Validated Caloric Expenditure Estimation using a Single Body-Worn Sensor

Jonathan Lester, Carl Hartung, Laura Pina, Ryan Libby, Gaetano Borriello
Computer Science and Engineering
University of Washington
Seattle, Washington, USA
{jlester,chartung,lpina,libby, gaetano}@cs.washington.edu

Glen Duncan
Epidemiology and Nutritional Sciences
University of Washington
Seattle, Washington, USA
duncag@u.washington.edu

ABSTRACT
In 2007, approximately 30% of US adults were obese, with related health care costs exceeding 100 billion dollars. Clearly, the obesity epidemic represents a growing societal concern. One challenge in weight control is the difficulty of tracking food calories consumed and calories expended by activity. This paper presents a system for automatic monitoring of calories expended using a single body-worn accelerometer. Our system uses activity inference combined with signal analysis to estimate calories expended in real-time using regression formulas developed by the American College of Sports Medicine. To validate our system, we have collected data from 51 subjects in a laboratory setting using a treadmill and a more natural field test. Actual caloric expenditure was determined using the medical "gold standard" measurement, of oxygen consumption. We are able to achieve 89% accuracy with lab data and 79% with field data – both high enough to be useful in practice.

Author Keywords
pervasive health, long term health monitoring, wellness, caloric balance

General Terms
Human Factors, Measurement, Verification

ACM Classification Keywords
H5.m. Information interfaces and presentation: Miscellaneous.

INTRODUCTION
The United States and many other industrialized countries currently face an epidemic of obesity. An overabundance of highly palatable, energy-dense foods, and a lack of regular exercise has resulted in 35% of the US population being overweight, and the percentage of obese adults has doubled to 30% over the past decade [2]. Often viewed as a US problem, the World Health Organization estimates that there are more than one billion overweight adults in the world; 300 million of whom are actually obese [1]. In addition to social and lifestyle effects, obesity is associated with increased risk of chronic and fatal diseases such as heart disease, stroke, some forms of cancer, type 2 diabetes, and hypertension. The costs for these diseases in terms of human life and medical expenses are staggering. In 1998 it was estimated that 9.1% of all US medical expenses, $78.5 billion dollars, was spent on overweight/obesity attributed expenses.

While there is no single solution to the obesity epidemic, research has shown that overweight/obese individuals tend to underestimate the calories they consume [3,4] and overestimate the calories they burn [5]. This is, of course, the worst possible error; people consume more calories and burn fewer calories than they believe, and remain overweight. This misconception is far from the only reason for the obesity epidemic; however, a better idea of the calories consumed and expended by a person would be a valuable piece of information for individuals to monitor their energy balance, as well as for doctors and other caregivers. With our busy lifestyle, it is difficult to imagine that most people would, or are even capable of, accurately and continuously tracking their eating and exercise habits. Systems which can automatically track and provide a user with feedback about their energy balance might have the potential to help obese, overweight, and even normal weight individuals.

To create such a system, we are working on two projects, funded by the National Institutes of Health, to develop an up-to-the-minute display of a user's caloric balance: what they've eaten versus what they've burned through physical activity. There are two components to our energy balance monitoring system: 1) an interface used to enter foods consumed and 2) the system presented here, algorithms to automatically compute caloric expenditure. For some preliminary details regarding food entry and some of our initial lab tests results please see [6]. In this paper we present a system for computing an estimate of the calories expended using a wearable system and its validation against the medical gold standard measurement. While the eventual end users of our system are obese and overweight individuals, the current target audience is primarily medical researchers and practitioners. These researchers are not amenable to overly complex algorithms or "black box" systems. Rather than use more complex models, which may perform better, we use more standardized algorithms and techniques to make our system more understandable, useful, and more likely to be adopted by this audience.

RELATED WORK

The area of fitness and exercise is filled with a variety of research projects [19] and commercial systems. Several commercial products exist which can measure calories expended during exercise. The most basic systems are those found on gym equipment. There have also been several portable systems for measuring more free living conditions. One popular device among researchers has been the RT3 accelerometer, a wearable accelerometer which computes activity counts that can approximate the energy level of the activity being performed by the user based on established cut-points [18]. However, the coarse grain output (often minute long averages of acceleration) of the RT3 means that it is impossible to gain detailed insight into the behavior of individuals, such as activity type and location, among others. The Philips Respironics' Actical is a similar device which can compute activity counts as well as an estimate of total energy expenditure [1].

In addition to more clinical devices; several casual systems have been released, including the Nike+ system, a shoe worn piezoelectric sensor that communicates wirelessly with a receiver plugged into an Apple iPod [8]. The Nike+ system tracks the distance, pace, estimates the calories burned during a run, and provides users with audio feedback while running.

Nokia provides two systems for activity monitoring: Activity Monitor and Sports Tracker [9, 10]. Sports Tracker is a GPS based system that uses either the phones built-in GPS or a Bluetooth GPS device to track users' trips. Users are provided with maps overlaid with their GPS tracks; information about the distance traveled, average speed, and duration is provided in real time. Activity Monitor uses the built-in accelerometer on Nokia's N95 and N82 phones to count the number of steps taken by a user. It provides tracking information for users recording the number of steps taken per day, providing an estimate of the calories consumed and the distance traveled by a user. Both the Nike+ and Nokia systems are closed platforms which use fairly simplistic methods for estimating calories without any extensive validation.

The SenseWear/BodyBugg system by SenseMedia is a wearable armband-based system equipped with a dual-axis accelerometer, skin temperature, Galvanic Skin Response, and heat flux sensors [7]. While the SenseWear system has been used in several experiments its cost and closed black box nature prevents it from widespread use or integration into other systems.

Choi et al. [11] conducted a similar laboratory study to ours, in which they collected wireless accelerometer data from 94 subjects walking on a treadmill hooked up to a portable $\dot{V}O_2$ gas analyzer to measure total energy expenditure. Accelerometer data was then filtered and integrated and used as an input into a regression analysis to match their accelerometer output with the energy expenditure computed by the gas analyzer. The main purpose of their study was to validate accelerometer readings, based on correlation with energy expenditure measured by $\dot{V}O_2$. The study consisted of a controlled laboratory experiment tested on primarily fit college-age students. The main differences between Choi *et al.* and the results presented here are that we used a wider variety of test subjects (ages/body types), are able to classify more natural activities using a field experiment, and that we use accepted algorithms to estimate energy expenditure instead of simply modeling the data we've collected.

All of these devices are closed systems, they cannot be modified or scrutinized, and do not provide extensive proof by validation. While many of these systems are appropriate for the casual user, we are focused on developing a system useful both to researchers, medical doctors, and practitioners alike, many of whom have qualms—perceived and real—about the utility, operation, and reliability of these commercial systems.

OVERVIEW

In this paper we first discuss the experimental design, algorithms used for computing caloric expenditure, step counting, and approach for estimating walking grade. We then present three sets of results, results from our laboratory experiments, field data collections, and a synthetic average day's worth of data. We are able to achieve 89% accuracy with lab data and 79% with field data.

EXPERIMENT DESIGN

While we could take one of the systems described in the related work section, we do not have a quantifiable understanding of how accurate these systems are or have access to their methodology. Researchers who wish to use the system need to have confidence that the system uses a reasonable approach and works well enough for them to conduct their research.

System Setup

Our system consists of a body-worn 3-axis accelerometer. The primary source of data for our experiments is the University of Washington/Intel Mobile Sensing Platform (MSP) [15], which is worn on the waist of subjects during data collection. As one of the goals of our project is to create a system which can be carried around by users throughout the day, we would eventually like to be able to use a cell phone as sole platform. To this end, we also collected a subset of accelerometer data from the Apple iPhone and the Nokia N95, in addition to the MSP data. While cell phones have many nice affordances, they still lack some sensors which we find useful, such as barometric pressure sensors. Their main limitation is their inability to collect and process data throughout an entire day (at least a 12 hour time span). However, in the future these issues are expected to be less of a problem.

Computing Caloric Expenditure

Caloric expenditure is calculated using the algorithms provided by the American College of Sports Medicine (ACSM) [14]. These algorithms, described in detail in a later section, require four pieces of information: activity type being performed, subject weight, speed of movement (horizontal), and the grade at which they are moving (vertical). Except for subject weight, the other pieces of information are provided by the sensors.

Validating Caloric Estimates

To understand the performance of our system and to provide evidence to medical practitioners that the system provides reasonable and reliable outputs, we need to test our system against the "gold standard" of energy expenditure. Ground truth is obtained using measures of oxygen consumption, $\dot{V}O_2$, using either a stationary or mobile metabolic measurement system.

Experiment Setup

Our sensors will provide us with the information needed to compute the caloric expenditure, and $\dot{V}O_2$ will provide us with the ground truth caloric expenditure. We collected data in both a lab and a field setting. Each subject was asked to perform one field and one lab session, with a third session randomly assigned to be either lab or field to ensure repeatability. Subjects were screened for any health issues which might prevent them from participating, demographic data was collected, and they were weighed and had their height measured at their first session.

For both the field and lab data collections, each subject wore a MSP on their waist, Polar heart monitor, Actical data logger, a $\dot{V}O_2$ mask or data collection system, and optionally a fanny pack containing an iPhone and N95. If subjects heart rates exceeded 85% of their computed maximum (based on age) the experiment was stopped. Several subjects were unable to complete all stages of the laboratory experiment because the metabolic demand exceeded their capacity.

Laboratory Data Collection

Participants walked and ran on a treadmill for the lab sessions. $\dot{V}O_2$ data was collected from a bench mounted stationary system, a metabolic cart, shown in Figure 3 (top). Table 1 shows the organization of the lab data collection. Each person completed 9 stages, the first two stages sitting/standing stages and the final sitting stages were always the same. Subjects performed the walking and running stages in a predetermined order with a random starting stage to help eliminate transitional effects in the data. During the experiment they recorded the time when the user switched between stages and set the appropriate speed/grade settings on the treadmill.

Stage	1	2	3	4	5	6	7	8	9
MPH	Sit	Stand	1.8	2.5	3	3.5	4.5	4.5	Sit
Grade	n/a	n/a	0%	5.0%	0%	7.5%	0%	2.5%	n/a
Duration	3 min	2 min	3 min	3 min	3 min	3 min	3 min	3 min	5 min
			Walking				Running		

Table 1. Laboratory data collection stages, each subject started and ended with a sitting stage. While the six walking and running stages were always executed in a random order.

Field Data Collection

Field data was collected to capture more realistic real world conditions. Subjects were asked to perform a scripted set of common day-to-day activities, with each session lasting about 15 minutes. Two experimenters conducted each field data collection. One would guide the subject and monitor them, while the other would record the ground truth activities and times on a timesheet. The script for field data was:

- Start data collection sitting down for several minutes on the 5th floor
- Stand and walk to the elevator
- Go down and walk through the atrium to get outside
- Walk down a small flight of ~4 stairs and then walk around an inclined area and up a set of steps
- Pick up a heavy object (~20lb) and walk around with it
- Place the object down and sweep the ground for approximately a minute
- Walk back inside to the elevator and up to the 5th floor
- Walk back to the setup area and sit for several minutes

Analysis and Real Time Classification

Although the results presented in this paper are from offline analysis, the MSP system runs in real time. Offline analysis is presented here owing to the need to align data to the recorded $\dot{V}O_2$ data and ground truth annotations. The differences between the offline/desktop version and the online version of the system are negligible.

PARTICIPANT STATISTICS

In order to ensure that our system functions across as wide a range of the population as possible, we carefully recruited subjects from a large representative group of ages, genders, and body types. In particular we were interested in gathering data from older and obese individuals as they are often left out of such studies because many of them will have difficulty completing the experiment. To date we have collected data from 51 subjects, of varying age, race/ethnicity, and body type (i.e., normal weight, overweight, and obese); 28 are female and 23 are male. Figure 1 illustrates the range of ages and genders who have participated in the study while Figure 2 shows the distribution of BMI across the subjects. The current lab data set contains 28 hours of laboratory data and 16 hours of field data.

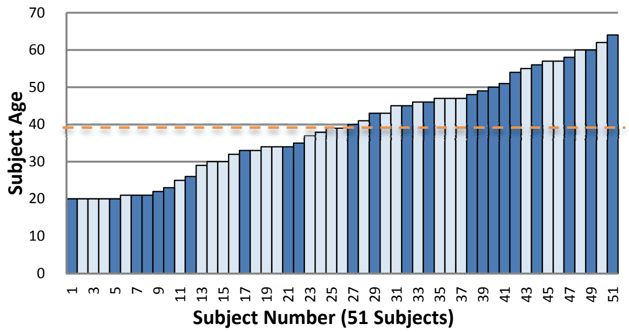

Figure 1. 51 subjects participated in the study; the average age is 39.4 years (std. dev. 13.28 years). Darker colored bars are male subjects (n=23, 45%) while lighter colored bars are female subjects (n=28, 55%).

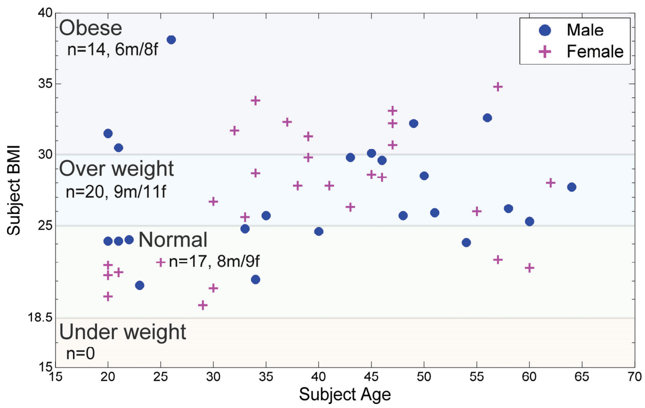

Figure 2. Subjects' age/gender displayed versus their Body Mass Index (BMI).

$\dot{V}O_2$ MEASUREMENTS FOR CALORIC EXPENDITURE

To measure the actual calories burned by a human body it would be necessary to measure the energy stored and consumed by every cell in the body. Unfortunately, it is not possible to measure at this granularity, instead we estimate metabolic rate by measuring by-products of respiration. Our bodies consume oxygen and produce carbon dioxide to metabolize energy. $\dot{V}O_2$ is a measure of the rate of oxygen consumption and is related to the rate of carbon dioxide production by the chemical reaction of cellular respiration during aerobic metabolism.

$\dot{V}O_2$ allows us to measure the energy consumed by having the subject wear a mask through which they breathe. The system measures the flow, O_2 content, and CO_2 content of the inspired and expired gases. With this data we can estimate how much O_2 the body has consumed and how much CO_2 it has produced, and so infer the metabolic rate. $\dot{V}O_2$ is the most readily available and reliable measure of caloric expenditure. It provides an accurate measure because our bodies primarily provide energy through aerobic metabolism and have very little oxygen storage. Figure 3 contains images of the stationary and portable field $\dot{V}O_2$ mask units we use in our work, along with an image of the MSP.

High-intensity activities can trigger the body to exceed its anaerobic threshold at which point the body can no longer produce enough energy aerobically and begins producing energy anaerobically. Anaerobic metabolism confounds our $\dot{V}O_2$ estimate because it does not require O_2 nor produce CO_2. To avoid these complications, steady-state activities lasting longer than a minute are required and the subject's heart rate is kept below 80-90% of the maximum heart rate.

Figure 3. (left) Subject wearing $\dot{V}O_2$ mask connected to a stationary metabolic system; nose clips ensure that all breathing is through the mouthpiece. (right) MSP, inside a black plastic case, clipped onto the waist. (bottom) Portable field $\dot{V}O_2$ unit. Subjects wear a face mask that is secured tightly against the skin to prevent leaks. The metabolic gas analyzer is worn around the chest.

Due to the fact that we have to capture all air exhaled and inhaled, the breathing devices worn during $\dot{V}O_2$ must be as airtight as possible. As a result they are uncomfortable to wear for long periods of time. Due to this discomfort, and the battery limitations on portable systems, it is not possible or practical to collect $\dot{V}O_2$ for more than a few hours.

ACSM CALORIC ESTIMATION FORMULAS

While $\dot{V}O_2$ provides a very accurate measure of the calories burned during exercise, $\dot{V}O_2$ equipment is not a practical way to measure caloric expenditure for day to day settings. To estimate the calories burned, regression equations have been developed to model the caloric expenditure of humans performing various activities [12, 13]. We use the equations adopted by the American College of Sports Medicine (ACSM) as they are the standard medical measure for caloric expenditure [14]. The ACSM measure caloric expenditure using the basic formula:

$$\dot{V}O_2 = R + H + V$$

where R is *Resting Metabolic Rate*
H is *Horizontal Component*
V is *Vertical Component*

At rest, a person's metabolic rate is defined as one metabolic equivalent (a MET) and is dependent on their body weight. The resting metabolic rate is computed as:

$$R = 3.5ml * kg^{-1} * min^{-1}$$
$$Calories = \dot{V}O_2 \frac{L}{min} * 5.01 \frac{kcal}{L} = \frac{kcal}{min}$$

Calories (kcal) are computed by using the conversion factor 5.01kcal/liter, which assumes that all calories are derived from burning carbohydrates. For non-resting activities an activity specific formula is used, which add an additional horizontal and vertical component:

Walking:

$$H = 0.1 * speed\left(in \frac{m}{min}\right) ml * kg^{-1} * min^{-1}$$
$$V = 1.8 * speed\left(in \frac{m}{min}\right) ml * kg^{-1} * min^{-1} * grade$$

Running:

$$H = 0.2 * speed\left(in \frac{m}{min}\right) ml * kg^{-1} * min^{-1}$$
$$V = 0.9 * speed\left(in \frac{m}{min}\right) ml * kg^{-1} * min^{-1} * grade$$

The horizontal component provides the energy expended based upon the subjects weight and speed, while the vertical component accounts for the additional effort required to move up an incline, and depends on weight, speed, and the grade of the incline. An average 5'10" person with a normal BMI would weigh around 160lb (73kg). At rest they would consume 1.28kcal/min, when walking on level terrain at 3MPH would consume 4.2kcal/min, and running at 5MPH 7.2kcal/min. An incline of 5% would increase this to 6.9kcal/min and 8.5kcal/min respectively.

COMPUTING CALORIC EXPENDITURE

To use the ACSM algorithms to compute caloric expenditure we need the following four pieces of information: subject weight, the type or mode of activity being performed (rest, walking, running, cycling, or stair stepping), subject speed, and subject grade. We can measure the subject's weight, and the other three pieces of information we can obtain from our sensor data. The following sections discuss our techniques for activity inference, a step counter to infer speed, and grade estimation techniques.

Activity Inference

We need to know the activity the user is performing to apply the correct ACSM algorithm. While there are a number of different classification techniques and activities which could be inferred, in this case, we're interested in the primary activities for which we have equations (rest, walking, and running). Because we are using the MSP platform, we use a Naive Bayes classifier, as the MSP provides a real time inference library for Naive Bayes classification.

On the MSP the accelerometer is sampled at 512Hz and features are computed at 4Hz, on cell phone platforms the sample rate is typically around 30Hz (100Hz for the iPhone) and the feature rate is 1Hz. The features being used are based on the magnitude of the 3D acceleration vector: the sum of one second FFTs using bins of 2-4Hz, 1-10Hz, variance, standard deviation, range, and a step speed estimate from our step counter (discussed in the next section).

	Precision	Recall	Accuracy
Lab: Leave one out	98.41%	96.64%	99.06%
Lab: 10-fold cross validation	96.67%	98.31%	98.73%
Field: Leave one out	95.82%	95.97%	95.97%
Field: 10-fold cross validation	96.14%	96.06%	96.06%

Table 2. Classification accuracies for the leave-one-out and 10-foled cross validation. Subjects were not split across folds in cross validation. In leave-one-out all of a subject's data was held for testing and not used during training.

The Naive Bayes classifier provides good discriminative classification (is able to separate classes well) and can be improved using a simple smoothing filter or a HMM to create smoother transitions. In our system we actually do not use additional smoothing as the instantaneous classification is used to generate caloric estimates, which will then be smoothed. Pre-smoothing of activities is often not beneficial and in some cases detrimental as it can miss short duration activities or inflate the duration of longer activities.

Classification Accuracy

Two classifiers are used in this system: one trained for the lab data and one for the field data. We could simply train one classifier off of the lab (or field) data and use that for both – the lab classifier achieves 94% accuracy on the field data. We use two classifiers to avoid any bias between the datasets. Ten folded cross validation was performed on the lab/field data sets, where each subject's data was placed all within one of the ten folds. We also performed leave-one-out subject training. The activity classification results used in this paper all come from leave-one-out training.

Table 2 lists the precision, recall, and accuracy numbers for these experiments. The classification accuracy is quite high, the lab data is quite easy to classify because it is so well structured. Some of the field data results are likely a bit optimistic simply because some of the more difficult to label activities have no ground truth. For example, if the subject stops to tie their shoe the observer will annotate this as 'null' and the data isn't used for training or testing a classifier. It is, however, all used during caloric expenditure computation because we need to classify from the start of the session to the end. In this case we use the classification output that best explains the data.

Step Counting

Once we have the activity the user is performing the next piece of information we need is the speed at which they are walking or running. To get from the accelerometer signal to walking/running speed we can use a number of simple zero crossing or peak detection algorithms to detect steps and then use the subject's stride length (determined by their

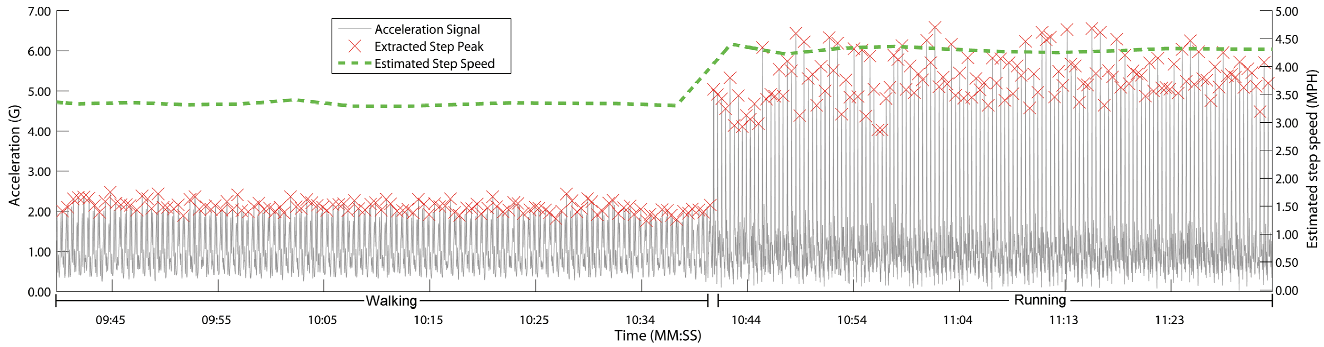

Figure 4. Example accelerometer trace with the magnitude of the accelerometer shown in solid line at the bottom, extracted footfall peaks are marked with X's, and the estimated speed from the steps is shown as a dotted line. During the first half of the trace the subject was walking at around 3MPH and then began running at 4.5MPH near the 10:44 mark

height) to compute speed. While it's easy to get one that works in one set of circumstances it is difficult to get an algorithm to work across all walking/running speeds, ages/body types, be immune to bumps or noise in the signal, not require parameter tuning, and yet be computationally simple enough to be implemented in real time.

Yang et al. [16] implemented a step counter for a dual-axis accelerometer worn on the foot by using three methods, Pan-Tompkins method, template matching, and peak detection based on combined dual-axis signals. Template matching proved to be too computationally heavyweight and the peak detection method was not appropriate because it made assumptions about the signals seen on the foot. The Pan-Tompkins method was implemented with several modifications; however, for a hip worn accelerometer it was too sensitive to noise in the signal and would produce spurious footfalls. The largest problem with this method was an appropriate low pass filter that could remove noise without distorting the underlying low frequency signal. Due to the differences in individual walking gates and differences in how the sensor is worn we wanted to avoid over-specifying parameters or requiring per user tuning. Instead we developed a method that uses an adaptive FFT energy based filter, which extracts steps by selecting the lowest frequency bins that account for 10% of the energy in the FFT output. This energy threshold allows us to robustly use the same method for a variety of walking conditions, users, sensor positions, loose or tight clothing, and terrains.

Example Step Detection Output and Accuracy

The algorithm used to extract footfalls from the acceleration magnitude is shown in the "Step Computation" side bar. Figure 4 displays an example trace of a user walking on a treadmill at 3MPH for the first half who then begins running at 4.5MPH near the 10:44 mark. There are many steps/footfalls detected in a relatively short amount of time and no obvious errors.

For normal walking speeds, the step detector achieves 91% accuracy (std. dev. 8.1%) and 90% (std. dev. 8.9%) for running. In general the estimated step speeds are stable and when there are errors, the error is the estimate being either too high or too low. One source of error in our estimate is stride length. Our system assumes people are walking at their natural stride length; however, walking at slow speeds, such as the 1.8MPH segment in the lab treadmill setup, causes most individuals to walk at a non-natural stride. Even though we cannot estimate stride length without knowing the distance traveled by a person; it is possible to learn an individual calibration for each user's walking pattern. However, it is not practical to calibrate every individual user and this process only increases our step speed accuracy by around 4%, and is not used in this paper.

Step Computation

- Compute magnitude of the accelerometer: magnitude = $\sqrt{X^2 + Y^2 + Z^2}$
- Compute the FFT using a 2 second window of data.
- Count the FFT's output bins, lowest first, until we have 10% of the total FFT energy (the sum of the FFT). Bins that are outside this range are set to 0:

```
threshold = sum(FFT_output)*10%;
curEnergy = 0;
for( cur=1 ; cur<NFFTLength ; cur++ )
      curEnergy += FFT_output[cur];
      if( curEnergy >= threshold )
            FFT_output[cur] = 0;
      end
end
```

- Next take the inverse FFT of the modified output to recover filtered acceleration windows. This signal is an energy based low pass filtered version of the magnitude.
- Construct a new filtered signal using 50% overlapping triangular windows. And then compute the derivative of this signal using Yang *et al.*'s derivative function
$$D(n) = \big(2x(n) + x(n-1) - x(n-3) - 2x(n-4)\big)$$
- Step intervals are found by looking for zero crossings in the derivative (D). An average of steps per second is computed using the average step interval for the last 5 steps.
- Step speed is then computed using
$$speed = K * height * avg\_stride\_per\_second$$
  - The constant K is selected by the gender of the subject 0.415 for males and 0.413 for females. Height is subject height in centimeters.

Grade/Slope Computation

The final input we need to our equations is slope, this information serves two important purposes. First it improves the overall accuracy of the system. Secondly, and more importantly, rewards users for more strenuous activities. Previous work [17] has shown how important it is for users to feel that they are appropriately rewarded by the system. If walking up a steep hill provides the same caloric expenditure estimate as walking along a flat street users will be understandably frustrated with the system, feel cheated, and have little motivation to expend additional effort.

Grade information is required by the ACSM algorithms to compute the additional workload required to move the body up against the force of gravity. They have not, however, been validated for walking down declines, mainly because there are very few treadmills capable of simulating declines. While our system currently uses the accelerometer as the primary data source we also collected barometric pressure and GPS information. Combinations of these two sensors plus additional data sources can be used to provide our system with an estimate of indoor and outdoor grade estimation. Barometric pressure can be converted into an elevation estimate, while barometric pressure varies and the elevation estimate may be off +/- an offset value the shape it traces out will be a fairly accurate measure of elevation. To get an idea of the grade of slope we are walking on we need two parameters: the rise (the change in elevation) and the run (the distance covered). Using these sensors our system has four methods available for computing the grade/slope of a surface.

Barometric Pressure. Using the barometric pressure sensor on its own we can estimate elevation. Combining this with a simple form of dead reckoning using activity-inference and step speed estimates, we can obtain an estimate of the distance traveled. This method has the advantage of being the lowest power consumer, the least data intensive solution.

Barometric Pressure + GPS. When GPS information is available, we can use the distance traveled with the barometric elevation estimate to estimate the terrain profile. We have had mixed results cleaning up GPS readings using accelerometer and GPS readings with a Kalman filter; however, the GPS readings can be used to generate a reasonable estimate of distance traveled.

GPS Only. GPS devices are capable of estimating their position in a 3D space when they have a good lock on four or more satellites. Depending on the quality of the signal and the device, the GPS position and the altitude estimate can be used to estimate the grade. GPS only solutions are appealing to mobile phone platforms where no barometer is currently available.

GPS + GIS Sources. Several GIS (Geographic Information Systems) surface elevation profiles are available. The USGS-DEM (US Geological Survey – Digital Elevation Model) covers most of the US with 15meter accuracy.

	Accuracy	Improvement
Original Estimate	86.29%	-
Barometer Only	95.88%	9.59%
Barometer + GPS	95.25%	8.96%
GPS + USGS DEM	95.65%	9.36%
GPS + LIDAR Scans	88.61%	2.32%
GPS Only	92.98%	6.69%

Table 3. Improvement in caloric estimate using the different sources of grade estimation.

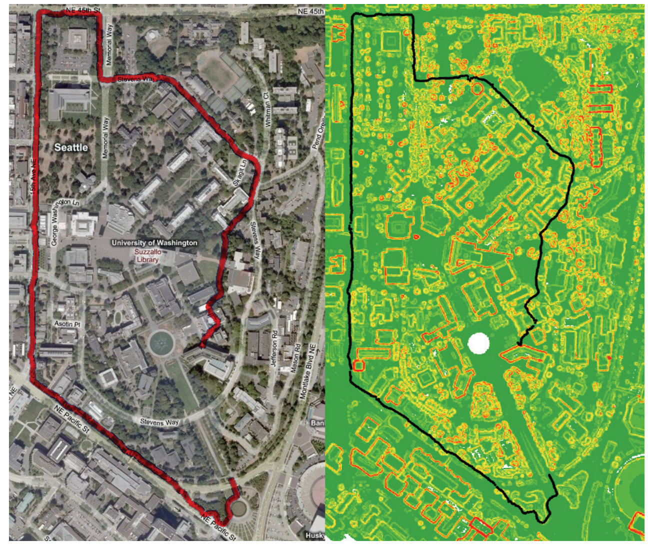

Figure 5. (left) A GPS trace plotted on top of a satellite view. (right) A LIDAR slope contour plot of the same area, green areas have very little slope while red areas (like the edges of buildings) have large slopes.

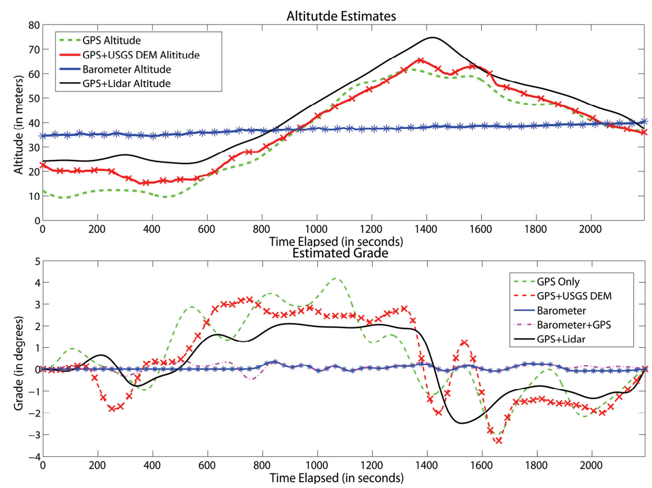

Figure 6. (top) A plot of the estimated altitudes from GPS, Barometric, and GIS data sources. Note, that the Barometric pressure had an offset added to better illustrate how they all produce reasonable elevation profiles. (bottom) A plot of the estimated grades the person in Figure 6 walked on.

Additionally, most metropolitan areas have LIDAR (air-based laser range finder scans) which provide high resolution elevation profiles accurate to within 30cm. In this case, higher elevation accuracy does not necessarily mean better grade estimation. For example, if you walk next to a building, your GPS position could get misreported as being on top of the building adding noise to your measurement, where as DEM data is more coarse grain but smoother. Figure 6, shows a GPS trace plotted on top of a satellite image, and to the right shows the same trace plotted on top of a LIDAR slope contour plot of the area.

While slope information is not a requirement for a functional system, it provides a necessary enhancement to

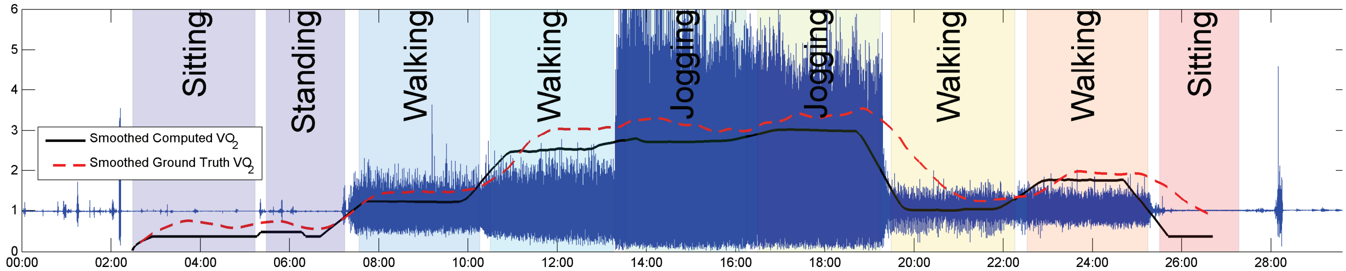

Figure 7. A plot of the accelerometer magnitude (solid waveform) versus the smoothed ground truth $\dot{V}O_2$, dotted line, reported by the metabolic cart for an example laboratory data trace. The smoothed computed $\dot{V}O_2$ is shown as a solid dark line. As you can see the computed and estimated $\dot{V}O_2$ match fairly well for most of the trace. However, after the subject has jogged for a few minutes there is a noticeable cool down visible in the ground truth data that is not captured by our computed estimate. The ground truth activities are shown as shaded bars behind the plotted data, containing all 9 stages listed in Table 1.

both increase accuracy and to capture important aspects of a users' normal daily activity. The methods described above can be used to increase the accuracy of the caloric expenditure calculation. Using the sample shown in Figure 5 and the computed slopes shown in Figure 6, we can improve our original caloric estimate, by 5-10%, as shown in Table 3, and help to alleviate user frustration with the system.

CALORIC EXPENDITURE ON MOBILE PHONES

The emergence of mobile phones like the Apple iPhone, Nokia N95, HTC Fuze, and Android G1, which all provide software interfaces to on-board accelerometers opens up the possibility of using these devices to develop software which can enable commodity devices to replace custom sensing platforms like the MSP. We have collected around 30-50 field and lab sessions with an iPhone and/or an N95 recording accelerometer data. Using this accelerometer data in place of the MSP data our accuracy only changes by around +/-3%, meaning that these mobile phones are capable of replacing the MSP accelerometer sensor and allowing us to deploy our system on both custom platforms and mobile phones. However, one of the largest obstacles to using commodity phones as our sole platform is the issue of battery life and software control. On the iPhone, we are unable to run our software in the background and our battery life is limited to 2 hours with GPS on and around 6 hours when it is off. By contrast the MSP will run for approximately 12 hours running its activity inference software (using an additional battery this can be increased to 24 hours). The real time caloric expenditure calculations only increase the power consumption by around 6%, reducing our battery life by only around a half-hour.

RESULTS

The data from our experiments was analyzed by aligning the sensor data with the ground truth and $\dot{V}O_2$ data. We then ran our data through our system, classifying the activity to determine whether the subject was sitting/standing, walking, or jogging. Then depending upon the activity we use the corresponding equation. Step speeds are provided by the step counter. For lab tests we used the provided treadmill slope and for field data we didn't use any slope information. Weight is used for the resting metabolic rate while height and gender are used to compute stride length, as part of the computation for walking speed.

The output from the ground truth and our computed caloric expenditure was then smoothed over one minute windows and the error computed as shown in the following equation.

$$Error = \frac{|\sum computed\,VO_2 - \sum ground\ truth\,VO_2|}{\sum ground\ truth\,VO_2} \times 100$$

Laboratory Data Results

An example trace is shown in Figure 8 where the labeled activities in the background denote the different stages of the experiment, the accelerometer magnitude, estimated step speed, and the estimated $\dot{V}O_2$ and computed $\dot{V}O_2$ are shown. As you can see from the solid and dotted lines, the estimated $\dot{V}O_2$ does follow the actual $\dot{V}O_2$ expenditure fairly closely. However, it does not completely follow the ground truth $\dot{V}O_2$. At around the 20 minute mark the dotted ground truth $\dot{V}O_2$ line stays higher than the solid estimated value. This is due to the "warm up" effect, called excess post-exercise oxygen consumption. Our bodies cannot go from strenuous activity back to a resting state instantaneously; instead they take several minutes to "cool down". This is not taken into account in the ACSM algorithms because they are designed to approximate steady state and not transitions such as these. However, because there is usually an associated ramp up delay the two effects tend to mostly cancel out for the laboratory tests, allowing us to estimate the caloric expenditure with 89.52% accuracy (std. dev. 7.25%).

Fields Data Results

To analyze the field data we used the same system employed in the laboratory setting, the only differences being that we used a classifier trained using leave-one-out field data instead of lab data, and that we had no ground truth grade so we did not provide it to our system. An example plot of the computed $\dot{V}O_2$ versus the ground truth $\dot{V}O_2$ is shown in Figure 7. As you can see the solid line (estimated calories) does follow the dotted line fairly well until the 21 minute mark when the dotted line increases quite a bit. This error is caused by the people picking up a heavy object and walking around with it (a 20lb box). Because the system cannot determine when a person is

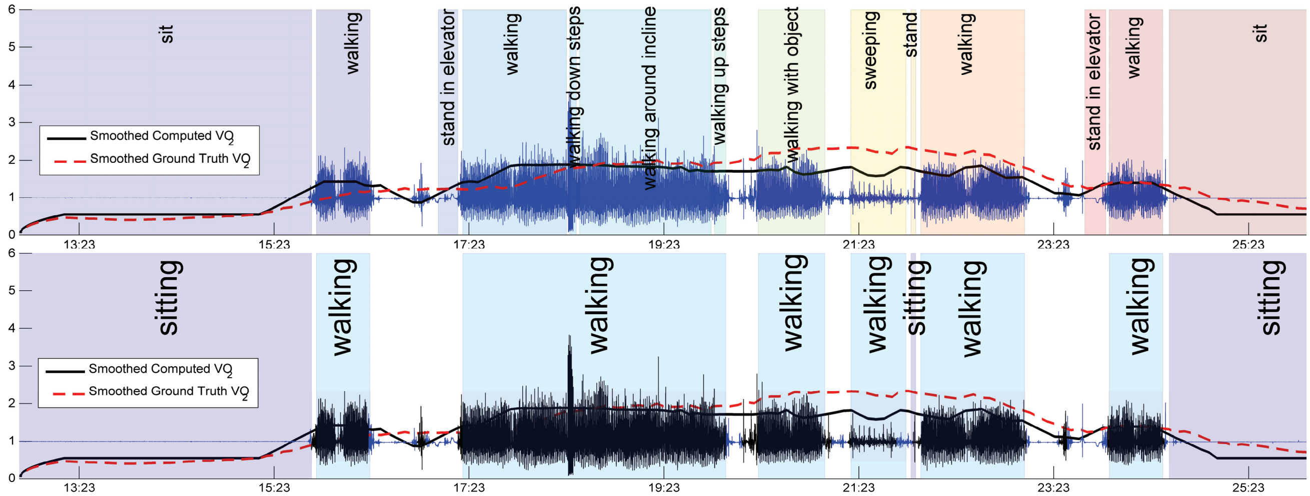

Figure 8. (top) Example field data with the ground truth labels, as provided by the experimeters. (bottom) The same trace with the re-labeled activities listed as either sitting (still) or walking (moving) activities. The darkend portions of the acceleroemter trace are those infered by the activity inference as being walking. As you can see there are some instances where the ground truth labels are somewhat misleading and there are portions where the inference could be improved to better infer the activities.

carrying an object or the mass of that object it cannot model this change. However, despite this we are still able to model the caloric expenditure in the field data with 79.8% accuracy (std. dev. 9.48%). While this is lower than the lab data, if we computed accuracy instantaneously across the entire trace, rather than looking at the sum total, the accuracy increases to 84.9%. Most of the error comes from the warm up effect from the extra strain induced by carrying the box, which the sensors can't pick up, if we ignore the $\dot{V}O_2$ data after the subject picked up the box our accuracy is 87.8% for the field data. As in both the lab and field data the system underestimates caloric expenditure, and does so for all test traces we have collected.

A SYNTHETIC DAY

Laboratory data gives us a very controlled environment with which to test our algorithms; and field data gives us a more natural idea of what would happen in real life. However, while the field data is a more natural representation it doesn't capture the ratio at which activities occur in everyday life. Intuitively, because energetic activities like running consume more calories they seem like a natural focus for our efforts. However, this fails to capture the fact that nearly all of our day is spent engaged in sedentary activities, such as sitting at a desk, broken up with brief periods of walking, and some longer journeys. While an average 160lb adult would consume three times more calories walking than sitting, they are likely to spend less than two hours walking while spending 20 hours or more in sedentary activities. While we are currently unable to measure $\dot{V}O_2$ for an entire day; we've constructed a synthetic day consisting of one subject's data stitched together to form a 24 hour period to better understand the portion of calories consumed by resting metabolism. Our synthetic day might look like the following scenario:

- Sleep until 8AM, go for a half hour jog around their neighborhood before returning to get ready for work
- Walk about half a mile to the bus stop and go to work, where they spend most of their time sitting at their desk making relatively few short walking trips to other co-workers offices and down to the lunch room
- After work they, catch a ride to watch a movie with friends, before walking to a restaurant, eating dinner, and then heading home, where they do a bit of work and watch TV before heading off to bed.

As illustrated in Figure 9, our synthetic person burned 2,500 calories whereas our system estimated around 2,010 calories (80% accuracy). 320 of these calories were burned during the morning run. However, the majority of their calories were burned sitting, 1200, and sleeping, 620; this is counterintuitive to most people's expectations of where we burn the most calories. It may be advantageous to have a system capable of recognizing a very diverse set of activities; however, modeling the most common case, sitting, and a few very common activities has the most impact. While exercise is clearly beneficial, it's important to understand that most of the time the system will have to model the resting rate accurately because we spend so much of our time sitting and working. The reason we focus so much on activities rather than rest is that people can increase their activities to burn more calories and people have an expectation to get credit for their exercise. And we have little control over our resting metabolic rate.

CONCLUSION

The caloric estimation system and the data we have collected have shown that by using the standard ACSM algorithms, we can achieve good results from laboratory (89% accuracy) and field data (79% accuracy). We have shown that mobile phones like the iPhone and N95 are capable of performing the same computations as we

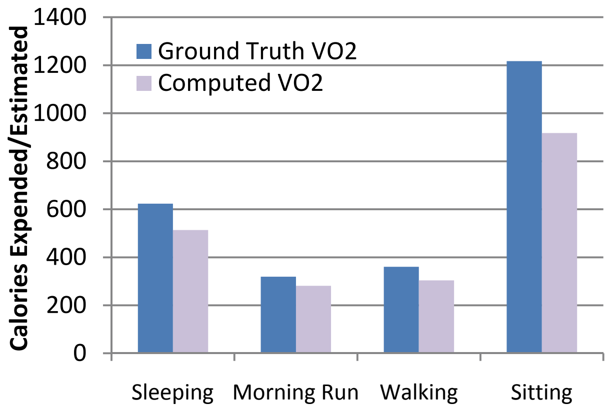

Figure 9. Histogram showing the calories burned during a hypothetical data created using a synthetic data set from one of our test subjects. The vast majority of calories and time are spent in sedentary activities, sitting and sleeping.

currently use with their own accelerometers and very nearly the same accuracy (+/-3%), opening up the possibility of widespread user studies or use of our system. We have also shown how the system can use GPS, GIS, and barometric pressure data to enhance the performance of our caloric estimates by 5-10% during more natural free-living conditions. As the system currently underestimates caloric expenditure, there are improvements which can be made to the system; however, doing so adds complexity and additional points of failure. In its current form the system is useful to medical practitioners and researchers, and provides them with a system which is understandable, open, and verified against a trusted standard.

ACKNOWLEDGMENTS

This work is funded in part by NIH R21AG032232 (Duncan, GE), BALANCE: Bioengineering Approaches for Lifestyle Activity and Nutrition Continuous Engagement, and NIH R21AG028719 (Duncan, GE), Ubiquitous Computing for the Measurement of Physical Activity.

REFERENCES

1. WHO Obesity and overweight. www.who.int/dietphysicalactivity/publications/facts/obesity
2. Obesity and Overweight: Trends: U.S. Obesity Trends. http://www.cdc.gov/nccdphp/dnpa/obesity/trend/maps/
3. Johansson, G., Wikman, A., Ahrn, A.M., Hallmans, G., and Johansson, I. Underreporting of energy intake in repeated 24-hour recalls related to gender, age, weight status, day of interview, educational level, reported food intake, smoking habits and area of living. *Public Health Nutrition,* 4(4), 919—27, 2001.
4. Brunner, E., Stallone, D., Juneja, M., Bingham, S., and Marmot, M. Dietary assessment in Whitehall II: comparison of 7 d diet diary and food-frequency questionnaire and validity against biomarkers. *British Journal of Nutrition*, 86:3, 405—14, 2001.
5. Duncan, G.E., Sydeman, S.J., Perri, M.G., Limacher, M.C., and Martin, A.D. Can sedentary adults accurately recall the intensity of their physical activity? *Preventative Medicine,* 33, 18—26, 2001.
6. Denning, T., Andrew, A., Chaudhri, R., Hartung, C., Lester, J., Borriello, G., and Duncan, G. BALANCE: Towards a Usable Pervasive Wellness Application with Accurate Activity Inference. In *International Workshop on Mobile Computing Systems and Applications* (Feb 23, 2008). HotMobile.
7. Andre, D., Pelletier, R., Farringdon, J., Safier, S., Talbott, W., Stone, R., Vyas, N., Trimble, J., Wolf, D., Vishnubhatla, S., Boehmke, S., Stivoric, J., Teller A. The Development of the SenseWear armband, a Revolutionary Energy Assessment Device to Assess Physical Activity and Lifestyle. http://www.bodybugg.com/pdf/wp_accuracy_ee.pdf
8. Nike+. http://nikeplus.nike.com/nikeplus/
9. Activity Monitor | Nokia Research Center. http://research.nokia.com/projects/activity_monitor
10. Sports Tracker | Nokia Research Center. http://research.nokia.com/research/projects/SportsTracker/
11. Choi, J.E., Lee, J., Hwang, H.T., Kim, J.P, Park, J.C., and Shin, K. Estimation of Activity Energy Expenditure: Accelerometer Approach. *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, (2005), 3830-3833.
12. Nagle, F.J., Balke, B., and Naughton, J.P. Gradational step tests for assessing work capacity. *Journal of Applied Physiology* 20, 4 (1965), 745-8.
13. Bassett, D.R., Giese, M.D., Nagle, F.J., Ward, A., Raab, D.M., and Balke, B. Aerobic requirements of overground versus treadmill running. *Medicine and Science in Sports and Exercise* 17, 4 (1985), 477-81.
14. American College of Sports Medicine. *ACSM's guidelines for exercise testing and prescription.* (M Whaley, senior editor, 7th ed.). Philadelphia, PA: Lippincott Williams & Wilkins; 2006.
15. Choudhury, T., Borriello, G., Consolvo, S., Haehnel, D., Harrison, B., Hemingway, B., Hightower, J., Klasnja, P., Koscher, K., LaMarca, A., Landay, J., LeGrand, L., Lester, J., Rahimi, A., Rea, A., and Wyatt, D. The Mobile Sensing Platform: An Embedded Activity Recognition System. *IEEE Pervasive Computing* 7, 2 (2008), 32-41.
16. Ying, H., Silex, C., Schnitzer, A., Leonhardt, S., and Schiek, M. Automatic Step Detection in the Accelerometer Signal. *4th International Workshop on Wearable and Implantable Body Sensor Networks*, Springer Berlin Heidelberg, 2007.
17. S. Consolvo, D.W. McDonald, T. Toscos, M.Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, & J.A. Landay. Activity Sensing in the Wild: A Field Trial of UbiFit Garden. In *Proceedings of the Conference on Human Factors and Computing Systems: CHI '08*, (Apr 2008), pp. 1797-806.
18. Rowlands, A.V., Thomas, P.W.M., Eston, R.G., and Topping, R. Validation of the RT3 triaxial accelerometer for the assessment of physical activity. *Medicine and Science in Sports and Exercise* 36, 3 (2004), 518-24.
19. Tsai, C., C., Lee, G., Raab, F., Norman, G., J., Sohn, T., Griswold, W., G., and Patrick, K. Usability and feasibility of PmEB: a mobile phone application for monitoring real time caloric balance. In *Mobile Networks and Applications*, 2007.
20. Phillips Respironics. Actical. http://respironics.com/, Nov 2008.