Combining Wearable Accelerometer and Physiological Data for Activity and Energy Expenditure Estimation

Marco Altini[1,2], Julien Penders[1], Ruud Vullers[1]
[1]Holst Centre - imec The Netherlands
Eindhoven, The Netherlands
altini.marco@gmail.com

Oliver Amft[2]
[2]ACTLab, Signal Processing Systems
Eindhoven University of Technology
Eindhoven, The Netherlands

ABSTRACT

Physical Activity (PA) is one of the most important determinants of health. Wearable sensors have great potential for accurate assessment of PA (activity type and Energy Expenditure (EE)) in daily life. In this paper we investigate the benefit of multiple physiological signals (Heart Rate (HR), respiration rate, Galvanic Skin Response (GSR), skin humidity) as well as accelerometer (ACC) data from two locations (wrist - combining ACC, GSR and skin humidity - and chest - combining ACC and HR) on PA type and EE estimation. We implemented single regression, activity recognition and activity-specific EE models on data collected from 16 subjects, while performing a set of PAs, grouped into six clusters (*lying, sedentary, dynamic, walking, biking* and *running*). Our results show that combining ACC and physiological signals improves performance for activity recognition (by 2 and 8% for the chest and wrist) and EE (by 36 - chest - and 35% - wrist - for single regression models, and by 18 - chest - and 46% - wrist - for activity-specific models). Physiological signals other than HR showed a coarser relation with level of physical exertion, resulting in being better predictors of activity cluster type and separation between inactivity and activity than EE, due to the weak correlation to EE within an activity cluster.

Categories and Subject Descriptors

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

General Terms

Algorithms, Experimentation

1. INTRODUCTION

Accurate monitoring of physical activity (PA) is key in unveiling the relation between aspects of human behavior and health status. Especially in today's industrialized societies, the population's physical activity level is decreasing below the recommended levels. As a result, new epidemics (e.g. obesity, diabetes) are spreading all over the world.

New technologies seamlessly integrated in everyone's life, able to monitor behavior objectively and non-invasively, can provide unprecedented insights on the relations between PA and health. Currently, epidemiologists use accelerometers (ACC) and heart rate (HR) monitors to objectively gather information about PA [11, 8]. For ACC, 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 ACC worn close to the body center of mass to detect low and upper body motion, the low accuracy of HR monitors during sedentary behavior and the need for individual calibration [20, 4].

Recent developments in objective monitoring of PA are moving towards activity-specific EE algorithms [2, 6, 20]. Activity-specific EE algorithms first recognize the activity performed, and then apply a model developed for the specific activity, showing consistent improvements compared to previous methods. Among these activity-specific models, the ones combining HR and ACC consistently showed better results compared to ACC only [2]. However, important limitations on previous studies are impairing our understanding of the role of physiological signals in PA assessment. While the relation between HR and EE has been covered in literature, only recently HR has been introduced in activity-specific models to improve EE estimation accuracy during moderate to intense activities [2, 17]. Other signals (e.g. Galvanic Skin Response (GSR), skin temperature) have been used to develop proprietary models not accessible by the research community, thus limiting our understanding of how these signals can contribute to the estimate [22].

Another aspect that should be taken into account, is the relation between physiological data, ACC data and activity type, since PA monitoring is not anymore limited to the assessment of one single parameter (i.e. EE). The relation between activity type and physiological signals was explored only partially in previous research [20, 15], and should be extended, since the recently widespread use of single sensor devices in locations where movement is weakly related to EE (e.g. the wrist) can cause inaccuracy in PA assessment. Finally, differences found in literature on PA monitoring using wearable sensors, regarding protocols (e.g. activities performed), subjects (e.g. anthropometric characteristics), reference systems (e.g. indirect calorimeters, DLW, etc.) and evaluation metrics (Root Mean Square Error (RMSE), ac-

Wireless Health '13, Nov 1-3, 2013, Baltimore, MD, USA

curacy, etc.), prevent meaningful comparison of previously published results.

In this paper we analyze the benefit of ACC and physiological data measured at different on-body sensor locations on PA type and EE estimation. We collected inertial (ACC), physiological (HR, GSR, respiration, skin humidity) and reference (activity type, EE) data using two wearable sensors, one on the chest (ACC, ECG, HR) and one on the wrist (ACC, GSR, skin humidity) and an indirect calorimeter (Cosmed $K4B^2$ - VO_2, O_2 and respiration rate), while a total of 16 subjects performed a set of physical activities, clustered in groups representative of human behavior in daily life. More specifically, our contribution is twofold:

1. We implemented single and activity-specific models combining ACC and physiological data, and selected metrics to evaluate the impact of each signal (or combination of signals) in activity type recognition and EE estimation, analyzed on the same set of subjects and activities. We evaluated the performance regarding activity type recognition accuracy and explained variations in EE within an activity cluster as well as explained variations in EE over all activities.

2. We analyzed trade-offs between on body position, activity recognition and EE estimation accuracy. We show that HR is the best predictor of EE, given the direct relation to oxygen uptake, while other physiological signals (GSR, skin humidity, respiration) have a coarser relation to level of physical exertion, which makes them better predictors of activity type clusters (when combined with ACC), than EE. We provide guidelines for future PA research that makes use of wearable sensors to monitor PA, using wearable sensors which combine ACC and physiological data.

2. RELATED WORK

2.1 Single Models

ACC and HR monitors are the most commonly used single sensor devices in epidemiologic studies. 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 [11]. As a result, even when motion intensity (or *activity counts*) is representative of EE, the output can be misleading. In [10] 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. HR monitors suffer from different problems, the most common being the low accuracy during sedentary behavior [8], given that HR is affected by other factors (e.g. stress), and the need for individual calibration [4].

2.2 Machine Learning-Based Models

The latest algorithms go towards three directions, all using pattern recognition and machine learning techniques.

1) Some authors applied machine learning methods to directly estimate EE from ACC features, using for example neural networks [16, 12]. However these approaches suffer from the same limitations of the counts-based approaches, being unable to capture the peculiarities of the relation between ACC features and EE during different activities [7].

2) Others, extended the single model approach, performing activity recognition over a pre-defined set of activities, and then applying different methods to predict EE [6, 2, 21, 17, 20] (activity-specific models).

3) Finally, unsupervised approaches were introduced by the authors in [9]. Unsupervised clustering was used to avoid exact activity detection and the need for time consuming activity labeling during data collection, still providing the advantage of dividing the EE estimation problem into subproblems, as done by activity-specific models [7]. However, this approach showed sub-optimal performance compared to activity-specific models [9].

Given the significant amount of work adopting activity recognition as a first step to estimate EE (method *2*), and the consistent improvements obtained [7, 9, 2, 1], we believe this is the best methodology to follow when developing EE estimation algorithms. Activity-specific models extended approaches based on single models by performing activity recognition over a predefined set of activities, and then applying different methods to predict EE [2, 20, 17, 6], based on the activity. One approach [20] is to apply a different regression equation for each activity classified. The regression models typically use ACC features and anthropometric characteristics as independent variables. Another approach is to assign static values from the compendium on physical activities to each one of the clusters of activities [6]. Assigning static values showed limitations during moderate to vigorous activities in a recent comparison between activity-specific models, since static values cannot capture differences in EE within one cluster [2]. Intra-individual differences in EE for an activity are caused by the fact that moderate to vigorous activities can be carried out at different intensities (e.g. walking at different speeds), resulting in different levels of EE. Activity-specific linear regression models require ACC and HR features to capture these differences. In [1], a multi-sensor system composed of three ACC was developed. The authors extended the static approach of [6], developing a custom MET table, which takes into account the HR at rest. In [17] HR and ACC were combined as well. The system consisted of three sensors, two ACC and a HR belt, and could classify six types of activities. In [2] we introduced a combined approach using static values for sedentary clusters, and regression equations for moderate to vigorous clusters, reducing EE estimation errors up to 31% compared to other state-of- the-art activity-specific algorithms.

2.3 Comparisons

A recent review reimplemented different ACC and HR based methods [2]. Activity-specific multiple linear regression models combining ACC and HR showed consistent improvements in EE estimation accuracy compared to algorithms using or ACC-only data - during moderate to vigorous physical activities. Other papers compared multi-sensor devices (e.g. the Actiheart which combines ACC and HR or Bodymedia's armband which combines GSR, skin temperature, heat flux, and ACC) during different activities [19]. Results showed that combinations of physiological and ACC data are better predictors of EE than ACC alone [22]. Unfortunately, some of these devices do not provide details on the algorithms, preventing PA researchers from understanding how different signals contribute to the final results.

3. RELATING ACCELEROMETER AND PHYSIOLOGICAL DATA TO PHYSICAL ACTIVITY TYPE AND EE

This section covers the motivations behind the use of different ACC and physiological data in single regression as well as activity-specific EE models. We cover the signals used in our analysis; ACC, HR, respiration rate, GSR and skin humidity. These sensor locations and types were selected based on current availability and user comfort, in order to provide useful insights for research at the algorithmic level. Fig. 1 shows the behavior of each signal during the activities included in our protocol, as well as reference activity type (color coded) and reference EE (in gray).

3.1 Acceleration

ACC have been used since the beginning of PA research. ACC typically use *activity counts*, a unit-less measure representative of whole body motion, as the independent variable to predict EE [11]. Alternatively, ACC have been used in both stages of activity-specific models, to first recognize an activity, and then model inter-individual differences in EE for an activity cluster [2]. Fig. 1 shows the weak relation between activities involving limited motion and EE (e.g. *biking*) or high level of motion and low EE (e.g. *sedentary* or *High Whole Body Motion/Dynamic* activities for ACC at the wrist), as well as the good correlation between activities involving high level of motion and EE (e.g. *walking*).

3.2 Heart Rate

The widespread use of HR monitoring is due to its ease of measurement and its reflection of the relative stress placed on the cardiopulmonary system due to PA. However, HR can also be elevated by emotional stress, which is independent of any change in oxygen uptake [8]. The high correlation between HR and EE for active clusters (e.g. *walking, biking, running*) is shown in Fig. 1. The relation is weaker for inactive clusters (e.g. *lying* and *sedentary*).

3.3 Respiration

The use of respiration rate in PA research is very limited. However, in [18] authors improved the EE estimation accuracy by combining ECG derived respiration rate and HR, compared to HR alone. During certain dynamics, for example when going from moderate to light activities, respiration rate might better represent the return of oxygen uptake to baseline, compared to HR. The relation between respiration rate, EE and activity type is shown in Fig. 1.

3.4 Galvanic Skin Response

Galvanic Skin Response (GSR) measures skin conductivity. Skin conductivity is affected by sweat due to physical exertion as well as emotional stimuli such as psychological stress. GSR is widely used in stress research, but given the strong relation between skin conductance and sweat it has been used as predictor of EE [22]. The coarse relation between GSR, EE and activity type is shown in Fig. 1.

3.5 Skin Humidity

Skin humidity is also affected by sweat, and looking at continuous measurement of skin humidity in conjunction with data collected from other sensors could reveal the body's level of physical exertion (see Fig. 1).

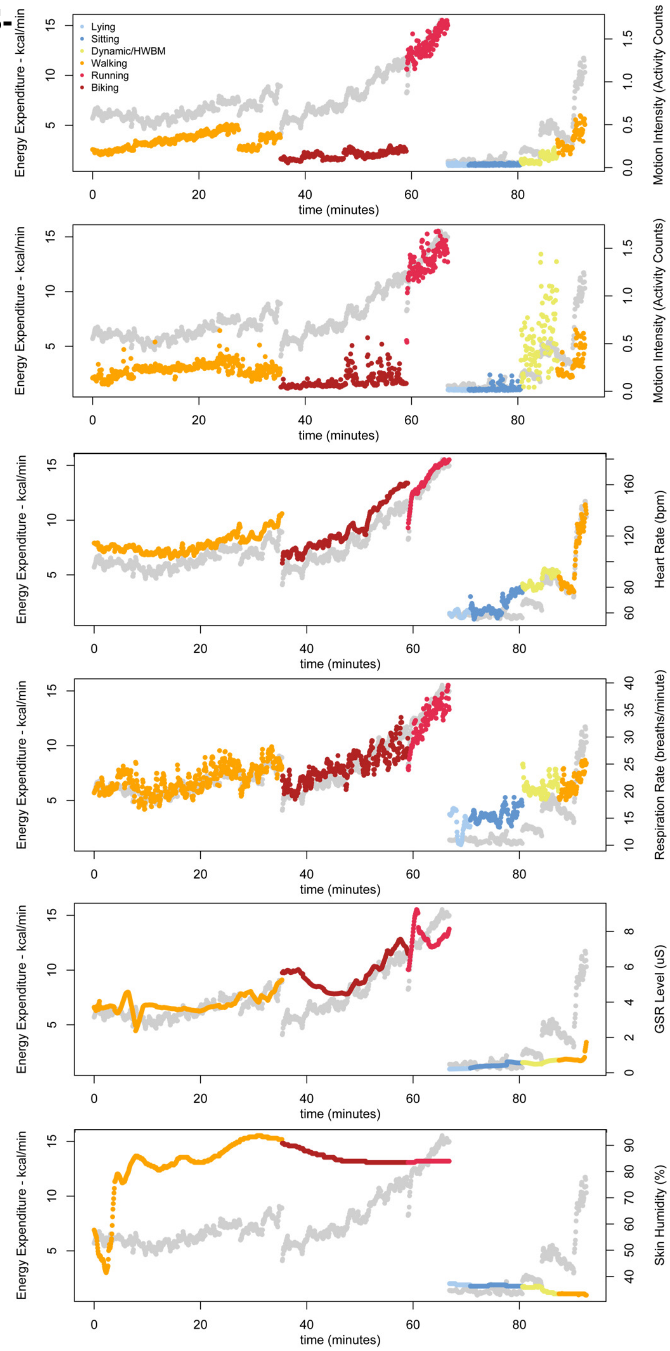

Figure 1: ACC and physiological signals during all the activities of our protocol, for one subject. From top to bottom: motion intensity from the chest, and wrist, HR, respiration rate, GSR level, skin humidity. Reference activity type is color coded, reference EE is shown in gray. Note the coarser relation between GSR, skin humidity and EE, and the higher correlation of HR with EE. ACC data is highly correlated to EE for some activities (e.g. running), and weakly correlated for others (e.g. biking). ACC data at the wrist is weakly correlated to EE compared to ACC data at the chest. GSR and skin humidity seem to be good predictors of activity and inactivity.

4. APPROACH

This section covers the approach we used to analyze the role of different physiological signals in PA monitoring. More specifically, we selected a set of methods (single regression and activity-specific) and metrics to evaluate how and when PA assessment can be improved by combining ACC and physiological data. Our analysis is structured as follows:

1. Each signal (ACC or physiological) was evaluated using Pearson's correlation coefficient between the signal and EE, over all the activities of our protocol. This evaluation highlights which signals are good predictors of EE during a range of activities of daily life, regardless of individual differences. Additionally, for each signal a single regression model is implemented to evaluate the accuracy of subject independent models using such signal to estimate EE, and the Root Mean Square Error (RMSE) as metric.

2. Each physiological signal is evaluated as feature in support to ACC features for activity type recognition models. This evaluation highlights which signals can provide useful insights for activity type. The analysis of the relation between activity type and physiological signals is of rising importance, since the recently widespread use of single sensor devices in locations where movement is weakly related to EE (e.g. the wrist) can lead to inaccuracy in PA assessment. Additionally, the coarser relation between physiological signals other than HR and EE (see Fig. 1), shows that physiological signals might be better predictor of activity cluster type than EE changes within an activity cluster.

3. Each signal is evaluated using Pearson's correlation coefficient between the signal and EE, for a single cluster of activities. This evaluation highlights which signals show a linear relation to EE for a specific activity, thus could serve as predictors of EE changes within a cluster, since some clusters of activities can be carried out at different intensities (e.g. walking at different speeds). For each activity cluster a multiple linear regression model is implemented to evaluate the accuracy of such signal in estimating EE for a specific cluster of activities, using the RMSE as metric.

All the analysis are performed considering single sensor devices, to maximize user comfort. More specifically, we combined ACC at the wrist with GSR and skin humidity, and we combined ACC at the chest with HR. All the activity-specific models are implemented using the methodology of [2] et al., including anthropometric characteristics based on the type of activity performed.

5. METHODS

5.1 Participants

The target population considered for our work is healthy adults, since we focus on disease prevention. Therefore, sixteen (12 male, 4 female) healthy young adults took part in the experiment. Mean age was $31.6 \pm 5.9 years$, mean weight was $73.1 \pm 9.9 kg$, mean height was $176.3 \pm 10.1 cm$ and mean BMI was $23.46 \pm 1.69 kg/m^2$. Imec's internal Ethics Committee approved the study. Each participant signed an informed consent form.

Figure 2: The two wearable sensors used in this experiment. Imec's ECG Necklace and Wristband.

5.2 Instruments

Two wearable sensors were used for data collection, imec's ECG Neckalce and Wristband (see Fig. 2). The ECG Necklace was configured to acquire one lead ECG data at $256 Hz$, and ACC data from a three-axial ACC at $32\ Hz$. Two gel electrodes were placed on the participant$'$s chest. Imec's Wristband was configured to acquire phasic and tonic GSR data at $128 Hz$, ACC data from a three-axial ACC at $32 Hz$ and skin humidity at $10.24\ Hz$.

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 [13]. The system was manually calibrated before each experiment according to the manufacturer instructions.

5.3 Experiment Design

Participants invited for recordings had to refrain from drinking (except for water), eating and smoking in the two hours before the experiment. The protocol included sedentary, lifestyle and sport activities (see Table 1). All activities were carried out for a period ranging from 3 to 10 minutes, with a 1 or 2 minutes break between sedentary activities.

We performed our comparative analysis on data collected 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.

5.4 Pre-processing

The dataset acquired in this work consists of reference VO_2, VCO_2, three axial ACC from chest (A-C) and wrist (A-W), ECG, respiration rate, GSR and skin humidity. 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. Breath-by-breath data acquired from the Comsed $K4b^2$ was resampled at $0.5\ Hz$. EE was calculated from O_2 and CO_2. The first minute of each activity were discarded to remove non-steady-state data.

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

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

5.4.1 *Activity Type Clusters*

We split the activities of our protocol into inactive and active clusters. In inactive clusters most of the EE is due to Basal Metabolic Rate, once the activity has been recognized, other features provide no extra information on EE. On the other hand, active clusters account for a higher share of Physical Activity Energy Expenditure, allowing for an analysis of how wearable sensors can provide information about differences in EE during activities of daily life. We further manually split the activities into six clusters related to the activity type and involved motion patterns. We included *lying* and *sedentary* as inactive clusters. We grouped sitting and standing activities into one cluster since the three main postures (lying, sitting and standing) are not recognizable with a single sensor device. Additionally, we included four active clusters, one representative of household activities and dynamic transitions between activities, namely the *high whole body motion cluster (HWBM or Dynamic)* and other three related to locomotion and active transportation, namely *walking*, *biking* and *running* (see Table 1).

5.4.2 *Feature Extraction and Selection*

Features extracted from the ECG necklace and Wristband raw data were used to derive activity recognition (on the six activity clusters covered in Section 5.4.1) and EE models.

Accelerometer Feature Extraction: ACC data from both sensors were segmented in 4 second windows, band-pass (BP) filtered between 0.1 and $10\,Hz$, to isolate the dynamic component, and low-pass (LP) filtered at 1 Hz, to isolate the static component. Time and frequency features were extracted from each window. Time features included; *mean, mean of the absolute signal, magnitude, mean distance between axes, variance, standard deviation, inter-quartiles range* and *median.* Frequency features included; *spectral energy, entropy, low frequency band signal power (*$0.1 - 0.75$ Hz*), high frequency band signal power (*$0.75 - 10$ Hz*), frequency and amplitude of the FFT coefficients.* These features were selected due to high accuracy showed in past research [20].

Physiological Feature Extraction: Three features were extracted from R-R intervals; *mean, variance and standard deviation.* Features extracted from phasic and tonic GSR data were; *mean skin conductance level*, *signal power*, *skin conductance response rate* and *mean Ohmic Perturbation Duration.* The only feature extracted from skin humidity was the *mean.* All features were extracted over non-overlapping 15 seconds windows.

Accelerometer Feature Selection: Feature selection for activity type recognition was based on mutual information [5]. The feature set includes; *mean of the absolute band-passed signal, magnitude and inter-quartile range*, *median*, *variance and standard deviation* and *main frequency peak and amplitude of the main frequency peak.* Feature selection for activity-specific EE models was based on how much variation in EE each feature could explain within one cluster of activities. The process was automated using linear forward selection. Anthropometrics features were added depending on the cluster, following the methodology for activity-specific EE modeling presented in [2]. Additionally, we used *activity counts*, calculated as *mean of the absolute band-passed signal summed over the three axis* to implement single ACC-based regression models.

Physiological Feature Selection: Features derived from physiological signals were used for both activity recognition and EE models. The most discriminative features were selected based on correlation. Selected features were; *mean HR*, *mean skin conductance level*, *mean skin humidity* and *respiration rate.*

5.5 Models Implementation

We implemented activity recognition, single EE models and activity-specific EE models (see Fig. 3).

5.5.1 *Activity Type Recognition*

We adopted a constant set of parameters for sliding window, sampling rate and classifier type of the activity recognition. We selected a time window of 4 seconds, which is short enough to detect short breaks in sedentary time, and long enough to capture the repetitive patterns of certain activities (e.g. walking or running). 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 $= 1$).

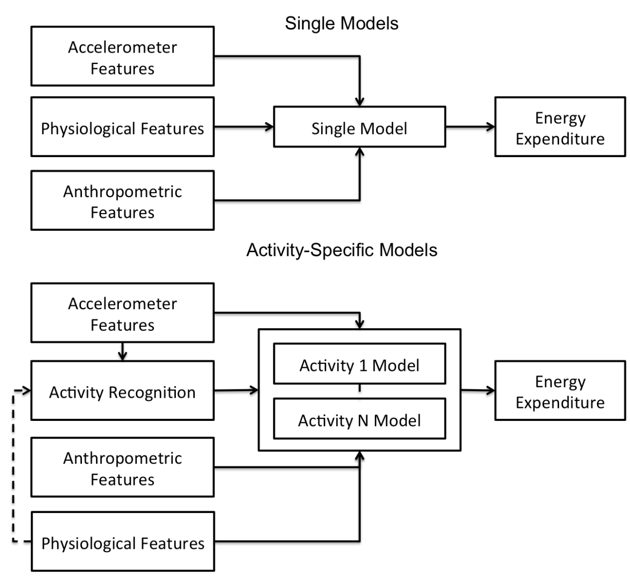

Figure 3: Models implemented in this work. Single models combine all features in one regression model. Activity-Specific models combine activity recognition and multiple regression models. Additionally, we explored the use of combined ACC and physiological signals for activity recognition (dashed line).

5.5.2 *Activity Intensity Estimation (EE)*

Within one activity cluster, EE can be better estimated using activity-specific features, representative for EE changes within the activity cluster (e.g. walking speed for walking). Depending on the signals included in the model (ACC and physiological) we created different EE activity-specific multiple linear regression models for each activity cluster.

5.6 Statistics and Performance Measures

Performance of the activity recognition models was evaluated using the percentage of correctly classified instances. Performance measures used for EE were Pearson's correlation coefficient, and the RMSE. Results are reported in terms of RMSE (leave one subject out cross validation) because of the large inter-individual variability that is typical for EE estimates. Normalization procedures do exist (e.g. kcal/kg), but do not take into account that EE during different activities is affected differently by body weight [3]. Pearson's correlation coefficient is used to quantify the predictive power of different signals in estimating EE, without being affected by inter-individual differences (average between subjects). As statistical analysis, we performed paired t-tests. Significance level α was set to 0.05 for all tests.

6. RESULTS

6.1 Single Models

Correlation coefficient between the signals and EE was 0.74 for A-C, 0.64 for A-W, 0.93 for HR, 0.76 for GSR, 0.72 for respiration rate and 0.70 for skin humidity. Subject independent RMSE was 2.16 $kcal/min$ for A-C, 2.59 $kcal/min$ for A-W, 1.88 $kcal/min$ for HR, 2.50 $kcal/min$ for GSR, 2.78 $kcal/min$ for respiration rate and 2.52 $kcal/min$ for skin humidity (see Fig. 4). Combining signals for the ECG Necklace (A-C and HR) reduced the RMSE to 1.59 $kcal/min$, while combining signals for the Wristband (A-W, GSR and skin humidity) reduced the RMSE to 1.95 $kcal/min$.

6.2 Activity Clusters Classification

Subject independent classification accuracy of activity type for the ECG Necklace using ACC features only was 93%. Recognition accuracy on each cluster is shown in Fig. 5. Performance was improved by 2% when HR was included ($p = 0.07 > \alpha$). Accuracy for the Wristband was 72% (see Fig. 5). Accuracy increased by 8% when GSR and skin humidity were included in the model ($p = 0.0019 < \alpha$). Inclusion of the respiration signal in the ECG Necklace model did not improve accuracy. Fig. 7 shows the misclassification rate for each cluster. Activity misclassification for the Wristband was higher for each activity cluster.

6.3 Activity-Specific EE Models

The relation between ACC, physiological signals and EE within an activity cluster was analyzed for active clusters only (i.e. *dynamic, walking, biking* and *running*). Correlation coefficient between HR and EE for each cluster of activities outperformed all other signals (0.57 for *dynamic*, 0.92 for *walking*, 0.96 for *biking* and 0.92 for *running*). Correlation coefficient between GSR, skin humidity and EE was below 0.21 for *dynamic* and ranged between 0.32 and 0.67 for *walking, biking* and *running.* Correlation ranged between 0.40 and 0.77 for A-C, while was always weak for A-W (< 0.51).

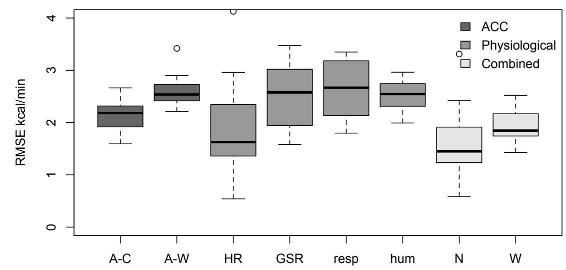

Figure 4: RMSE for all the signals included in this study, subject independent evaluation of single models. ECG Necklace (N) combines A-C and HR, Wristband (W) combines A-W, GSR and humidity.

6.3.1 *Assuming Perfect Activity Recognition*

Subject independent RMSE for activity-specific multiple linear regression models, assuming perfect activity recognition, are shown in Fig. 6. Activity-specific models including HR for the ECG Necklace outperform ACC only models for *walking* and *biking* clusters. The Wristband's RMSE was not significantly reduced by combining ACC and physiological signals in any activity cluster.

6.3.2 *Including Cluster Misclassification*

Misclassification for the Wristband often concerns confusion between inactive (e.g. *sedentary*) and active (e.g. *biking*) clusters, leading to higher EE estimation errors due to the application of the wrong activity-specific model. However, confusion between inactive and active clusters is highly reduced when physiological signals were used for activity recognition (*lying* and *sedentary* misclassification as *biking* goes from 20% to 1%). Fig. 8 shows the reduction in RMSE for all clusters, when physiological signals were included. Overall, misclassification causes RMSE to increase from 0.94 (average of the six clusters) to 0.99 $kcal/min$ for the ECG Necklace, and from 1.05 to 1.25 $kcal/min$ for the Wristband.

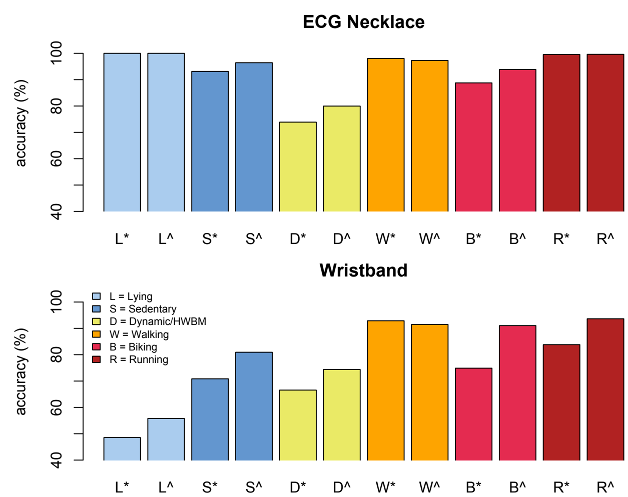

Figure 5: Accuracy for the activity recognition models implemented in this study. Results are shown per cluster of activities. * indicate ACC models, Λ indicate combined models (ACC+physiological).

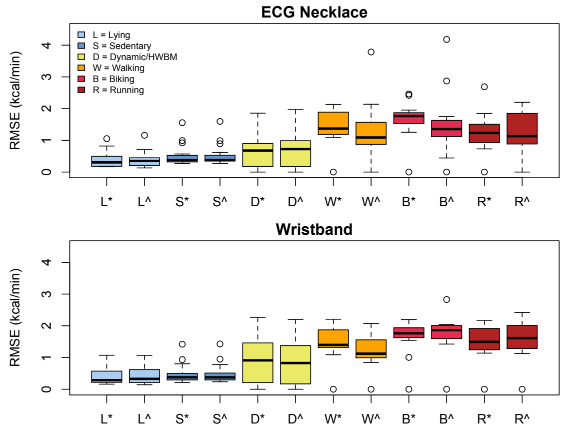

Figure 6: RMSE of activity-specific EE models for the ECG Necklace and Wristband sensors, assuming perfect activity recognition. * indicate ACC models, Λ indicate combined models.

However, RMSE after misclassification is 35 and 36% lower for the ECG Necklace ($p = 0.0008 < \alpha$) and Wristband ($p = 0.00001 < \alpha$) activity-specific models compared to single models. RMSE is reduced by 18 - ECG Necklace - and 46% - Wristband - when including physiological signals.

7. DISCUSSION AND CONCLUSION

In this study, we evaluated the impact of ACC and multiple physiological signals acquired from two body locations on activity type recognition and EE estimation. To the best of our knowledge, this is the first time that state of the art activity recognition and activity-specific EE models are jointly evaluated to determine benefits of using different physiological signals. Especially when developing activity-specific models, evaluating the benefit of multiple signals at each stage of the estimation process (activity recognition and activity-specific EE models) is important, since different signals relate differently to levels of physical exertion, and can contribute distinctively. We report three main findings.

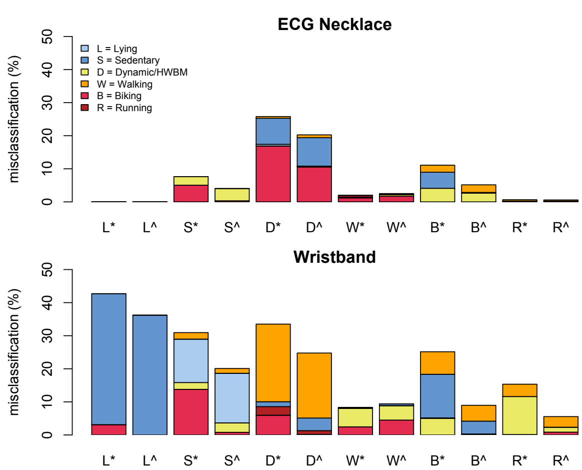

Figure 7: Misclassification of the activity recognition models per cluster of activities.* indicate ACC models Λ indicate combined models.

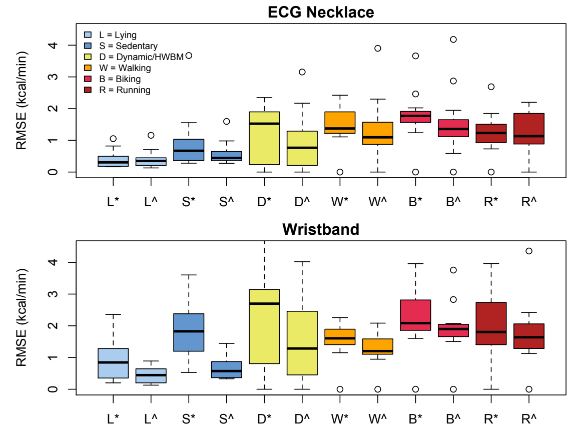

Figure 8: RMSE of activity-specific EE models for the ECG Necklace and Wristband sensors, including misclassification effects.* indicate ACC models, Λ indicate combined models.

a) On physiological signals: HR is the best predictor of EE, given the direct relation to oxygen uptake, while other physiological signals have a coarser relation to level of physical exertion, which makes them better predictors of activity and inactivity conditions, or activity type clusters (when combined with ACC). Given the coarse relation between physiological signals other than HR and EE, these signals (GSR level, respiration rate, skin humidity), do not provide extra information on differences in EE for one activity cluster, compared to ACC only. Correlation coefficient between EE and GSR, skin humidity, respiration was relatively high when computed over the complete set of activities (range 0.7 to 0.76). However, when the analysis is broken down to the activity cluster level, these signals are weakly correlated to EE (range 0.21 to 0.67), compared to HR.

b) On activity recognition: As expected, the ECG Necklace showed better performance compared to the Wristband ($15 - 21\%$ difference), due to the positioning close to the body's center of mass, where motion is more representative of PA levels. Activity misclassification of the Wristband is due to the fact that not only movement at the wrist is weakly related to EE, but also to activity type (high intensity of wrist movement can be detected even at rest, while e.g. writing). By combining ACC with GSR and skin humidity, the misclassification error between inactive and active clusters could be significantly reduced. Previous research underestimated the importance of physiological signals in activity type recognition, since multiple accelerometer were used [20]. However, when dealing with single sensor devices (a condition necessary to improve user comfort), physiological data brings significant improvements (e.g. *lying* and *sedentary* misclassification as *biking* went from 20% to 1%).

c) On combining signals: Combining ACC and physiological signals improved performance of both activity type recognition and EE estimation, since the two sensor modalities are often complementary, regardless of the models used (single regression or activity-specific). Especially when the sensor is located where motion is weakly related to activity type and EE (e.g. the wrist), combining ACC and physiological signals showed the most significant improvements (8% in activity recognition, 46% in EE estimation).

We recognize limitations in this study. Even though performance of PA monitoring systems should be assessed in free living conditions, there is currently no reference system able to measure breath-by-breath EE in unconstrained settings. Only by using indirect calorimetry and supervised settings we can record data which allows us to analyze how multiple signals affect the EE estimate process in both activity recognition and EE within one activity cluster.

The aim of our analysis was to understand what physiological signals can be used for physical activity assessment, in terms of both activity type recognition and EE estimation. Some signals showed high correlation with EE but failed to reduce RMSE compared to ACC alone when analyzed in a subject independent manner. This behavior is expected, since physiological signals are highly dependent on the subject, and typically require individual calibration to predict EE accurately. The causes behind these individual differences can be of different nature depending on the signal taken into account (for example cardiorespiratory fitness is considered the main factor behind variability in HR during physical effort). Therefore, to fully exploit the discriminative power of physiological signals for both activity type recognition and EE estimation, new procedures able to automatically normalize differences in physiological signals between individuals are necessary. We are focusing our future work on such personalization techniques, aiming at introducing methodologies able to normalize various physiological signals independently of the causes which are generating differences between individuals. Towards this aim, we already introduced a technique able to personalize HR based EE estimations [4]. Additionally, non-steady-state and transitions between activities, where changes in physiological signals lag behind changes in activity type and EE, should be investigated and modeled.

8. REFERENCES

[1] 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.

[2] M. Altini, J. Penders, and O. Amft. Energy expenditure estimation using wearable sensors: A new methodology for activity-specific models. In *Wireless Health '12*, 2012.

[3] M. Altini, J. Penders, and O. Amft. Body weight-normalized energy expenditure estimation using combined activity and allometric scaling clustering. In *EMBC '13*, in-press.

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

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

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

[7] 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.

[8] 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.

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

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

[11] 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.

[12] 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.

[13] J. E. McLaughlin. Validation of the COSMED K4 b2 portable metabolic system. *International journal of sports medicine*, 22(4):280–284, May 2001.

[14] J. N. Morris and M. D. Crawford. Coronary heart disease and physical activity of work. *British Medical Journal*, 2(5111):1485, 1958.

[15] T. Pawar. Body movement activity recognition for ambulatory cardiac monitoring. *IEEE transactions on biomedical engineering*, 54(5):874–882, 2007.

[16] 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.

[17] 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.

[18] J. Smolander. A new heart rate variability-based method for the estimation of oxygen consumption without individual laboratory calibration: application example on postal workers. *Applied Ergonomics*, 39(3):325–331, 2008.

[19] 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):659–667, 2011.

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

[21] 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.

[22] G. J. Welk, J. J. McClain, J. C. Eisenmann, and E. E. Wickel. Field validation of the mti actigraph and bodymedia armband monitor using the ideea monitor. *Obesity*, 15(4):918–928, 2007.