eindhoven, The Netherlands (e-mail: amft@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2014.2313039

approaches and activity-specific using METs lookup approaches for active clusters.

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

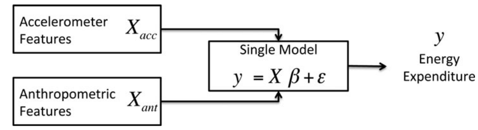

Fig. 1. Block diagram of counts-based estimation methods. Accelerometer and anthropometric features are used independently of the activity type.

II. Related Work

A. Counts-Based Estimation Methods

Counts-based methods were the first EE estimation algorithms developed, given the relation between motion intensity close to the body's center of mass and EE [6]. However, single regression models are unable to fit all the activities, since the slope and intercept of the regression model change based on the activity performed while data are collected [10]. As a result, even when motion intensity (*activity counts*) is representative of EE, the output can be inaccurate. Additionally, the inability of these systems to recognize high or low body movement (e.g., biking or arm exercises) caused high estimation error for activities not involving whole body motion. In [11], the authors had to remove biking activities from their evaluation, due to the inability of their system to capture EE changes when there is limited motion close to the body's center of mass.

B. Activity-Specific Estimation Methods

The latest algorithms extended estimation methods based on single models by performing activity recognition over a predefined set of activities—or clusters of activities, and then applying different methods to predict EE [3]–[5], [8], [10], based on the activity detected (activity-specific EE approaches). Other machine-learning-based methods were developed [7], trying to directly estimate EE from accelerometer features, using for example neural networks [7], [12]. However, these approaches suffer from the same limitations of the counts-based estimation methods, being unable to capture the peculiarities of the relation between accelerometer features and EE during different activities [13]. The most common activity-specific approaches are the following:

1) Activity-Specific Using METs Lookup: One approach is to assign static MET values from the compendium on physical activities [14] to each one of the clusters of activities [3], [8], and use anthropometric features or other static features (e.g., heart rate at rest) to personalize the activity-specific models for different individuals.

2) Activity-Specific Using Accelerometer Features: Another approach is to apply a regression equation for each activity classified [5], [10], extending counts-based approaches to multiple clusters of activities. The regression models typically use accelerometer features and anthropometric characteristics as independent variables.

C. Comparisons

1) Comparisons of Estimation Methods: Altini *et al.* [3] showed that activity-specific estimation methods using METs lookup outperform counts-based approaches when a single sensor is used. In [8], the authors extended the static approach of [3], developing a custom MET table, which takes into account the heart rate at rest, to predict EE, showing 15% improvement in performance compared to the best counts-based estimation methods. In both our previous work [4] and in [8], activity-specific estimation methods using METs lookup and accelerometer features were implemented and compared. However, while we proposed a combined approach using METs lookup for sedentary clusters of activities and using accelerometer features for active clusters, the authors of [8] opted for using METs lookup only. The two systems used a different sensor setup. One single sensor on the chest was used in [4], while three sensors placed on the upper arm, thigh, and waist were used in [8]. The different activity types, sensors number and positioning might have motivated the different choices made by the authors. Thus, it is unclear which estimation method works best as well as if different estimation methods require different sensors number.

2) Comparisons of Sensors Number and Positioning: When it comes to sensors number and positioning, comparisons are lacking. Some works investigated the accuracy of sensors placed on different parts of the body to detect a specific set of activities [2], [5], [15]–[17]. However, none of these works considered how sensors number and positioning affects EE. Some researchers showed high accuracy in EE estimates adopting one sensor placed on the lower back [3] or chest [4]. Others used two or three accelerometers [5], [8]. Small differences between protocols used to collect data, algorithms evaluation metrics, as well as the inclusion of extra sensors in only some of the systems (e.g., heart rate), limit our understanding of what is the best solution in terms of sensors number and positioning.

III. Analysis Approach

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

1) *Estimation Methods:* We compared three common methods to estimate EE: 1) counts-based, activity-specific using 2) METs lookup and using 3) accelerometer features (see Figs. 1 and 2).

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

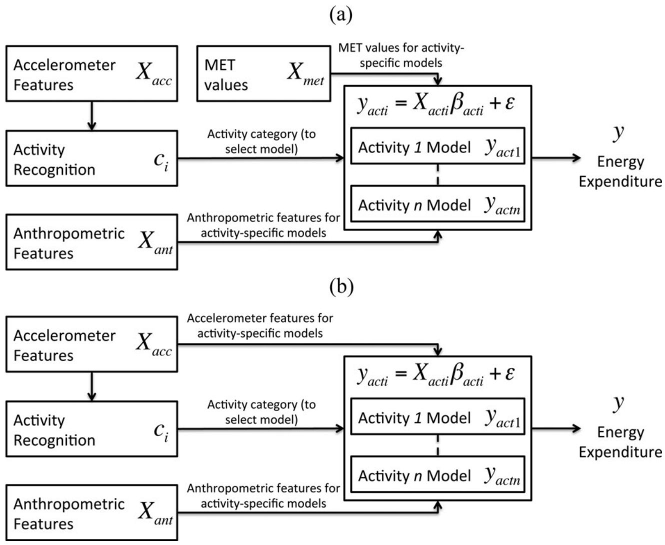

Fig. 2. Block diagram of the activity-specific estimation methods considered for comparison in this work. a) shows approaches using METs lookup while b) shows approaches using accelerometer features as predictors.

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

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

2) *Sensors Number and Positioning:* We evaluated all possible combinations of five sensors (see Section V-B for details). Our analysis is structured as follows:

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

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

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

A. Statistics and Performance Measure

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

IV. Implementation

A. Activity-Type Clusters

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

B. Features Extraction and Selection

Features extracted from the sensors' raw data were used to derive activity recognition and EE models. Accelerometer data from the three axes of all five sensors were segmented in 4 s windows, bandpass (BP) filtered between 0.1 and 10 Hz, to isolate the dynamic component caused by body motion, and low-pass filtered at 1 Hz, to isolate the static component, due to gravity. Feature selection for activity recognition was based on correlation, due to the hypothesis that a good feature set includes features correlated with the class, but uncorrelated to each other. The final feature set included: *mean of the absolute BP signal*, *interquartile range*, *mean distance between axes*, *median*, *variance*, *standard deviation*, *zero crossing rate*, *main frequency peak*, *low and high-frequency band signal power*. Feature selection for EE was based on how much variation in EE each feature could explain within one cluster. The process was automated using linear forward selection. Features to be selected depended on the combination of sensors considered for a model. Additionally, anthropometrics features (*body weight* and *resting metabolic rate (RMR)*, estimated with the Harris–Benedict formula [18]) were added depending on the cluster, following the methodology for activity-specific EE models presented in [4].

C. Activity Recognition

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

D. Energy Expenditure

1) Counts-Based Methods: We implemented single regression models using data from all activities and motion intensity (i.e., *mean of the absolute BP signal summed over the three axis*) as the only accelerometer feature, together with anthropometric characteristics (*body weight* and *RMR*), as typically done in epidemiological studies (see Fig. 1).

2) Activity-Specific Estimation Methods Using METs Lookup: Activity-specific estimation methods using METs lookup relied on the activity recognition system of Section IV-C. METs values were used together with anthropometric features (*body weight* and *RMR*), for the activity-specific linear regression models [see Fig. 2(a)]. METs values were chosen based on compendium values for the activities included in each cluster, resulting in 1 for *lying*, 1.3 for *sitting* and *standing*, 3.5 for *HWBM*, 3 for *walking*, 6.7 for *biking,* and 11 for *running*.

3) Activity-Specific Estimation Methods Using Accelerometer Features: Within one activity cluster, EE can be estimated using other features, representative of EE changes within the activity cluster [4], [5], [10]. Depending on sensors selected, we created different EE activity-specific linear models, using the selected set of features for those sensors [see Fig. 2(b)].

V. Evaluation Study

A. Participants

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

B. Instruments

1) Body Area Network: The sensor platform used was the ECG Necklace. Five ECG Necklaces were synchronized in a wireless network [19]. One ECG Necklace was placed on the chest (C) and configured to acquire one lead ECG data at 256 Hz, and accelerometer data at 64 Hz (ADXL330). Sampling frequency was chosen as 64 Hz since it is considered to be much higher than typical human motion. The other four ECG Necklaces were configured to acquire only accelerometer data at 64 Hz and placed on the dominant ankle (An), dominant thigh (T), dominant wrist (W), and waist (Wa)—at the right hip. All sensors were attached to the body using elastic bands. ECG data were not used for this study. Activity type was annotated manually by experimenter.

2) Indirect Calorimeter: Breath-by-breath data were collected using the Cosmed $K4b^2$ indirect calorimeter. The Cosmed $K4b^2$ weights 1.5 kg and showed to be a reliable measure of EE [20]. The system was manually calibrated before each experiment according to the manufacturer instructions.

C. Experiment Design

Participants were invited for recordings on two separate days. They reported to the lab at 8:00 A.M., after refraining from drinking (except for water), eating and smoking in the 2 h before the experiment. The protocol included a wide range of sedentary, lifestyle, and sport activities. Each activity was carried out for a period from 4 to 12 min, except for running (1 to 4 min). The first minute of each recording was removed to discard nonsteady-state data.

VI. Results

Given the high number of models implemented, we report only results for the best combinations of 1 to 5 sensors [see Figs. 3(a), (b), 4(a), (b), and 5(a)], as well as information on exactly which sensors provide these optimal performance, together with the worst performance obtained with the same number of sensors, for comparison.

A. Estimation Methods

Fig. 3 shows the effect of different feature sets on EE estimation performance for activity-specific EE models, assuming perfect activity recognition. Only one activity-specific model using METs lookup is needed for comparison, since these approaches do not use accelerometer features. The RMSE obtained for activity-specific estimation methods using METs lookup was 1 kcal/min, while for activity-specific estimation methods using accelerometer features it ranged between 0.84 and 0.86 kcal/min ([18% error reduction, $p < 0.05$, Fig. 3(a)]. 23% error reduction was shown for active clusters using accelerometer features. Fig. 4 shows performance of the EE estimation models in combination with activity recognition, as well as counts-based estimation methods. For clarity, results for the activity-specific estimation methods using METs lookup were omitted in Fig. 4. Activity-specific estimation methods using METs lookup rely on the same activity recognition algorithms used by the activity-specific method using accelerometer features, thus the METs-based method would still perform suboptimally. The RMSE for activity-specific estimation methods using accelerometer features ranged from 0.85 to 0.89 kcal/min. RMSE for counts-based estimation methods was between 1.6 and 2.6 kcal/min depending on sensor position (88% error increase for the best-performing sensor, C, $p < 0.05$). The error obtained using the counts-based estimation was significantly higher compared to activity-specific models even when counts were considered separately for sedentary and active clusters [see Fig. 5(b)].

B. Sensors Number and Positioning

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

1) Activity Recognition: Fig. 5(a) shows the performance of the activity recognition models. Additionally, the impact of sensor location (best versus worst for each number of sensors) is shown [see Figs. 5(b)–(e)]. Accuracy varied between 85 for 1 sensor, and 98% for 3 or more sensors [see Fig. 5(a)]. Accuracy for active clusters was always above 98%, with differences of only 1% between the best single sensor system and a five sensors body area network [see Fig. 5(a)]. Sedentary clusters accuracy ranged between 69.9 and 97%. Sensor location affected the accuracy by 12% for a single sensor, while the decrease in performance was reduced to 7%, 5%, and 4% for two, three, and four sensors, respectively [see Figs. 5(b)–(e)].

2) Differences in EE Within an Activity Cluster: Fig. 3 shows the effect of different feature sets on EE estimation performance for activity-specific estimation methods using accelerometer features, assuming perfect activity recognition. No significant differences were found when different locations on the body were considered to extract activity-specific features. However, differences are found when analyzing separately sedentary and active clusters, showing higher errors in sedentary clusters using accelerometer features from four or five sensors [see Fig. 3(b)].

3) EE Estimation: Fig. 4 shows performance of the EE estimation models in combination with the activity recognition. In this analysis, differences in performance are due to 1) higher misclassification rates of models based on a smaller number of sensors and 2) different feature sets used for activity-specific estimation methods using accelerometer features, depending on the sensors that are part of the system. Sensor location analysis shows the Chest sensor as the best single sensor for EE estimation, while the Wrist sensor seems to perform worse than any other combination [see Figs. 5(c)–(f)].

VII. Discussion

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

A. Estimation Methods

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

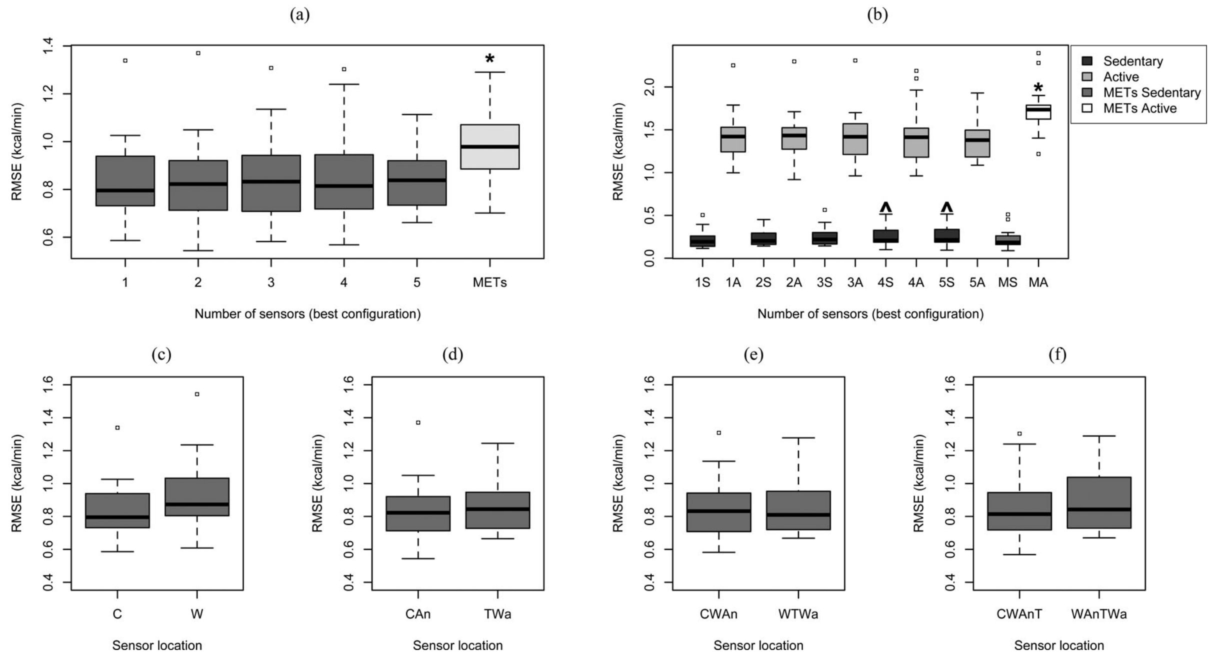

Fig. 3. EE estimation RMSE for sedentary and active clusters when perfect activity recognition is assumed in activity-specific estimation methods. Boxplots 1 to 5 in plot a as well as $1S - 1A$ to $5S - 5A$ on plot b concern activity-specific estimation methods using accelerometer features, while an activity-specific model using METs lookup is shown as *METs* on plot a and $MS - MA$ on plot b. $*$ indicates significant differences between the annotated model (activity-specific model using METs lookup) and all of the other models, i.e., the ones using accelerometer features ($p < 0.05$). Λ indicates significant differences between models the annotated models, i.e., 4S (four sensors, sedentary clusters) and 5S (five sensors, sedentary clusters) and model sedentary model when only one sensor is used, i.e., 1S ($p < 0.05$). RMSE for the best and worst activity-specific model using accelerometer features for each number of sensors is shown on the bottom row. C is Chest, T is Thigh, An is Ankle, W is Wrist, and Wa is Waist. (a) All clusters-best. (b) Sedentary and Active clusters-best. (c) 1 sensor-best VS worst. (d) 2 sensors-best VS worst. (e) 3 sensors-best VS worst. (f) 4 sensors-best VS worst.

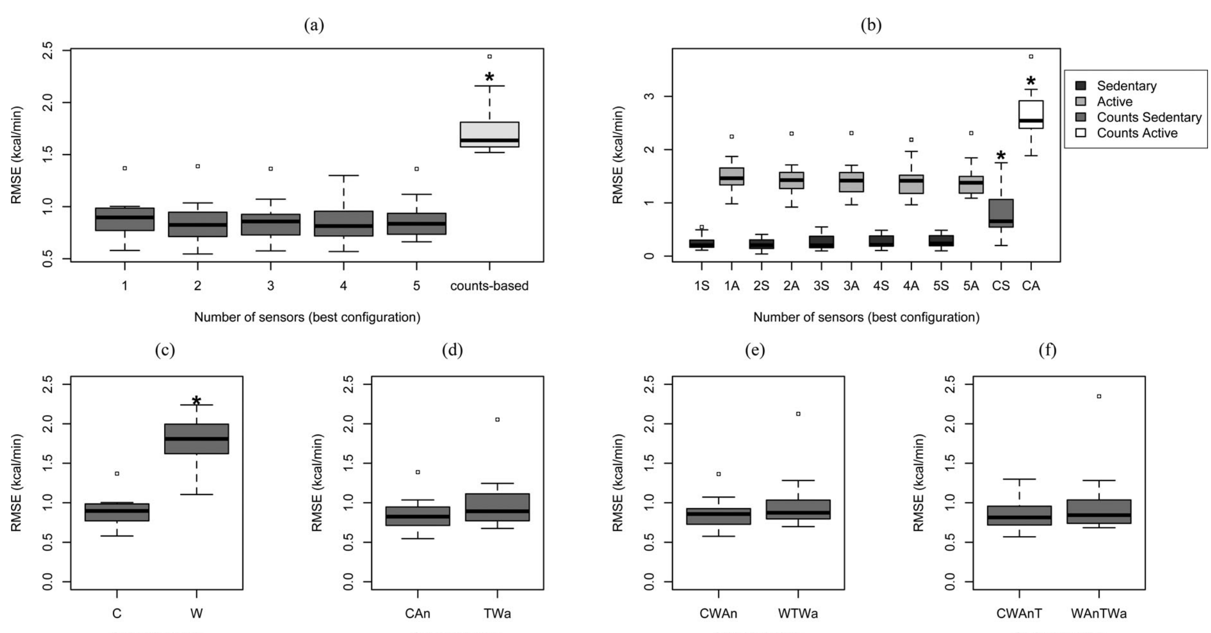

Fig. 4. EE estimation RMSE for sedentary and active clusters in activity-specific estimation methods using accelerometer features, after activity classification. Activity-specific estimation methods using METs lookup are not shown due to subperforming results. Comparison with a counts-based model is shown in a) as *counts-based* and b) as *CS* and *CA*. $*$ indicates significant differences between the annotated counts-based model and all of the other models, i.e., activity specific models using accelerometer features ($p < 0.05$). RMSE for the best and worst activity-specific models using accelerometer features for each number of sensors is shown on the bottom row. C is Chest, T is Thigh, An is Ankle, W is Wrist, and Wa is Waist. (a) All clusters-best. (b) Sedentary and Active clusters-best. (c) 1 sensor-best VS worst. (d) 2 sensors-best VS worst. (e) 3 sensors-best VS worst. (f) 4 sensors-best VS worst.

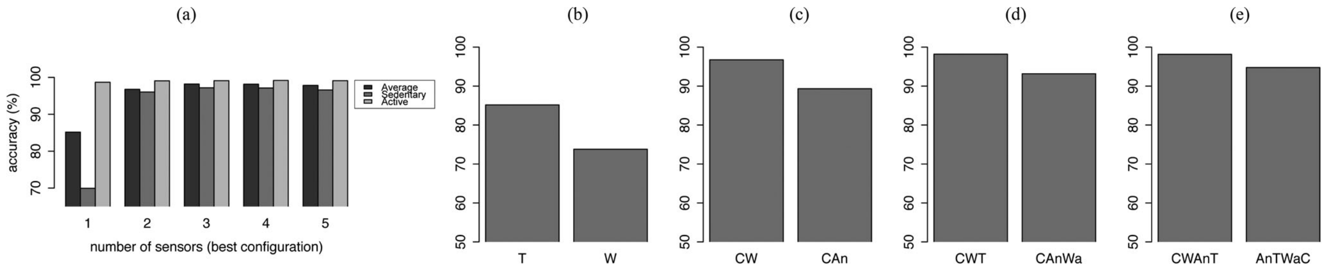

Fig. 5. Activity recognition accuracy for sedentary (average accuracy of *lying, sitting,* and *standing*) and active (average accuracy of *HWBM, walking, biking,* and *running*) clusters, and their average. Classification accuracy for different sensors number and positioning are shown on the right. C is Chest, T is Thigh, An is Ankle, W is Wrist, and Wa is Waist. (a) Sedentary and Active clusters-best. (b) 1 sensor-best VS worst. (c) 2 sensors-best VS worst. (d) 3 sensors-best VS worst. (e) 4 sensors-best VS worst.

B. Sensors Number and Positioning

Our results on the sensors number and positioning point out three main findings: 1) On activity recognition: if properly chosen, two sensors are sufficient to provide accurate PA type assessment (see Fig. 5). 2) On differences in EE within an activity cluster: Adding features from more than one sensor in the activity-specific models using accelerometer features does not improve the accuracy of the EE estimate (see Fig. 4). 3) On EE estimation: Applying a wrong EE model due to misclassification of the activity type has a small (nonstatistically significant) impact on the EE estimate accuracy provided that an optimal sensor positioning is chosen (e.g., the Chest sensor, see Fig. 4). Thus, choosing the best performing single sensor does not reduce performance for EE estimation compared to a five sensors system.

1) Activity Recognition: Our results on the sensor number for activity recognition confirm previous works that considered multiple accelerometers [5], [15]–[17]. Adding more sensors improves accuracy, until a plateau is reached, when two or more sensors are used, in our case, 97/98% accuracy using Chest and Wrist or Chest and Thigh sensors. It is of interest for our analysis, how activity recognition influences EE estimates as discussed next.

2) Differences in EE Within an Activity Cluster: Our second finding concerns the accelerometer features needed to explain differences in EE within one cluster. To determine such features, we developed EE models assuming perfect activity recognition (see Fig. 3). We showed that accelerometer features from one sensor are sufficient to explain differences in EE within one cluster of activities. This finding can be explained by the fact that within one cluster of activities (for example *walking*) the variation in EE is explained mainly by the level of motion intensity of the *whole body*. Other features, such as *motion intensity of the wrist sensor*, can lead to errors, since high level of motion (e.g., while writing), do not correspond to high EE. This reasoning might explain why in Fig. 3 the error is shown to increase when features from 4 or 5 sensors are used for sedentary clusters [see Fig. 4(b)].

Even though adding features from more sensors does not reduce EE estimate error, accelerometer features from at least one sensor should be used for active clusters (23% error reduction compared approaches using METs lookup). In a recent review on activity-specific EE estimation [21], the controversy between applying static values (i.e., MET values) and the need of including accelerometer features in linear models had been raised. Past research showed inconsistency in the approach used for activity-specific models even after implementing and comparing estimation methods using METs lookup or accelerometer features [4], [8]. With this analysis, we show that accelerometer features are relevant only for active clusters, and most importantly this is true regardless of the number of sensors used [see Fig. 3(b)]. Our findings are consistent with our previous work using one sensor [4], indicating that the best approach to obtain high accuracy and limit model complexity, is to use a combined approach. Activity-specific models using METs lookup can be used for sedentary activities, where static METs values and anthropometric features are sufficient to accurately estimate EE.

3) EE Estimation: Provided that the best performing sensor is chosen, no significant error reduction was found when more than one sensor was used for EE estimation. This is due to the fact that errors are mainly due to misclassification of posture (one single sensor is unable to recognize all of the three postures in the sedentary cluster), resulting in applying a very similar activity-specific EE model. Thus, the EE estimation RMSE for a single sensor placed on the Chest is similar when compared to a five sensors system (no statistically significant difference), even if activity classification accuracy is decreased by up to 13% on average, and 28% for sedentary clusters. This is an important finding since past work showed good accuracy using one single accelerometer and activity-specific approaches [3], [4], but no previous work could compare performance of EE estimation methods when different sensors number and positioning were used, preventing us from understanding if systems relying on multiple sensors for activity recognition [5], [8] could still provide better results.

C. Limitations

Performance for activities that were not part of the dataset should be assessed outside of the lab. However, there is currently no reference system able to measure breath-by-breath EE in unconstrained settings. Only by using indirect calorimetry and supervised settings, we can record data which allow us to analyze how multiple sensors affect the EE estimate process in both activity recognition and intraindividual differences within one activity. Another limitation was to limit the number of MET values used for our analysis to the ones associated to the activity clusters, while more fine grained values could be used for certain activities (e.g., walking at different speeds). However, we believe that using individual MET values for activities may not generalize, since some activities (e.g., related to household) show different EE but cannot be accurately subdivided when using a limited number of sensors. Hence, some activity clusters would still require a single MET value to be used, while actual EE varies widely.

VIII. Conclusion

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

References

[1] N. Butte, U. Ekelund, and W. KR, "Assessing physical activity using wearable monitors: Measures of physical activity," *Med. Sci. Sports Exercise*, vol. 44, no. 1, Suppl. 1, pp. S5–S12, Jan. 2012.

[2] L. Atallah, B. Lo, R. King, and G.-Z. Yang, "Sensor positioning for activity recognition using wearable accelerometers," *IEEE Trans. Biomed. Circuits Syst.*, vol. 5, no. 4, pp. 320–329, Aug. 2011.

[3] A. G. Bonomi, "Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer," *J. Appl. Physiol.*, vol. 107, no. 3, pp. 655–661, 2009.

[4] M. Altini, J. Penders, and O. Amft, "Energy expenditure estimation using wearable sensors: A new methodology for activity-specific models," in *Proc. Wireless Health Conf.*, 2012.

[5] E. Tapia, "Using machine learning for real-time activity recognition and estimation of energy expenditure" Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2008.

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

[7] P. Freedson, "Evaluation of artificial neural network algorithms for predicting METs and activity type from accelerometer data: Validation on an independent sample," *J. Appl. Physiol.*, vol. 111, pp. 1804–1812, 2011.

[8] F. Albinali, S. S. Intille, W. Haskell, and M. Rosenberger, "Using wearable activity type detection to improve physical activity energy expenditure estimation," in *Proc. 12th ACM Int. Conf. Ubiquitous Comput.*, 2010, pp. 311–320.

[9] S. J. Morris and J. A. Paradiso, "Shoe-integrated sensor system for wireless gait analysis and real-time feedback," in *Proc. 2nd Joint EMBS/BMES Conf.*, 2002, vol. 3, pp. 2468–2469.

[10] 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*, vol. 17, no. 6, pp. 1287–1292, 2009.

[11] S. Crouter, J. Churilla, and D. Bassett, "Estimating energy expenditure using accelerometers," *Eur. J. Appl. Physiol.*, vol. 98, no. 6, pp. 601–612, Dec. 2006.

[12] 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.*, vol. 103, no. 4, pp. 1419–1427, Oct. 2007.

[13] A. G. Bonomi and G. Plasqui, ""Divide and conquer": Assessing energy expenditure following physical activity type classification," *J. Appl. Physiol.*, vol. 112, no. 5, p. 932, Mar. 2012.

[14] B. E. E. A. Ainsworth, "Compendium of physical activities: An update of activity codes and met intensities," *Med. Sci. Sports Exercise*, vol. 32, no. 9, pp. S498–504, 2000.

[15] L. Bao and S. Intille, "Activity recognition from user-annotated acceleration data," in *Pervasive Computing*. Berlin, Heidelberg: Springer, 2004, vol. 3001, pp. 1–17.

[16] I. Cleland, B. Kikhia, C. Nugent, A. Boytsov, J. Hallberg, K. Synnes, S. McClean, and D. Finlay, "Optimal placement of accelerometers for the detection of everyday activities," *Sensors*, vol. 13, no. 7, pp. 9183–9200, 2013.

[17] D. O. Olguın and A. S. Pentland, "Human activity recognition: Accuracy across common locations for wearable sensors," 2006.

[18] D. Frankenfield, L. Roth-Yousey, and C. Compher, "Comparison of predictive equations for resting metabolic rate in healthy nonobese and obese adults: A systematic review, vol. 105, pp. 775–789, May 2005.

[19] M. Altini, S. Del Din, S. Patel, S. Schachter, J. Penders, and P. Bonato, "A low-power multi-modal body sensor network with application to epileptic seizure monitoring," in *Proc. IEEE Eng. Med. Biol. Soc.*, 2011, pp. 1806–1809.

[20] J. E. McLaughlin, "Validation of the COSMED K4 b2 portable metabolic system," *Int. J. Sports Med.*, vol. 22, no. 4, pp. 280–284, 2001.

[21] V. T. van Hees and U. Ekelund, "Novel daily energy expenditure estimation by using objective activity type classification: Where do we go from here?," *J. Appl. Physiol.*, vol. 107, pp. 639–640, 2009.

Marco Altini received the M.Sc. degree in engineering and computer science from the University of Bologna, Bologna, Italy, in 2010. He is currently working toward the Ph.D. degree with the Technical University of Eindhoven, Eindhoven, The Netherlands.

His current research interests include machine-learning techniques for the development and implementation of activity recognition, energy expenditure, and cardiorespiratory fitness estimation algorithms based on acceleration and physiological signals acquired using wearable sensors.

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, USA, in 2004 and 2006, respectively.

He is currently the Program Manager with the Holst Centre/IMEC, Eindhoven, The Netherlands. He has authored or coauthored more than 30 papers in journals and conference proceedings on body area networks and autonomous wireless sensor networks, and is the author of two book chapters.

Ruud Vullers received the M.Sc. degree in physics at the Radboud University of Nijmegen, Nijmegen, The Netherlands, and the Ph.D. degree at the K.U. Leuven, Leuven, Belgium, in 1994 and 2000, respectively.

He then started as a Researcher for Philips Research Laboratories in the area of Optical Storage. In 2006, he joined IMEC/Holst Centre, Eindhoven, The Netherlands, where he is Principal Researcher and Activity Leader in the micropower program. This program focuses on the development of energy harvesters, design of power management, and evaluation of energy storage systems for use in autonomous wireless sensor nodes.

Oliver Amft received the M.Sc. degree from the Chemnitz University of Technology, Chemnitz, Germany, in 1999, and the Ph.D. degree from ETH Zurich, Zurich, Switzerland, in 2008, both in electrical engineering and information technology.

He is a Full Professor (W3) and Chair of Sensor Technology at University of Passau, Passau, Germany. He is also with the Wearable Computing Lab, ETH Zurich and the Signal Processing Systems section at TU Eindhoven, Eindhoven, Netherlands. His research interests include multimodal activity recognition and human behavior inference algorithms and has coauthored more than 90 publications in this field.