Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: laboratory-based calibrations are not enough

Thomas Bastian,1 Aurélia Maire,1 Julien Dugas,1 Abbas Ataya,2,3 Clément Villars,1 Florence Gris,2,3 Emilie Perrin,4 Yanis Caritu,4 Maeva Doron,2,3 Stéphane Blanc,5 Pierre Jallon,2,3 and Chantal Simon1,6

1CarMeN INSERM U1060, University of Lyon 1, INRA U1235, Centre de Recherche en Nutrition Humaine Rhône-Alpes, Centre Européen pour la Nutrition & la Santé, Pierre-Bénite, France; 2University of Grenoble Alpes, Grenoble, France; 3Commissariat à l’Énergie Atomique, Leti, Département Microtechnologies pour la Biologie et la Santé, Laboratoire Électronique et Systèmes pour la Santé, MINATEC, Grenoble, France; 4Movea, Grenoble, France; 5Hubert Curien Pluridisciplinary Institute, Department of Ecology, Physiology and Ethology, University of Strasbourg, UMR CNRS 7178, Strasbourg, France; and 6Service d’Endocrinologie, Diabètes, Nutrition, Centre Hospitalier Lyon Sud, Pierre-Bénite, France

Submitted 29 October 2013; accepted in final form 26 December 2014

Bastian T, Maire A, Dugas J, Ataya A, Villars C, Gris F, Perrin E, Caritu Y, Doron M, Blanc S, Jallon P, Simon C. Automatic identification of physical activity types and sedentary behaviors from triaxial accelerometer: laboratory-based calibrations are not enough. J Appl Physiol 118: 716–722, 2015. First published January 15, 2015; doi:10.1152/japplphysiol.01189.2013.—“Objective” methods to monitor physical activity and sedentary patterns in free-living conditions are necessary to further our understanding of their impacts on health. In recent years, many software solutions capable of automatically identifying activity types from portable accelerometry data have been developed, with promising results in controlled conditions, but virtually no reports on field tests. An automatic classification algorithm initially developed using laboratory-acquired data (59 subjects engaging in a set of 24 standardized activities) to discriminate between 8 activity classes (lying, slouching, sitting, standing, walking, running, and cycling) was applied to data collected in the field. Twenty volunteers equipped with a hip-worn triaxial accelerometer performed at their own pace an activity set that included, among others, activities such as walking the streets, running, cycling, and taking the bus. Performances of the laboratory-calibrated classification algorithm were compared with those of an alternative version of the same model including field-collected data in the learning set. Despite good results in laboratory conditions, the performances of the laboratory-calibrated algorithm (assessed by confusion matrices) decreased for several activities when applied to free-living data. Recalibrating the algorithm with data closer to real-life conditions and from an independent group of subjects proved useful, especially for the detection of sedentary behaviors while in transports, thereby improving the detection of overall sitting (sensitivity: laboratory model = 24.9%; recalibrated model = 95.7%). Automatic identification methods should be developed using data acquired in free-living conditions rather than data from standardized laboratory activity sets only, and their limits carefully tested before they are used in field studies.

actimetry; physical activity; sedentary behaviors; machine learning and classification methods; field study

DEEPENING OUR UNDERSTANDING of the impacts of physical activity and sedentary behaviors on health requires objective methods to measure these behaviors both quantitatively and qualitatively in free-living conditions (2, 8). This is even more critical as it becomes increasingly clear that sedentary occupations and low nonexercise activities can have significant health impacts, independent of exercise (10, 18).

For several years now, accelerometers have been one of the main instruments used by physical activity researchers and epidemiologists to investigate these questions (7, 13). Over the past few years, new generations of accelerometers have emerged, which have enabled raw triaxial acceleration signals to be recovered at high frequency. Combined to advances in signal processing techniques and machine-learning algorithms, this paved the way for the development of methods capable of automatically identifying postures or types of physical activity from raw acceleration signals (6). A lot of research has been conducted in that area, and a wide range of machine-learning approaches and classification techniques have been developed and tested in controlled conditions (27). Most activity sets used to generate the data necessary to calibrate these tools in the laboratory now include real-life situations of the continuum from light activities to full sedentariness, which still remain challenging to accurately dissociate (6). Although there have been attempts to include sitting (e.g., watching TV, reading, desk-based occupations), “household” (e.g., hovering, washing dishes), and locomotory (e.g., walking, climbing stairs) activities (25) in activity sets used in laboratory studies, only a few studies have considered the case of riding a bicycle on the street (19) or sitting in a motorized vehicles (24). Those activities are now recognized as important targets for public health campaigns aiming at decreasing sedentary time and increasing daily physical activity levels at population level, but further studies are needed (26). Clearly, the identification of these activities might also be useful to improve the assessment of activity-related energy expenditure (5). Most importantly, studies testing the algorithms obtained in controlled laboratory conditions on independent samples out of highly controlled laboratory conditions (i.e., in the field) remain very scarce (5, 12, 14, 23).

In the present work, two versions of an automatic posture identification tool based on a machine-learning algorithm were tested on data collected in free-living conditions. The first version tested was “trained” only with data acquired in the laboratory, whereas the second version was initialized with additional triaxial acceleration signals from activities performed out of the laboratory, including cycling the street and sitting in motorized vehicles (car, bus, subway). Comparing the performances of the two tested versions highlighted the importance of using data acquired in free-living conditions to pa-
rarameterize and test activity identification and classification algorithms before using them in the field, but also demonstrated some remaining limits of this approach.

**MATERIAL AND METHODS**

**Classification algorithm.** The algorithm used in the present study is based on a Bayesian machine-learning approach, which enables an automatic identification of postures or activity classes from the raw signals recorded by a single hip-worn triaxial accelerometer (9). The devices used were MotionLogs (Movea, Grenoble, France). Acceleration signals were acquired at 100 Hz and subsequently downsampled to 25 Hz (by keeping one of every four data points) to ease computations, while keeping a high enough frequency to assess physical activities [most of human movements occur at < 10 Hz (33)].

Three features from the time (mean) and frequency domains (total energy and dominant frequency) were extracted for the vertical axis, based on a Bayesian machine-learning approach, which enables an expectation-maximization algorithm (34). The number of components $K$ considered sufficient to model the activities.

**Out-of-the-laboratory tests.** The performances of this laboratory model were tested on an independent set of data collected in semi-free-living conditions, outside of the laboratory. Twenty other volunteers, aged 18–39 yr, engaged in a different set of activities, this time including more real-life field situations, e.g., walking in the street, riding a true bicycle, getting on the bus (full list in Table 1). Subjects were instructed to perform activities at their own pace in bouts of ~3 min or more. In other words, the subjects walked as they wanted, able to climb the stairs up and down as fast as it pleased them (one step at a time, or two at time), and while cycling they did as much turns as they wanted. Data for sitting or standing in transport were recorded while the subjects used the actual public transport system (bus, tramway, subway) of the city of Lyon (France), except for sitting in a car, during which the subject was sitting as a passenger in a private car engaged on highways and city-center small streets. The beginning and the end of each activity bouts were precisely timed (hour, minutes, seconds) and noted by an observer. Data were then extracted from the devices and treated with the classification algorithm. For each activity, classification outputs were compared with the corresponding a priori expected activity class (Table 1). To do so, the 1/6-Hz classification outputs were scaled to 1 Hz by constant interpolation and then compared second by second to the observers’ notes.

**Recalibration.** In an effort to improve the initial classification algorithm (the laboratory model), one-half of the “out-of-the-laboratory” dataset (see previous paragraph) was added to the initial training set, and a new learning phase was initialized with three additional classes: “riding a bicycle” (in contrast to an exercise bicycle), “sitting in a vehicle,” and “standing in a vehicle.” This resulted in a recalibrated algorithm (or model), which discriminates between 11 postures. Performances (sensitivity, specificity, and precision) of this new model were tested using the second half of the out-of-the-laboratory dataset and compared with performances of the laboratory-calibrated model on the same data.

**Data acquisition in the laboratory and initial learning phase.** The data to initialize the first version of the algorithm were collected as follows: Fifty-nine volunteers, aged 19–55 yr (mean ± SD = 37.3 ± 10.6 yr), were recruited to perform a set of standardized activities in the laboratory. Subjects were selected according to their body mass index (BMI) (26 with a BMI < 25, 16 with 25 < BMI < 30, and 17 with BMI > 30), sex (30 men, 29 women), and physical activity level (31 active, 28 inactive subjects), according to the Monica Optional Study of Physical Activity questionnaire (29). The activity set included 23 activities of various intensities (rest to vigorous activity), performed in consecutive bouts of 5 min, after an initial time period of 45 min lying down. Each activity was associated with one of the eight activity classes predefined for the training of the algorithm, and the learning phase was conducted. The algorithm resulting from this procedure will be hereafter referred to as the laboratory model.

A cross-validation using a take-one, leave-one-out procedure was performed to evaluate the performances of this algorithm. Results of this cross-validation are reported as the aggregate detection accuracy or “good detection rate” for each activity. This aggregate accuracy is obtained by averaging the activity detection accuracy for each activity over all the subjects. For an activity $i$, the aggregate accuracy is given by:

$$\frac{1}{N} \sum_{n=1}^{N} 100 \times \frac{1}{T_i} \sum_{t=1}^{T_i} 6(A_n^i - i)$$

where $N$ is the total number of subjects in the database, $\delta(x)$ is a function that is equal to 1, if $x = 0$, and 0 otherwise, $A_n^i$ represents the detected activity at instant $t$ for subject $n$, and $T_i$ represents the real number of instants during which subject $n$ practiced activity $i$.
Table 1. List of the activities performed during the laboratory trial (from which the initial laboratory-calibrated model was developed) and of the activities of the semi-free-living condition trial

<table>
<thead>
<tr>
<th>Activities</th>
<th>Details</th>
<th>Activity Class</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Laboratory conditions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lying down</td>
<td>Lying down on the back, on a bed, still</td>
<td>Lying down</td>
</tr>
<tr>
<td></td>
<td>Lying down on the back, on a bed, moving arms up and down every 30 s</td>
<td>Lying down</td>
</tr>
<tr>
<td></td>
<td>Lying down on the back, on a bed, moving arms up and down every 15 s</td>
<td>Lying down</td>
</tr>
<tr>
<td>Slouching</td>
<td>Slumped/slouched in a comfortable armchair</td>
<td>Slouching</td>
</tr>
<tr>
<td>Sitting</td>
<td>On a chair, back straight, still</td>
<td>Sitting</td>
</tr>
<tr>
<td></td>
<td>On a chair, lifting a 50-cl bottle up and down with both arms</td>
<td>Sitting</td>
</tr>
<tr>
<td></td>
<td>Throwing a ball to the experimenter</td>
<td>Sitting</td>
</tr>
<tr>
<td></td>
<td>At a desk, inactive, elbows on the desk</td>
<td>Sitting</td>
</tr>
<tr>
<td></td>
<td>At a desk, typing a text on a computer</td>
<td>Sitting</td>
</tr>
<tr>
<td>Standing up</td>
<td>Standing still</td>
<td>Standing</td>
</tr>
<tr>
<td></td>
<td>Lifting a 50-cl bottle up and down with both arms</td>
<td>Standing</td>
</tr>
<tr>
<td></td>
<td>Lifting a 100-cl bottle up and down with both arms</td>
<td>Standing</td>
</tr>
<tr>
<td></td>
<td>Making few steps, pacing/stalling</td>
<td>SUM</td>
</tr>
<tr>
<td>Interspersed ambulation</td>
<td>Standing up, occasionally making a few steps to sweep a small area of the rooms’ floor</td>
<td>SUM</td>
</tr>
<tr>
<td>Sweeping the floor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleaning windows</td>
<td>Standing up, cleaning a window</td>
<td>Standing</td>
</tr>
<tr>
<td></td>
<td>Walking at 20% VO₂ reserve, no inclination</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>Walking at 30% VO₂ reserve, no inclination</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>Walking at 40% VO₂ reserve, no inclination</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>Walking at moderate speed (subject dependent), 10% slope</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>Walking at moderate speed (subject dependent), 5% slope</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>Running at 70% VO₂ reserve</td>
<td>Running</td>
</tr>
<tr>
<td>Running on a treadmill</td>
<td>Cycling between 60 and 80 rpm at 50% VO₂ reserve</td>
<td>Cycling</td>
</tr>
<tr>
<td></td>
<td>Cycling between 60 and 80 rpm at 70% VO₂ reserve</td>
<td>Cycling</td>
</tr>
<tr>
<td>Semi-free-living conditions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lying down</td>
<td>On a bed</td>
<td>Lying down</td>
</tr>
<tr>
<td>Slouching</td>
<td>In a reclined chair</td>
<td>Slouching</td>
</tr>
<tr>
<td>Sitting</td>
<td>On a chair, at a table</td>
<td>Sitting</td>
</tr>
<tr>
<td>Housecleaning</td>
<td>Sweeping the floor, washing dishes</td>
<td>SUM</td>
</tr>
<tr>
<td>Climbing stairs</td>
<td>Up and down 3 levels, 2 or 3 times</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>Up and down 3 levels, 2 or 3 times</td>
<td>Standing</td>
</tr>
<tr>
<td>Taking a lift/elevator</td>
<td>On flat ground</td>
<td>Cycling</td>
</tr>
<tr>
<td>Cycling on the street</td>
<td>Downhill</td>
<td>Cycling</td>
</tr>
<tr>
<td></td>
<td>Uphill</td>
<td>Cycling</td>
</tr>
<tr>
<td>Running on the street</td>
<td>On flat ground, subject’s own pace, 3 min</td>
<td>Running</td>
</tr>
<tr>
<td>Walking on the street</td>
<td>On flat ground</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>Walking</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>On flat ground and carrying bags</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>Uphill</td>
<td>Walking</td>
</tr>
<tr>
<td></td>
<td>Downhill</td>
<td>Walking</td>
</tr>
</tbody>
</table>

**Table 1.—Continued**

<table>
<thead>
<tr>
<th>Activities</th>
<th>Details</th>
<th>Activity Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sitting in transport</td>
<td>In a car</td>
<td>Sitting</td>
</tr>
<tr>
<td></td>
<td>In a bus</td>
<td>Sitting</td>
</tr>
<tr>
<td></td>
<td>In the subway</td>
<td>Sitting</td>
</tr>
<tr>
<td></td>
<td>In a tramway</td>
<td>Sitting</td>
</tr>
<tr>
<td>Standing in transport</td>
<td>In the subway</td>
<td>Standing</td>
</tr>
<tr>
<td></td>
<td>In a tramway</td>
<td>Standing</td>
</tr>
<tr>
<td>Using mechanical stairs</td>
<td>Up and down without actively walking</td>
<td>Standing</td>
</tr>
<tr>
<td>Shopping</td>
<td>Walking in a shop intermittently stopping in front of shelves</td>
<td>SUM</td>
</tr>
</tbody>
</table>

VO₂, O₂ uptake; SUM, small utilitarian movements.

**Protocol approval.** The present study was approved by the French Sud-Est 2 Institutional Review Board. All subjects signed an informed consent.

**RESULTS**

Initial learning phase and cross-validation in laboratory conditions. Results of the cross-validation procedure on the initial training set are presented in Table 2. The best performances were achieved for lying down, walking, and running (93.4, 95.4, and 93.5%, respectively). The aggregate accuracy of cycling was 66.3%.

Application to semi-free-living condition data. When applied to data acquired in semi-free-living conditions, the true positive rate in the identification of lying down and running were very high (sensitivity of 96.6 and 100%, respectively), but the detection of walking decreased to 76%, and those of sitting, cycling, and standing dropped drastically (24.9, 5.4, and 2.4%, respectively) (see Tables 2 and 3). Low performances of sitting activities were due to the classification of sitting in a car, in the subway, or in the tramway being mostly identified as slouching (at 64.3, 44.13 and 46.6%, respectively). Cycling was mainly misclassified as SUM or walking (at 64.8 and 26.9%, respectively), and most of the periods when the subjects were standing in a motorized vehicle (in a bus, subway, or a tramway) were misclassified as SUM (65.4, 82.9, and 98.5%, respectively). Sitting at a desk was identified as sitting with a sensitivity of 59.9%, but misclassified as standing at 19.8%.

Tests after recalibration. After training the algorithm to recognize sitting times while in a vehicle (recalibrated model), the detection of all sitting activities (i.e., sitting and sitting in a vehicle considered altogether) was improved (see Tables 2 and 4). Specificity was high for all classes (minimum 90.9% for standing). The true positive rate of sitting on a chair was only 46%, but this time misclassification was with slouching (7.9%), sitting in a vehicle (30.9%), and standing (9.6%). Sitting in a vehicle was correctly identified with a sensitivity of 98.1%. The automatic classification worked relatively well for cycling on flat (sensitivity 75.9%), but cycling uphill and downhill proved more difficult to identify. Nonetheless, overall, the identification of cycling achieved high precision (90.9%) and specificity (99.6%). The performances in identifying lying and running were similar to those achieved with the initial laboratory-calibrated model. All walks considered together achieved a slightly lower true positive rate of 73.4%, but...
precision and specificity increased. Precision increased also for lying, slouching, sitting, standing, and cycling.

**DISCUSSION**

Previous studies have suggested that promising automatic posture and activity recognition tools developed from data acquired in highly controlled environments (i.e., in the laboratory) may not perform as well when applied to real-life data (17, 19). Results of the present work clearly support this. Indeed, the initial laboratory-calibrated algorithm performed poorly when applied to data acquired in the semi-free-living trial.

This loss of performances was induced by activities that were not present in the initial laboratory trial or not in similar conditions. Table 2 shows the performances of both algorithms when applied to semi-free-living data; the recalibrated algorithm performed substantially better than the original laboratory-calibrated classifier. Table 3 provides a detailed view of the activities recognized by both algorithms. The first 10 individuals who served as test individuals for the recalibrated algorithm were selected at random from the 10 individuals who served as calibration individuals. The last 10 individuals who served as calibration individuals were selected at random from the 10 individuals who served as test individuals for the recalibrated algorithm. Detailed results for each activity can be found in Tables 3 and 4.

Table 2. Performances of the original laboratory-calibrated classifier and of the version recalibrated with semi-free-living data, when applied to data acquired in semi-free-living conditions

<table>
<thead>
<tr>
<th>Posture</th>
<th>Laboratory Data, GDR, %</th>
<th>Semi-free-living Data, %</th>
<th>Sensitivity</th>
<th>Precision</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lying</td>
<td>93.4</td>
<td>96.6</td>
<td>100</td>
<td>83.4</td>
<td>99.4</td>
</tr>
<tr>
<td>Slouching</td>
<td>85.2</td>
<td>77.7</td>
<td>54.6</td>
<td>14.4</td>
<td>72.0</td>
</tr>
<tr>
<td>Sitting</td>
<td>52.7</td>
<td>24.9</td>
<td>95.7</td>
<td>72.4</td>
<td>85.7</td>
</tr>
<tr>
<td>Standing</td>
<td>81.3</td>
<td>2.4</td>
<td>56.5</td>
<td>13.2</td>
<td>45.2</td>
</tr>
<tr>
<td>SUM</td>
<td>79.4</td>
<td>51.6</td>
<td>14.2</td>
<td>9.0</td>
<td>9.4</td>
</tr>
<tr>
<td>Walking</td>
<td>95.4</td>
<td>76.1</td>
<td>73.4</td>
<td>84.7</td>
<td>87.4</td>
</tr>
<tr>
<td>Running</td>
<td>93.5</td>
<td>100</td>
<td>100</td>
<td>50.4</td>
<td>49.0</td>
</tr>
<tr>
<td>Cycling</td>
<td>66.3</td>
<td>5.4</td>
<td>45.9</td>
<td>29.7</td>
<td>90.9</td>
</tr>
</tbody>
</table>

N is the number of points (seconds) per activity. Lab. model, original laboratory-calibrated classifier; Recal. model, version recalibrated with semi-free-living data. Results of the initial cross-validation in laboratory conditions are also presented [reported as aggregate accuracy, or good detection rate (GDR)]. For the recalibrated algorithm, the categories “sitting” and “sitting in a vehicle” were pooled together, the same for “standing” and “standing in a vehicle,” and for “cycling” and “cycling an exercise bicycle.” To allow for direct comparison, the original laboratory-calibrated model performances were calculated only for the 10 individuals who served as test individuals for the recalibrated algorithm. Detailed results for each activity can be found in Tables 3 and 4.

Table 3. Outputs of the laboratory-calibrated classification algorithm, for each activity performed during the semi-free-living condition trial

<table>
<thead>
<tr>
<th>Detailed Activities</th>
<th>Identified As, %</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lying</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lying on a bed</td>
<td>96.62</td>
<td>1,743</td>
</tr>
<tr>
<td>Slouching</td>
<td>10.25</td>
<td>1,639</td>
</tr>
<tr>
<td>Sitting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sitting at a desk</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Sitting in a car</td>
<td>2.25</td>
<td>4.40</td>
</tr>
<tr>
<td>Sitting in the bus</td>
<td>0.00</td>
<td>10.58</td>
</tr>
<tr>
<td>Sitting in the subway</td>
<td>0.00</td>
<td>20.98</td>
</tr>
<tr>
<td>Sitting in the tramway</td>
<td>0.00</td>
<td>31.49</td>
</tr>
<tr>
<td>Standing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standing still</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Elevator down</td>
<td>0.00</td>
<td>20.00</td>
</tr>
<tr>
<td>Elevator up</td>
<td>0.00</td>
<td>18.42</td>
</tr>
<tr>
<td>Mechanical stairs down</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Mechanical stairs up</td>
<td>0.00</td>
<td>1.26</td>
</tr>
<tr>
<td>Standing in the bus</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Standing in the subway</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Standing in the tramway</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>SUM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUM</td>
<td>0.00</td>
<td>0.25</td>
</tr>
<tr>
<td>Walking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walking on flat ground</td>
<td>0.00</td>
<td>0.57</td>
</tr>
<tr>
<td>Walking carrying a bag</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Walking uphill</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Walking downhill</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Walking upstairs</td>
<td>0.00</td>
<td>1.51</td>
</tr>
<tr>
<td>Walking downstairs</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Running</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Running</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Cycling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cycling on flat</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Cycling uphill</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Cycling downhill</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

N is the total number of points per activity. To allow for direct comparison, results were calculated with only the data from the 10 individuals on whom the recalibrated algorithm (Table 4) was tested. Nos. in bold highlight a positive match between expected results and classification outputs.
conditions. In particular, sitting in a bus or a car ended up being largely misidentified as SUM instead of sitting. This was probably due to the movements induced by the vehicle motion and vibrations, which are obviously absent from the signal generated when sitting at a desk. With sitting times becoming an important criterion for estimating sedentariness (11), the inability to correctly capture the time spent in transport may lead to serious underestimations of total daily sitting times in the present study, since the annotations made by the external observer during the trial could not reach the required level of precision to record every stop-and-go movements interspersed with the current method. There may be a case in saying that the SUM class should be dropped, since interspersed cycling on flat ground; the sensitivity of downhill or uphill cycling detections remained low, even with the recalibrated algorithm.

However, perhaps the main limitation of the recalibrated algorithm was that the introduction of a “standing in vehicle” class further decreased the positive detection rate of SUM, showing that these two behaviors cannot be reliably discriminated with the current method. There may be a case in saying that the SUM class should be dropped, since interspersed ambulatory movements could be defined as a succession of standing and walking events, which are distinct classes that should be picked up by the algorithm. This was impossible to address in the present study, since the annotations made by the external observer during the trial could not reach the required level of precision to record every stop-and-go movements during SUM behaviors. This also highlights the difficulty to choose an adequate width for the time window used for extracting the signal features that define each activity class.

Moreover, cycling is usually considered difficult to capture with a single hip-worn accelerometer (13). However, these encouraging results were limited to the phases during which the subjects were cycling on flat ground; the sensitivity of downhill or uphill cycling detections remained low, even with the recalibrated algorithm.

### Table 4. Outputs of the recalibrated classification algorithm, for each activity performed by the 10 test subjects during the semi-free-living condition trial

<table>
<thead>
<tr>
<th>Activities</th>
<th>Identified As, %</th>
<th>Cycling</th>
<th>Exercise</th>
<th>Outdoor</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sitting</td>
<td>Standing</td>
<td>Trunk</td>
<td>Walking</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Lying</td>
<td>SUM</td>
<td>Running</td>
<td>SUM</td>
<td></td>
</tr>
<tr>
<td>Lying</td>
<td>Lying on a bed</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Slouching</td>
<td>10.74</td>
<td>54.61</td>
<td>0.00</td>
<td>33.50</td>
<td>0.00</td>
</tr>
<tr>
<td>Sitting</td>
<td>0.00 3.78</td>
<td>46.12</td>
<td>30.94</td>
<td>9.59</td>
<td>3.60</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>99.97</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>96.29</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>98.81</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 4.96</td>
<td>94.21</td>
<td>0.00</td>
<td>0.48</td>
<td>0.00</td>
</tr>
<tr>
<td>Standing</td>
<td>0.00 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>14.52</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>10.00</td>
<td>20.00</td>
<td>31.58</td>
<td>31.58</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>7.89</td>
<td>31.58</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.07 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.91</td>
<td>36.29</td>
</tr>
<tr>
<td>Walking</td>
<td>0.25 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>8.79</td>
<td>8.87</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>4.61</td>
<td>2.87</td>
<td>0.57</td>
<td>8.79</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>8.56</td>
<td>9.31</td>
<td>1.56</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>0.00 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Running</td>
<td>0.00 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Cycling</td>
<td>0.00 0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>0.18 3.30</td>
<td>0.00</td>
<td>5.31</td>
<td>52.36</td>
<td>5.62</td>
</tr>
</tbody>
</table>

N is the total number of points per activity. Bolded nos. highlight positive match between expected results and classification outputs.
real-time annotations, future validation studies could probably benefit from using devices enabling continuous video records throughout the trial (22) and a posteriori annotation procedures. Furthermore, with connected objects integrating accelerometers or other actuimeters (phone, watch, shoes) becoming more readily available to the public, and the quick development of crowd-sourcing techniques (20), the constitution of databases of annotated signals collected by volunteers in a wide range of everyday life situations (16) is now technically feasible (31). Our experience nevertheless shows that the level of details required for annotating the signals is difficult to achieve.

Several studies have considered the possibility of improving posture and activity detection by increasing the number of sensors worn by the subjects (15). It is indeed unrealistic to hope to capture every possible movement with a single device. Combining several sensors might, therefore, appear as a valid strategy, but, for studies in free-living conditions, this may seriously decrease compliance (6, 32). This is why a single device worn on the hip or the wrist is common practice in free-living trials, with hip-worn accelerometers still offering more reliable results (19).

In a recent study, van Hees et al. (19) quantified the errors in estimations of daily walking time resulting from the application of an algorithm “trained” with a limited number of activities to more complex daily routines. With their work, van Hees et al. provided an elegant proof of concept that study design can greatly affect the performances and evaluation of activity classifiers. The present work is complementary to this, in that it explicitly demonstrated the shortfalls of a laboratory-calibrated-only classifier when applied to data acquired outside the laboratory, but also how some of these shortfalls can be overcome by training the classifier with appropriate datasets. It, however, also showed that, as the complexity of the classifier and the learning set develops, other issues may arise, and we willingly acknowledge that even further issues may appear when moving toward total free-living conditions. It is, therefore, critical to test automatic classification tools in real life before they are used in the field. Finally, considering the complexity of human occupations and activities, intra- and interindividual variability in their execution, and practical limitations regarding the number of sensors one can bear in real-life situations (32), we believe that the idea of coming up with a comprehensive and 100% reliable classification method might be illusory. Choices and compromises will always have to be made in accordance with research questions.

ACKNOWLEDGMENT

Many thanks are expressed to Stéphanie Aubousser, Caroline Gastebois, and Ophélie Maire for help during data collection, and to all volunteers of the study.

Present addresses: T. Bastian, Université du Littoral Côte d’Opale, LOG UMR 8187, MREN, Avenue Foch, 62930 Wimereux, France. F. Gris, Movea, Grenoble, France.

DISCLOSURES

The accelerometer devices used in the study were prototypes from MOVEA SA, a company specialized in providing motion-processing technologies, including software, embeddable solutions, and semiconductor IP. E. Perrin and Y. Caritú were working with Movea at the time of the study. F. Gris was not, but is now. P. Jallón worked there briefly. However, data acquisition and performance tests presented in this article were conducted by researchers of the CRNH Rhone-Alpes (C. Simon, C. Villars, A. Maire, T. Bastian, J. Dugas), with no conflict of interest for the research.

AUTHOR CONTRIBUTIONS


REFERENCES


