Comparative evaluation of physical activity parameters based on wrist-wearable devices

T. Hübner1, G. García-Sáez1, J. M. Iniesta1, J. Tapia1, C. Pérez-Gandia2,1, M.E. Hernando1,2

1 Centro de Tecnología Biomédica, ETSI de Telecomunicación, Universidad Politécnica de Madrid, Spain
tilmann.huebner@gmail.com, {ggarcia,jminiesta,jtapia,cperez,mariaelena.hernando}@gbt.tfo.upm.es
2CIBER-BBN: Networking Research Centre for Bioengineering, Biomaterials and Nanomedicine, Madrid, Spain

Summary
Physical activity is a significant part of the treatment and prevention of a wide range of diseases. A modern approach to classification of physical activity is the use of accelerometers that are available in many devices. This work compares characteristic parameters for physical activity which are measured by wrist-wearable devices. The amount of calories burnt, the number of steps taken and the heart rate of four users were recorded while the user is exercising on a treadmill in a physical activity session that included different intensity levels estimated as metabolic equivalents of tasks (METs).

The measured amount of calories burnt and number of steps taken showed to be very similar over all of the devices and users. Higher intensity levels produce a greater variance in the measured data. The measurement of the heart rate showed indistinctive results as the data varied strongly over users and devices. A deeper analysis is needed to overcome synchronization problems.

The study showed that the measurement of physical activity with wrist-wearable devices is reliable and the results for commercial devices and a proprietary algorithm are comparable to each other.

1. Introduction
An active lifestyle is considered an essential part in the treatment and prevention for a wide range of chronic diseases that have become very common [1,2]. Monitoring the Physical Activity (PA) of a patient can help to analyze a patient’s behavior and give advice for its further treatment. This requires an objective assessment of the activity and has become an increasing topic in research [3,4].

An interesting part about PA is the amount of energy used, which is often also called energy expenditure (EE). Different methods have been established to measure the EE and quantify the PA. Currently direct calorimetry, indirect calorimetry and doubly labeled water are considered as the gold-standard techniques [4]. Their advantages are the very precise results, but they also have some disadvantages: All of these techniques are very expensive to use and no usable in free-living conditions as a complex setup of technologies is needed. An easier measurement method is the estimation of the EE from the heart rate (HR). It has been shown that the HR almost linearly follows the oxygen uptake ratio (VO2) and energy consumption in case of aerobic exercise [5]. A rather new but promising approach is the use of multi-axial accelerometers. The costs of these sensors are decreasing and they become more and more available in very common and non-specialized devices such as smartphones. The wide-spread distribution makes them a very attractive tool to be used in research [6].

The accelerometers are placed at key locations of the body (for example hip or wrist) to enhance detection. Basic personal characteristics like weight, height and sex are included in the calculations to estimate the EE by linear regression from the acceleration data, or using acceleration counts (also known as ‘activity counts’ based on pre-processed raw data with proprietary algorithms) [7-10].

A key part in the estimation of the energy expenditure efficiently is the classification of the intensity level of the detected physical activity as they are closely related. The energy expenditure induced by physical activity is primarily related to the body weight of the person. Better comparison across individuals can be done with normalization or with alternative magnitudes such as metabolic equivalents of tasks (METs). METs express energy costs as multiples of the basal metabolic rate (BMR) at rest. Several studies have shown that METs are simple descriptors of workload levels across activity modalities and populations [11,12]. Classification of the intensity level from very light to vigorous intensity is done by using multiples of METs. So far this has been done by using regression techniques with threshold values or ad hoc models with a priori knowledge about the nature of the classification problem [13].

Many different approaches to the exact implementation of the classification algorithm have been taken in research and as well in commercially available products. Most of the algorithms are designed for a specific device and its accelerometer. Our research group has developed a Physical Activity Detector (PAD) which is an algorithm integrated in an Android application that estimates and classifies the energy cost of human PA related to METs intensity values. The PAD algorithm uses the internal accelerometer of a smartphone to calculate the type of activity, the intensity of the exercise, the number of steps walked/run and estimates the energy used by means of the METs and the calories burnt. It was initially developed for a Samsung S4 device [14]. The same algorithm has been implemented in a smartwatch with Android Wear to detect physical activity based on the internal accelerometer.
This paper focuses on the comparison of classification results of the PAD algorithm to other commercial devices and algorithms.

2. Materials and Methods

2.1. Materials

For intra device comparison a number of different types of devices are used at the same time and the output of these devices are compared to each other. The devices used are:

- Samsung S4 (running PAD algorithm)
- Samsung S6 (running PAD algorithm)
- Motorola moto360 (smartwatch, running an adaption of PAD algorithm)
- Two Microsoft Band 2 (running proprietary algorithm, but saving raw data as well)
- Axivity AX3 (only for documentation of raw accelerometer data)
- Zephyr BioHarness (documentation of accelerometer data and heart rate)

The position of the sensors is shown in Figure 1.

![Figure 1. Sensor Position. Green = Microsoft Band 2, dark blue = Microsoft Band 2, pink = Motorola moto360, turquoise = Axivity AX3, purple = Zephyr BioHarness, light blue = Samsung S4, red = Samsung S6.](image)

2.2. Methods

The devices output differ slightly from each other. Therefore only the similar parameters will be compared. This includes the following:

- Steps taken by the subject.
- Calories burnt.
- Heart rate (when available in the device).

The evaluation group consisted of 4 users, all of them males, within an age range of 22 – 38, with different fitness levels (weight 65 kg – 82 kg, 1.75m – 1.85m). All participants performed a PA session following the same protocol that was designed to develop and test the PAD algorithm in a previous work. This consists of seven different levels of activities at different speeds while the user is exercising on a treadmill. Table 1 provides a list of the activity protocol. Each participant wore all the devices shown in Figure 1 at the same time.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Speed (km/h)</th>
<th>Duration (min)</th>
<th>Theoretical METs (Ainsworth [11])</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking</td>
<td>3.2</td>
<td>3</td>
<td>2.8</td>
</tr>
<tr>
<td>Walking</td>
<td>4.8</td>
<td>3</td>
<td>3.5</td>
</tr>
<tr>
<td>Walking</td>
<td>5.6</td>
<td>3</td>
<td>4.3</td>
</tr>
<tr>
<td>Running</td>
<td>6.4</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Running</td>
<td>8.0</td>
<td>3</td>
<td>8.3</td>
</tr>
<tr>
<td>Running</td>
<td>11.3</td>
<td>2</td>
<td>11</td>
</tr>
</tbody>
</table>

**Table 1. PA session - Activity protocol.**

3. Results

The amount of calories burnt during the PA session was calculated by each device and is displayed in Figure 2. Over all devices and users the curse is very similar at the beginning. As the speed increases, the variance between devices becomes greater. The variance within each device remains fairly steady.

![Figure 2. The calories burnt during the physical activity by each of the users and each device is shown over time.](image)

The mean and standard deviation (SD) was calculated for each user over all of the devices depending on the type of activity (walking or running). Table 2 shows the results of these calculations. The deviation for each user is small and the variation in between the users shows very similar results too. The relative percentage of the SD to the mean is about 20% for all the users.

<table>
<thead>
<tr>
<th>User</th>
<th>μ(calories walked)</th>
<th>SD(calories walked)</th>
<th>μ(calories run)</th>
<th>SD(calories run)</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>25.6</td>
<td>0</td>
<td>63.9</td>
<td>0</td>
</tr>
<tr>
<td>02</td>
<td>25.5</td>
<td>7.5</td>
<td>77.7</td>
<td>10.4</td>
</tr>
<tr>
<td>03</td>
<td>31.7</td>
<td>6.2</td>
<td>80.6</td>
<td>13.9</td>
</tr>
<tr>
<td>04</td>
<td>33.2</td>
<td>9.0</td>
<td>89.1</td>
<td>35.3</td>
</tr>
</tbody>
</table>

**Table 2. Mean and SD of the number of calories burnt by each user during both types of physical activity.**

The measured number of steps shows similar results and is shown in Figure 3. Although the variance in between the devices appears to be smaller. All curves show a very similar behavior. The total number of steps measured by each of the devices for a user is very similar.
Figure 3. The number of steps during the PA session by each user and each device is shown over time.

Table 3. Mean and SD of the steps taken by the user during both types of physical activity.

<table>
<thead>
<tr>
<th>User</th>
<th>(\mu) (steps walked)</th>
<th>SD (steps walked)</th>
<th>(\mu) (steps run)</th>
<th>SD (steps run)</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>760</td>
<td>0</td>
<td>1238</td>
<td>0</td>
</tr>
<tr>
<td>02</td>
<td>625</td>
<td>131.5</td>
<td>1386</td>
<td>361.3</td>
</tr>
<tr>
<td>03</td>
<td>754</td>
<td>95.6</td>
<td>1352</td>
<td>265.7</td>
</tr>
<tr>
<td>04</td>
<td>711</td>
<td>124.8</td>
<td>1317</td>
<td>192.4</td>
</tr>
</tbody>
</table>

The same mean and SD analysis as for the calories was performed for the number of steps taken. The data was classified into two types of activity (walking and running) and then the mean and the SD of the total amount of steps taken during each type of activity calculated (Table 3). It shows that the deviation of the number of steps is greater during the more intense activity of running. The relative percentage of the SD in comparison to the mean is 10-15% and increases to about 20-25%. The comparison in between users shows very similar results. The mean number of steps taken is quite similar and the SD during walking as well. During the running exercise the deviation in between the users rises significantly.

Figure 4. The measured hear rate during the physical activity by each user and each device is shown over time.

4. Discussion

During the study several problems with the measurement systems occurred. Sometimes data was not saved by the device because of connection problems or software bugs. The physical activity data from the Samsung S4 and Samsung S6 was not used for analysis because the parameters were only calculated every 5 minutes. The acceleration data was stored continuously but during the time of the study there was no offline program available to extract the physical activity parameters from the stored data.

The analysis process revealed as well a problem with synchronization of the data. Apparently the devices were set to slightly different times and therefore the data does not always start at the same time. These problems could not be resolved automatically during data analysis.

Despite all problems the results reveal a good similarity between the wrist-wearable devices. Both the burnt calories and the steps taken show for all devices the same trend. Figure 5 shows a boxplot of the available data for the burnt calories. The greater deviation of the burnt calories during running becomes visible. This is partly due to the differences in the users, as the height and weight of the user influences the amount of calories burnt and as well to greater variances in the data recording process.

Figure 5. Boxplot of the mean of the calories burnt and the SD for all users.

Figure 6 shows the same analysis for the number of steps taken. The variance in the mean number of steps taken appears to remain equal for both activities. The great outliers in the SD are because of a lack of data.

The measurement of the HR does not seem to be very reliable as there are times without a valid signal. In Figure 7 the measured heart rate is grouped into intervals of one minute. The great range of outliers and confidence interval support the assumption about the limited reliability of the HR sensors on wrist-based devices. As the marked means show a trend can be derived from the heart rate sensors but relying on only one of these devices can cause significant false assumptions of the current HR.
Figure 6. Boxplot of the mean number of steps and the SD for all the users.

Figure 7. Boxplot of the HR. Each box represents a time interval of one minute and accumulates the HR measured by all the devices and for each user.

5. Conclusion

The study showed that the measurement of PA with wrist-wearable devices is possible and the results for different devices are comparable to each other.

More work has to be done with deeper analysis of the available data and techniques to overcome synchronization problems have to be developed. The analysis of the data should be repeated for a greater number of users in order to achieve more reliable results.

For the HR analysis it has to be considered that these graphs show the measured HR of all the users. Because of different fitness levels each user reacts differently and the HR does not necessarily increase in the same manner.

Acknowledgments

We would like to thank the Department of Health and Human Performance at the Faculty of Physical Activity and Sport Sciences-INEF for their collaboration. This work has been funded by the Spanish grant “FitCLoop” (PI14/00109), co-funded by FEDER.

References


