

Machine Learning and Sensor Fusion for Estimating Continuous Energy Expenditure

Nisarg Vyas, Jonathan Farrington, David Andre, John (Ivo) Stivoric

■ In this article we provide insight into the BodyMedia FIT armband system—a wearable multisensor technology that continuously monitors physiological events related to energy expenditure for weight management using machine-learning and data-modeling methods. Since becoming commercially available in 2001, more than half a million users have used the system to track their physiological parameters and to achieve their individual health goals including weight loss. We describe several challenges that arise in applying machine-learning techniques to the health-care domain and present various solutions utilized in the armband system. We demonstrate how machine-learning and multisensor data-fusion techniques are critical to the system's success.

In the United States alone, approximately \$2.6 trillion was spent on health care in 2010. It is well recognized that regular and accurate self-monitoring of physiological parameters and energy expenditure (calorie burn) can improve self-awareness of personal health by providing important feedback. Such awareness and tracking are prerequisites for cost-effective health management, illness reduction, health-conscious decision making, and long-term lifestyle changes.

There exists a wide spectrum of technologies available for monitoring physical activity, tracking energy expenditure, and managing weight. While many of these technologies provide some degree of accuracy, the most accurate among them, metabolic carts and calorimetry chambers, are bulky, expensive, and limited to laboratory and clinical use (Holdy 2004). In contrast, those that are small and inexpensive are, by and large, inaccurate.

At the high-accuracy end of the body monitor space is the doubly labeled water technique, a medical procedure that is guaranteed to give accurate measures of energy expenditure (Schoeller et al. 1986), but is very expensive and only gives readings for a 10- to 14-day period, making it impractical for continuous or short-term monitoring. At the less precise end of body monitor devices are several single-sensor (predominantly accelerometer-based) devices currently in the consumer market

that are low cost and lightweight at the expense of accuracy (Beighle, Pangrazi, and Vincent 2001; Crouter et al. 2003).

We believe that a physiological monitoring device that provides estimates such as energy expenditure should be accurate, provide continuous user feedback, be user friendly, and be fully functional during all the activities of a user's daily life (free-living conditions). Moreover, the device should be cost-effective. The presented BodyMedia FIT armband system (BodyMedia 2011) achieves these goals. The effective integration of machine-learning methodologies and a multisensor technology used in a smart manner can rival medical-grade equipment in terms of clinical accuracy, at the same time surpassing such equipment by collecting data in real time under free-living conditions.

The BodyMedia FIT system is able to provide accurate free-living energy expenditure estimates for two principal reasons—usage of machine-learning-based algorithms and multiple-sensor technology. The system employs state-of-the-art data modeling and machine-learning techniques to implement a data-centered process to estimate, rather than measure, most key physiological parameters. Multiple sensors operate concurrently to provide a real-time user activity context, which, in turn, provides a context-sensitive estimate of the users' physiological parameters.

This article will describe some of the challenges associated with estimating energy expenditure, engineering the BodyMedia FIT armband, applying machine-learning techniques used in developing the estimation algorithms, as well as the results of several studies assessing accuracy of the device and the practical utility of the device in a weight-loss scenario.

Background

Figure 1 shows the armband device (model MF). It is worn on the upper arm. The current commercial version uses four types of sensors: a three-axis accelerometer tracks the movement of the upper arm and body and provides information about body position. A synthetic heat-flux sensor measures the amount of heat being dissipated by the body to the immediate environment. Skin temperature and armband-cover temperature are measured by sensitive thermistors. The armband also measures galvanic skin response (GSR), the conductivity of the wearer's skin, which varies due to sweating and emotional stimuli. The armband contains a transceiver radio and a Universal Serial Bus (USB) port, allowing wireless transmission as well as wired uploading of data. The armband is made predominantly of natural Acrylonitrile Butadiene Styrene (ABS) and 304 grade stainless steel

and attaches to the arm with an elastic Velcro strap. The armband is approximately 55 by 62 by 13 mm (2.2 by 2.4 by 0.5 inches) and weighs 45.4 grams (1.6 oz), it stores more than 14 days of continuous body data and has enough power for 5–7 days of wear from a rechargeable battery when worn 23 hours a day. Each sensor is sampled 32 times per second. Other BodyMedia armband monitors are available that record the same sensor information but differ in other features such as Bluetooth wireless or increased memory capacity.

The system collects physiological data on a continuous basis from the armband user. Data is conditioned, analyzed, interpreted, and stored within the device. The device's on-board algorithms provide real-time estimations of key physiological measures of interest such as the energy expenditure, total number of steps, and number of minutes of moderate and vigorous activity. These key measurements can be displayed wirelessly on a BodyMedia FIT display device (figure 1c), or on a phone using a mobile application such as the iPhone application shown in figure 1d. Additionally, the data can later be transferred electronically (through USB or wirelessly) to a computer or to a BodyMedia web account, where the software reanalyzes the data and makes a definitive high-level analysis of the data with algorithms that are too computationally expensive to run in the device's firmware.

The system collects physiological data on a continuous basis from the armband user. Data is conditioned, analyzed, interpreted, and stored within the device. The device's on-board algorithms provide real-time estimations of key physiological measures of interest such as the energy expenditure, total number of steps, and number of minutes of moderate and vigorous activity. These key measurements can be displayed wirelessly on a BodyMedia FIT display device (figure 1c), or on a phone using a mobile application such as the iPhone application shown in figure 1d. Additionally, the data can later be transferred electronically (through USB or wirelessly) to a computer or to a BodyMedia web account, where the software reanalyzes the data and makes a definitive high-level analysis of the data with algorithms that are too computationally expensive to run in the device's firmware.

Various versions of the armband have been in active use by hundreds of thousands of users over the last nine years. The earlier products were larger, heavier, and more expensive to manufacture. The earliest of these had only a two-axis accelerometer rather than the current three-axis model. Additionally, the algorithms have been updated numerous times over the years because more data enables more accurate and improved algorithms.

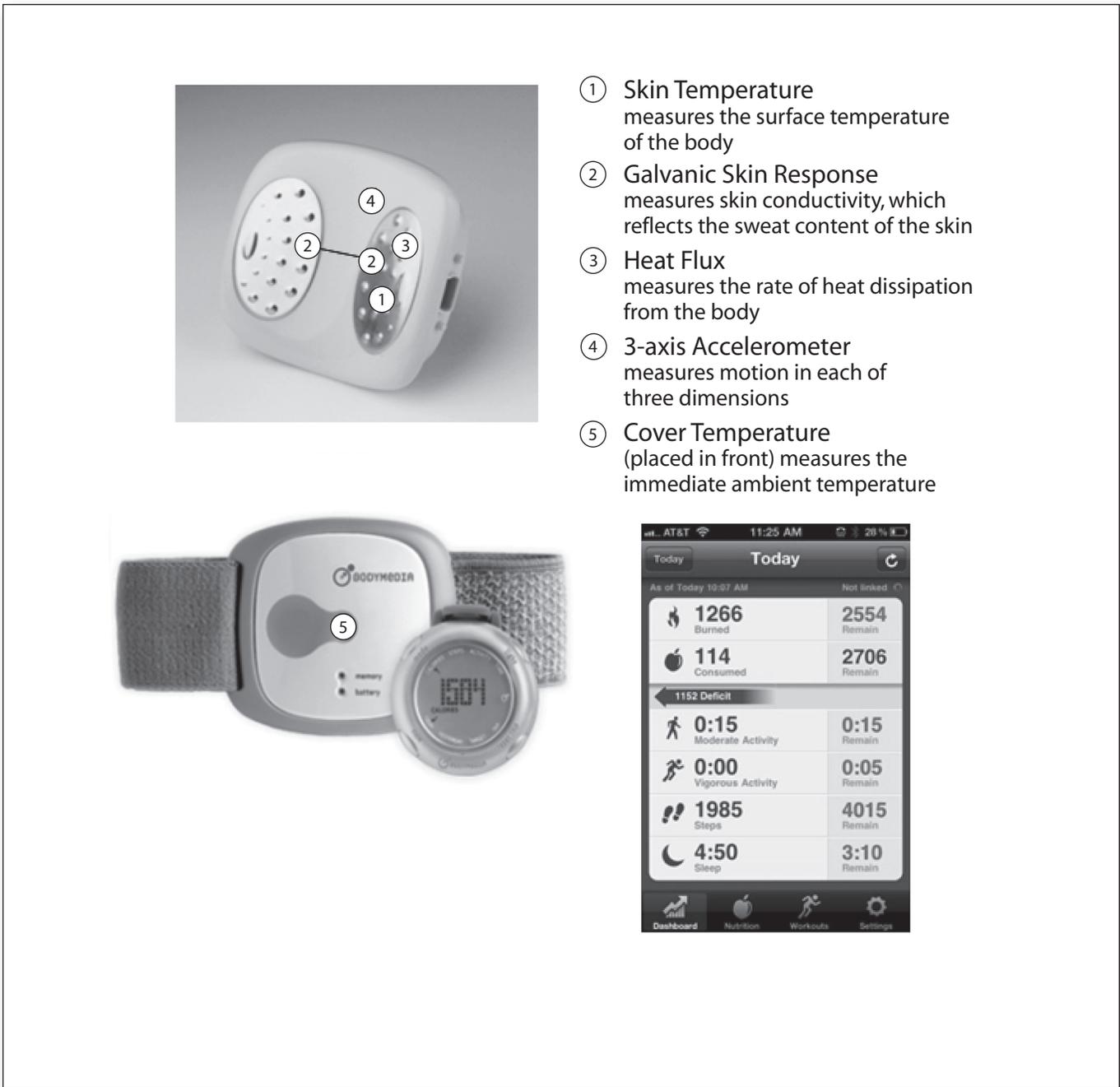


Figure 1. BodyMedia System Overview.

(a) (top left) BodyMedia armband device and its sensor layout. The figure shows the side of the device worn against the skin. (b) (top right) Sensor description. (c) (bottom left) Front of the armband (showing the cover temperature sensor) and a display device that provides real-time feedback of physiological parameters to the end user. (d) (bottom right) iPhone application that provides real-time feedback.

The BodyMedia armband is in use in several commercial applications, including the BodyMedia FIT product and the bodybugg product from 24 Hour Fitness, especially targeted for consumers trying to lose or maintain their weight and increase fitness.

Introduction to Energy Expenditure Measurement

The number of calories a person burns is an important and actionable parameter for achieving many health goals, such as weight control and sports per-

formance, as well as for managing many disease conditions including metabolic disorders, diabetes, obesity, and associated diseases. Due to the complex nature of metabolism, true total energy expenditure (TEE) is very difficult to measure. Nearly all techniques make use of approximations of one kind or the other. Indirect calorimetry, doubly labeled water, self-report techniques, pedometers, heart-rate monitors, and accelerometers are a few methods commonly used for energy expenditure estimation.

With indirect calorimetry, widely accepted in the sports medicine research community, metabolic carts measure the oxygen and carbon dioxide a person inhales and exhales and then indirectly compute calories burned by comparing the input and output volumes of the two gases. Based on a survey of the literature, devices in this category differ from one another, as well as between repeated measurements of the same activity on the same machine by 5–10 percent (Yates, Cullum, and Pittsley 2004; Wells and Fuller 1998). Most metabolic carts are rather large and bulky and are not suited for monitoring outside a laboratory while portable carts are not as accurate. These devices are expensive, costing upward of US\$20,000 for a basic system and US\$40,000 for a portable oxygen analyzer (Holdy 2004, Berntsen et al. 2010).

The doubly labeled water (DLW) stable isotope method is considered to be the gold standard for measuring TEE under free-living conditions (Schoeller et al. 1986). This method is based on the principle that in a loading dose of $^2\text{H}_2^{18}\text{O}$, the ^{18}O isotope is eliminated as CO_2 and water, while the deuterium isotope is eliminated as water. The rate of CO_2 production, and thus energy expenditure, is calculated from the difference in residual isotopes of hydrogen and oxygen remaining in urine at the end of the testing period. Limitations of the DLW method include high cost, the need for specialized equipment and expertise to implement the techniques, and the fact that the method can only be used to measure expenditure over a long period of time (for example, 10–14 days).

Self-report methods include questionnaires, interviews, and activity diaries. There are some advantages of using self-reports or 24-hour recalls, as they are inexpensive and easy to administer. However, estimating duration and energy expenditure with these can provide only a rough and inaccurate estimate of activity levels.

Pedometers, by definition, measure footfalls. The clear advantage of pedometers is their low cost, ranging from \$15 to \$300 (Beighle, Pangrazi, and Vincent 2001; Freedson and Miller 2000). In general, pedometers are not accurate when used for activities that do not involve footfalls (for example, weight lifting, biking, household activities).

Heart rate is one of the fundamental vital signs and is intimately related to the level of physical exertion. Especially for moderate to strenuous activity, a person's heart rate increases linearly with oxygen consumption (Freedson and Miller 2000, Welk 2002). Although heart-rate monitoring is often used as part of an exercise prescription, heart rate is not linearly related to energy expenditure for activities that do not involve large muscle groups, such as sitting, resting and sleeping. Moreover, chest-strap heart-rate monitors can be burdensome to users due to constriction across the chest necessary to maintain good skin contact. Electrode-based heart monitors are difficult to wear, because electrode placement, preparatory skin treatment, and skin irritation can be detrimental to long-term use.

Accelerometers operate by measuring acceleration along a given axis, using any of a number of techniques, including piezoelectric, micromechanical springs, and changes in capacitance. Often, multiple axis measurements are bundled into a single package, allowing two- and three-axis accelerometers. While there is no simple linear relationship between acceleration and energy expenditure, for certain well-understood activities, such as walking, well-understood basic principles of physics (work) and physiology (muscle efficiency) can aid in developing appropriate equations.

Modeling and Design

BodyMedia's approach to the estimation of energy expenditure is nonconventional and different from the approaches mentioned above, since it employs machine learning to solve the problem. Machine learning in this context is faced with some significant challenges including data quality, inherent variability, and acquisition expense; unavoidable disparities between training and testing distributions; real-time feedback requirements; reliable functionality during hardware/firmware upgrades; and necessary hardware size, power, and cost limitations.

First and foremost is the need for high-quality data, which can be very expensive to obtain on a sufficiently broad set of activities and subjects.

The second issue is in the inherent variability present in the data set and in the target user population. Each armband user is unique in physical characteristics such as age, weight, gender, and fitness level. Each of these characteristics affects the relationships between measured parameters and energy expenditure. Moreover, there is variation in data due to geographical and environmental differences, including different humidity and external temperatures. Additional sources of variation include differing calibration specifications and improper adjustment of medical gold-standard

equipment. In order to ensure precision, it is imperative that candidate models be sufficiently robust to accommodate the aforementioned variations.

The third challenge stems from inherent disparities between data used for building energy expenditure (EE) estimation models and data collected in actual use. The gold-standard data used for building the EE estimation models can only be collected from a limited set of simple, time-constrained and mostly indoor activities in lab settings, whereas their actual use occurs in free-living, real-world settings where the users perform a multitude of complex activities. Making EE estimation models in these circumstances violates a fundamental assumption of machine learning that both the training and testing distribution are drawn from the same distribution.

The fourth challenge is in the fact that the models themselves have to satisfy multiple objectives. For example, the model should be accurate for minute-to-minute real-time feedback for specific activities as well as for weeks-long free-living protocols composed of a multitude of activities. These different use cases can make model selection difficult.

Another challenge is that algorithms need to function reliably through hardware improvements, including miniaturization and simplifications, that result in reduced costs.

Finally, many models require the production of real-time results. In this event, an on-board processor that has limited memory and computational capability computes the algorithm. This influences the choices of underlying features and modeling methods, in that we prefer efficient methods with respect to time and space complexity.

Modeling Process

BodyMedia's modeling process can be defined in the following steps: data collection, data cleaning, feature generation, development of context detectors, development of regression models, and finally internal and external validation.

Data Collection

Any nontrivial machine-learning method needs good data. To meet the challenge of obtaining high-quality data, BodyMedia conducts data collection studies at multiple clinical sites spread across the globe. We have worked to enlist many academic researchers as colleagues and advisors, allowing us to obtain data from far more studies than we could collect ourselves. Data collection is designed to provide sufficient samples to capture variability in the domain. As specific examples, the data used in the algorithms range from 5-year-old children to retirees in their 70s; it represents unhealthy subjects suffering from multiple dis-

eases at one end of the spectrum to elite athletes participating in sports events at the other. We capture data from people engaging in many different activities as well, ranging from restful activities such as sleeping and lying down to highly vigorous activities such as sprinting, stair-master, rowing, and mountaineering. The collected data can come from either of the two environments: free-living with user-annotated activities or in a laboratory or controlled environment following a strict protocol. For most lab studies, data from high-accuracy gold-standard equipment, such as metabolic carts or metabolic chambers, is also collected for training and testing purposes. While free-living data consists of many activities and is used for activity classification, lab data is limited to a certain subset of activities and is used primarily for building EE estimation models.

Data Cleaning

We have developed a rigorous process for cleansing data and preparing it for modeling. Armband sensor data in each data file is verified; cases of sensor malfunction are detected by comparing the armband sensor data with the sensors' standard distribution. If outliers are found in the sensor values, those data points are discarded. In cases where gold-standard medical equipment data is also collected, each data point is carefully aligned using semiautomated procedures. In cases where gold-standard equipment either is not properly calibrated or exhibits a tendency to overestimate or underestimate, its compliance with standard MET (metabolic equivalent, essentially energy cost per unit of mass) ranges for the corresponding activities (Ainsworth et al. 2000) is utilized for data validation. Moreover, data sets are checked for correctness of the activity annotations, which are usually inserted manually by the user or experimenter. Activity annotations are verified for correctness by comparing the sensor values with the standard sensor value distribution for the activity and applying other such heuristics. As an example, most cases of resting activities cause minimal changes in motion sensors and GSR sensors. If the sensors recorded a high degree of motion and/or steep rises in GSR, it is very likely that the activities were annotated incorrectly. For cleaning of the activity annotations, the policy is to err on the side of caution, as we have found that allowing even small amounts of poorly labeled or misaligned data can have inordinate effects on algorithm performance.

Feature Generation

The sensors used in the armband are sampled at 32 Hz, whereas the armband records data every minute (this can be adjusted through software). Thus compressed and summarized features are cal-

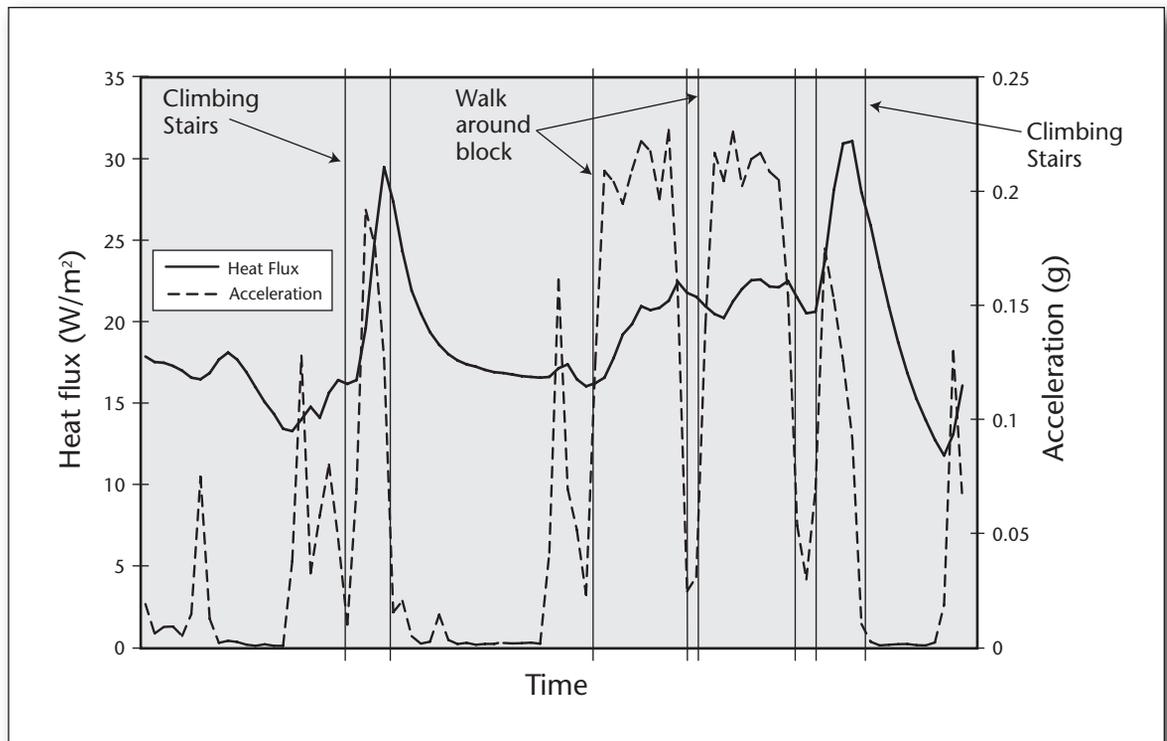


Figure 2. A Sample of Armband Signals during Various Activities.

Multiple sensors allow better activity classification.

culated and created from the raw data. More than 50 features of this multidimensional raw data stream are gathered as separate channels. For example, the variance of the heat flux is a channel, as is the average of the heat-flux values. Some channels are fairly standard such as standard deviation, frequency, peaks, and averages. Others are complex proprietary algorithms embedded in the on-board processor of the armband. Typically these summary features for each minute-epoch/duty cycle are stored and the raw data is discarded to conserve memory. Typically, we refrain from calculating computationally complex and storage-expensive operations, such as Fourier transforms, in order to conserve the on-board processor's memory and finish the full-feature calculation in each epoch. We also use approximations if the actual feature is computationally expensive to create.

The next stage of feature generation is performed on the data retrieved from the armband. This is done to find features useful for recognizing patterns of activity and for calibrating various measures against one another. For example, relative values of GSR are often more useful than absolute readings. Multiple methods are used to extract these features: some features are derived using domain knowledge of exercise and physiolo-

gy, some are derived using an automated feature-generation technique similar to genetic programming where features must pass a few statistical tests such as high correlation to the ground-truth EE in all or some activities, and some features are derived using standard machine-learning feature-generation techniques such as principal components analysis (PCA) and independent component analysis (ICA). Others are added based on intuition and visual observation. This phase results in a feature space with more than 500 variables.

Development of Context Detectors

The next stage of the modeling process is to develop a series of classifiers that break down a user's activity into primary components for which good models of energy expenditure can be created. Classifiers are created for the following basic activities: sleeping, resting, motoring, walking, running, weight lifting, stationary biking, and road biking. Many popular, computationally inexpensive machine-learning methods such as naive Bayes and decision trees are tried out for feature selection and training for the classifiers.

To avoid overfitting, all the feature selection and classification algorithms use k -fold cross-validation. In our experience, it is not sufficient to perform k -fold point-based cross-validation (that is, creating

folds at the level of individual data points) to ensure robustness and generalization capabilities, as there are many subject-specific traits present in the data. Instead, each subject's entire data is assigned to one of the k folds. We refer to this strategy as k -fold by-subject cross-validation. Using by-subject cross-validation results in algorithms that generalize well to unseen subjects. Moreover, to avoid overfitting by the classifiers, it was found to be necessary to avoid features that include subject-specific traits (such as demographic information).

The use of multiple sensors provides orthogonal sets of features in the feature space, helping to provide more discriminating capability to the classifiers. As an example, figure 2 represents signal values of one accelerometer-based feature and a heat-flux-based feature during various activities. The values for the accelerometer feature for "climbing stairs" and "walking around the block" are very similar, making it complex for the classifiers to infer the activity based on the accelerometer data. With the introduction of heat flux, even a simple classifier can distinguish between the two activities.

Typically, the classifier model is designed as a hierarchical combination of various subclassifier responses. A base classifier classifies the data into generic activities, and the next level of the classifier model provides a more fine-grained label to the activities. For example, one can think of the base classifier only classifying an activity as "biking," and the second-phase classifier classifying all the "biking" data points into "stationary biking" and "road biking." The classifier model makes use of several subclassifiers spread across multiple levels of hierarchy.

Development of Regression Models

In this phase, several regression models are built that provide energy expenditure estimates. Usually the models are built for a specific activity (or for a set of very similar types of activities). The regressions are then combined according to the probabilities output from the activity classifiers. Many prevalent AI-based regression techniques such as robust regression and locally weighted regression are used for fitting the data. Feature selection and training for the regression models is also performed using k -fold by-subject cross-validation.

Most of the physiological measures of interest estimated by the armband are dependent on subject-specific traits (for example, mass). Rather than predicting absolute measures, regressions are tuned to predict relative measures that are subsequently adjusted for the subject. For example, in the case of energy expenditure, the regression models are actually trained on the relatively subject-independent unit METs (Ainsworth et al. 2000) instead of absolute units such as kJoules or kcalories.

The steps of feature generation, context detector development, and regression estimation overlap one another and are addressed simultaneously. Multiple iterations of feature generation, classifier modeling, and regression modeling result in improved algorithms.

Internal and External Validation

For each algorithm release cycle, certain data sets are kept untouched for the entire development period, and performance of the model is evaluated on those validation sets. The models are approved and released only if they pass predefined criteria on the validation sets as well as the training sets. Similar to the training data sets, validation sets are subjected to a sufficiently large array of data samples and a sufficiently broad variety of subjects for each activity to ensure statistical validity. Some of the validation sets target particular areas of concern such as a demographic group such as children, unhealthy adults, or athletes. Accordingly, some of the validation sets target only specific activities. Some types of data sets are good only to serve as validation sets, for example, the doubly labeled water data set, where there is only one reading of total energy expenditure every two weeks. At the alpha and beta stages of the model release, results of the models are observed, and minor changes are made to the model if necessary. Many researchers also carry out independent external validation and performance evaluation of the system. In most cases, independent external validation is performed after the product is released, providing helpful cues to further improvement for the next version of the models (St. Onge et al. 2007, Welk et al. 2007, Jakicic et al. 2007, Malavolti et al. 2007).

Results

BodyMedia armbands have been commercially available since 2001. The fifth generation of the system is currently sold in the market. There are more than half a million users of the system throughout the world. To date, BodyMedia has collected more than 10 billion minutes of armband data. The system has recorded more than 170 billion steps and estimated more than 20 billion calories.

Data Sets

In the most recent energy expenditure algorithm created at BodyMedia, a data set with roughly 1 million minutes featuring around 800 users was used for training the context detectors. All the minutes were carefully annotated by the users or experimenters and cleaned to make the data suitable for modeling. Developing the regressions required a gold-standard data set that had 658 sub-

Activity	True-Positive Rate Percent	True-Negative Rate Percent	Accuracy
Sleeping	76.9	98.7	89.8
Resting	91	80.8	83.8
Motoring	75.3	97.4	96.3
Walking	96.9	94.3	94.4
Running	92.5	99.5	99.2
Weight Lifting	39.3	80.8	83.8
Stationary Biking	61.5	98.5	98.1
Road Biking	90.6	99.6	99.1

Table 1. Classification Results for Prominent Activities, Evaluated Using By-Subject Cross-Validation.

Activity	Most Predicted Classes per Activity (with True-Positive Rate Percent)
Sleeping	76.9 percent sleeping, 22.9 percent resting
Resting	91 percent resting, 5.5 percent weight lifting
Motoring	75.3 percent motoring, 18.7 percent resting, 3.7 percent stationary biking
Walking	96.9 percent walking, 3 percent running
Running	92.5 percent running, 7.4 percent walking
Weight Lifting	39.3 percent weight lifting, 20.2 percent stationary biking, 15.7 percent resting
Stationary Biking	61.5 percent stationary biking, 10.1 percent weight-lifting, 9.2 percent motoring
Road Biking	90.6 percent road biking, 2.9 percent stationary biking, 2.5 percent resting

Table 2. Four Most Predicted Classes per Activity.

jects and approximately 40,000 data points.

The data sets had a wide range of demographic variations: age ranged from 5 years to 78 years; weight ranged from 18 kg to 152 kg (40 lb to 335 lb). The data was collected from more than 50 different studies, conducted at clinical sites spread across the world as well as from studies conducted in-house.

Classification

Tables 1 and 2 show classification results for the most recent algorithm for major activities. The main purpose of classification is to assign the most appropriate regression model to the query data point. Some of the true-positive rates are seemingly quite low. Further inspection reveals that much of the misclassification happens between similar types of activities. For example, misclassification between motoring and resting occurs often, but from an energy expenditure standpoint the misclassification does not cost much because their EE ranges are very similar.

Table 2 shows the four most frequently predict-

ed classes for each true class (the confusion in classification). Tables 1 and 2 show results evaluated by using by-subject cross-validation. The model generalizes well to unseen subjects' data, with the overall accuracy of the unseen subjects' data set just 1 percent less than the accuracy obtained for by-subject cross-validation.

Regression Models

Figure 3 shows results of average METs for a new release candidate versus a recent model already in use per each activity. The METs value can be thought of as the relative activity intensity and energy requirement. The figure shows the release candidate's estimates are much closer to true average METs than the previous version, hence confirming the improvement.

Typically, the errors on the regression models are measured in Mean Absolute Percentage Errors (MAPEs). MAPEs are calculated averaging over each minute as well as averaging over each session of continuous observation. The new release candi-

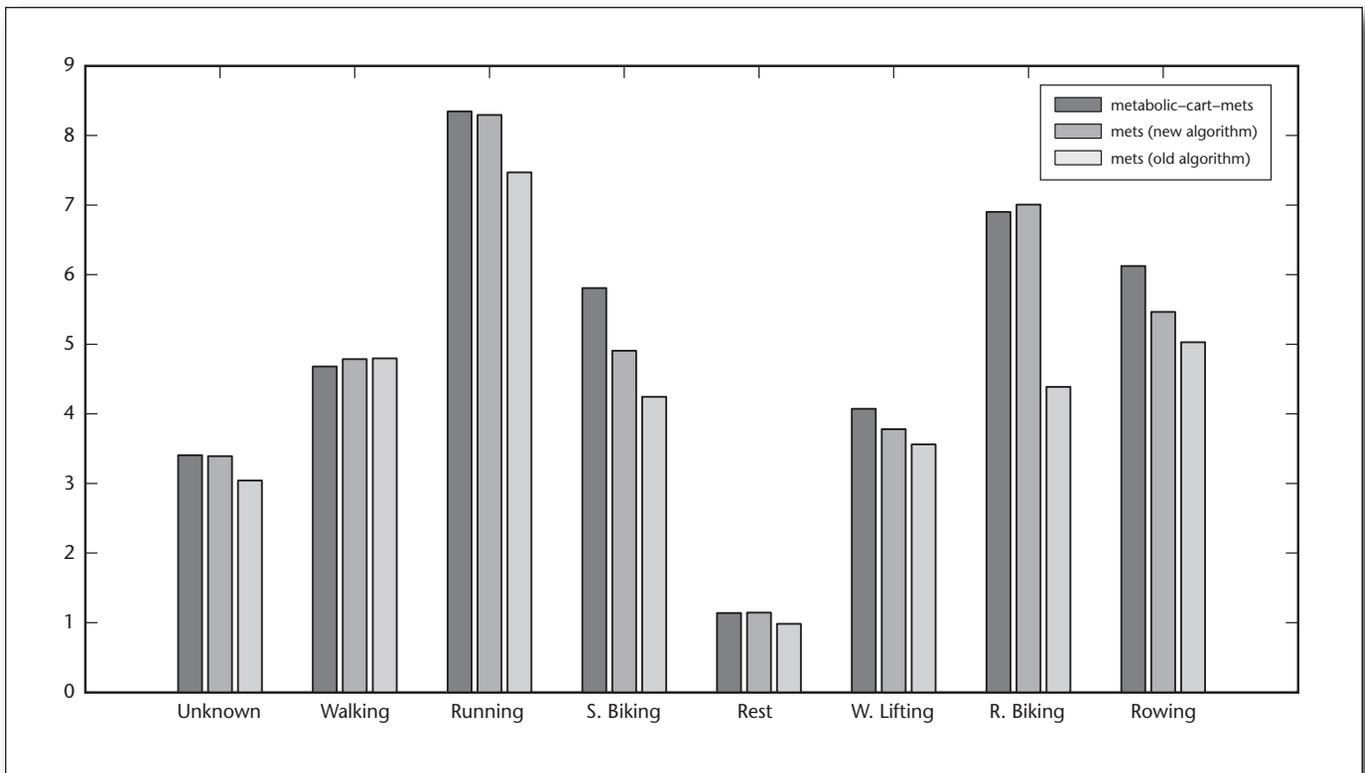


Figure 3. Average METs per Activity.

Ground truth computed from metabolic carts, for current algorithm and new release candidate.

date algorithm has 15 percent session MAPEs over all the lab data. The new release candidate algorithm provided robust estimates for children as well as adults, with the session MAPEs as low as 13.7 percent. Daily MAPEs for adults are expected to be lower than the lab data suggests. The lab data is composed predominantly of subjects engaged in exercise, whereas a typical day is made up of mostly sleep (about 30 percent), restful activities (about 60 percent), and a small amount of moderate to vigorous activity.

Doubly Labeled Water Data

As mentioned earlier, doubly labeled water is the most accurate method to estimate energy expenditure, but provides only one reading per 14 days.

Figure 4 shows a scatter plot with estimated TEE values that match well with the actual TEE calculated from the DLW method. The data was collected on 30 adult individuals and 30 children wearing two versions of armbands (2008–2010 model Pro3 and 2010 model MF), one on each arm for a period of two weeks. The MAPE is less than 10 percent for adults and the correlation between the true and estimated TEE is 0.88 (Johansen et al. 2010). A similar study was also conducted for children, in which the MAPE was under 15 percent (Calabró et al. 2011).

An independent study (Bernsten et al. 2010) val-

idated the accuracy of the armbands in simulated free-living conditions, where 20 subjects participated in 60 to 120 minutes of realistic daily activity. The estimation error from the armbands was less than 10 percent. These results demonstrate that the models are generic enough to work for unseen subjects performing free-living activities.

Results of doubly labeled water studies are of paramount importance during the prerelease evaluation of the candidate models because they provide valuable feedback on how an algorithm would behave in free-living settings.

Accelerometers-only Versus Multisensors: We conducted a study to measure the efficacy of models built based on the current sensor set versus models built only on the accelerometer- and motion-based signals, but still using BodyMedia's pattern-recognition methods. More than 30 subjects participated in various exercise activities. We found that models that used all the sensors had 8 percent per subject error, whereas the models that used only the accelerometers had 12–15 percent per subject error. A separate study also provided comparative evaluation of BodyMedia armband devices with other commercially available energy expenditure estimation devices and it was found that the BodyMedia armband system provided the most accurate results in comparison to other devices. The next best device had 14 percent per

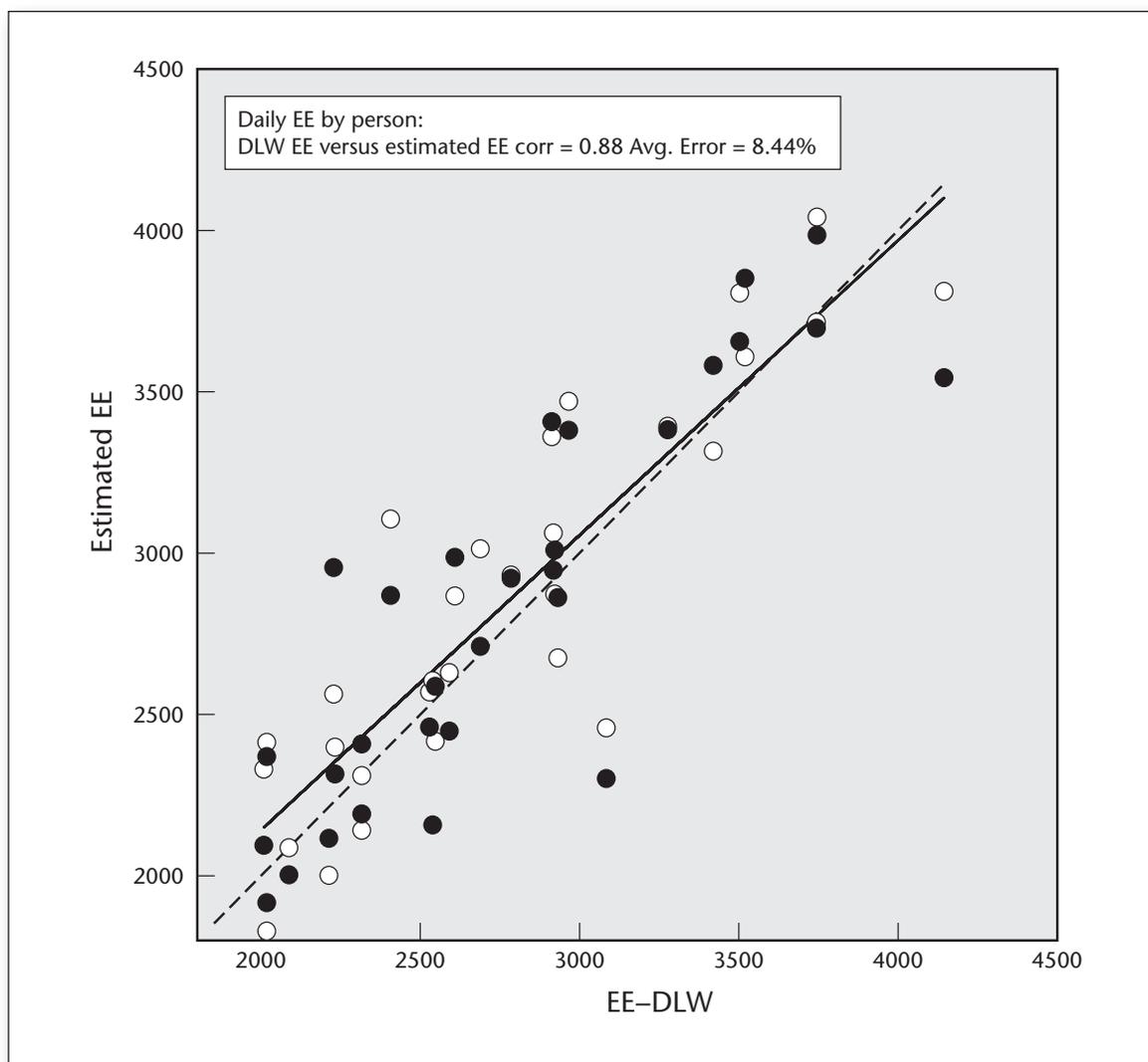


Figure 4. DLW Results Compared with Estimated EE for 30 Adult Subjects.
Fourteen days per subject. Black is armband model Pro3, white is model MF.

subject error, in comparison to the BodyMedia system's 8 percent per subject error.

Weight-Loss Results

Not only is the armband in use by many users, but it also appears to significantly help users in achieving their weight-loss and lifestyle goals. A study performed at the University of South Carolina showed that participants who used armbands in their weight-loss program lost more than twice as much weight as subjects who did not use the armbands (Barry et al. 2010, Sui et al. 2010). A weight-loss study done at the University of Pittsburgh achieved similar results (Pellegrini et al. 2010).

Real-Time Versus Offline EE Estimates

As mentioned earlier, the commercial system provides two types of estimates: one in real time, computed by the on-board processors, and the other

one with a richer set of features and modeling methods when users upload their data to their web account. For a good user experience, it is imperative that the real-time EE estimates match closely with the offline EE estimates. The data from the DLW experiment mentioned earlier (30 adult subjects, 14 days wear) found that the mean difference between the real-time EE (display EE) and the offline EE was 2.3 percent per day (about 66 kcal) with a median difference of 1.7 percent per day; see figure 5.

Conclusions and Future Work

With health-care costs growing each year, people can benefit from an effective, inexpensive, easily wearable, and accurate physiological monitoring device: BodyMedia armbands attempt to provide

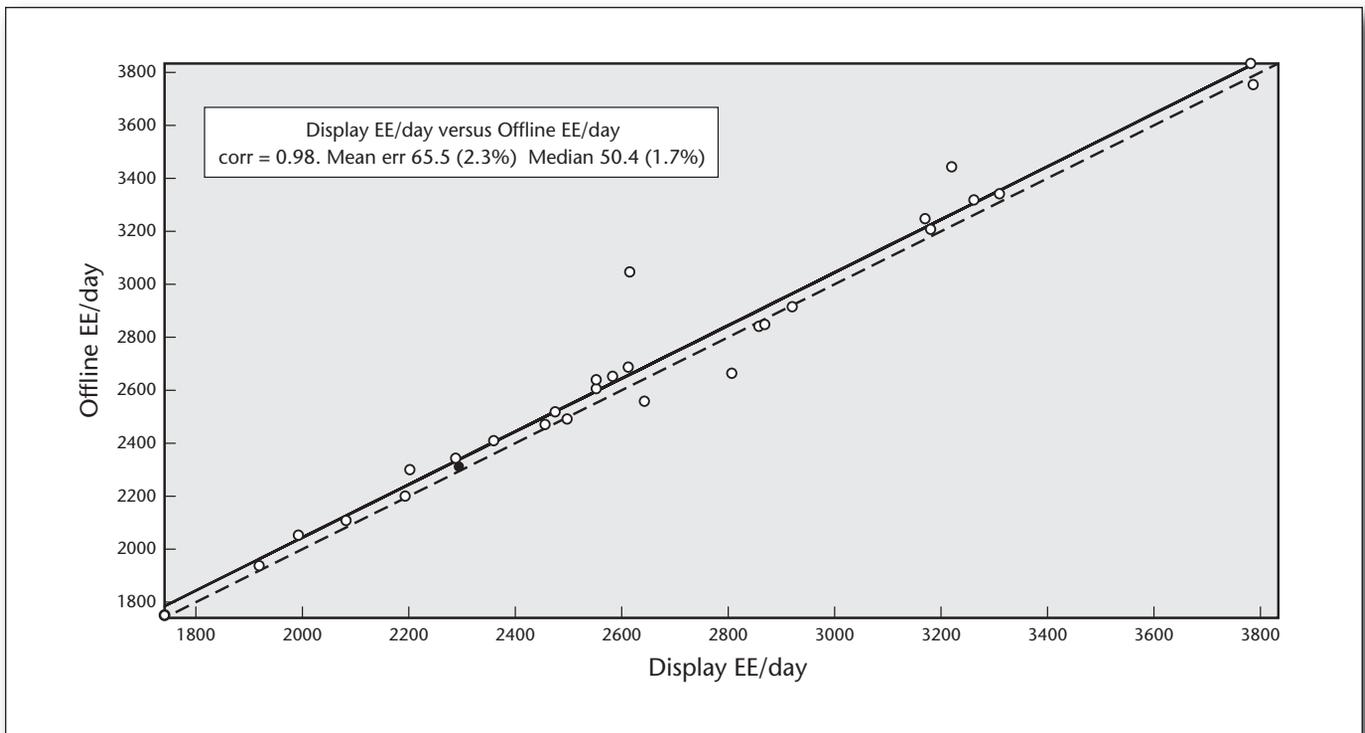


Figure 5. Comparison of Display EE Versus Offline EE.

The correlation is 0.98 and the mean difference is 2.3 percent.

that solution in a nonconventional way using sensor fusion and state-of-the-art machine-learning and artificial intelligence techniques.

In this article, we have described the modeling process for estimating the physiological parameters (especially energy expenditure). We also demonstrate that using AI and machine learning for solving physiological-monitoring-related problems poses unique challenges. Several minor and problem-specific adjustments to the traditional approaches helped us overcome the challenges, such as the use of disparate data sources; careful data cleaning; use of robust features; employing data validation tricks such as by-subject cross-validation; use of an iterative multistep process for building and evaluating models; and performing rigorous evaluation that validates the models for their multiple objectives.

The results presented here demonstrate the capability of the armband sensors and models to provide accurate results for various activities for a large range of users in both laboratory and free-living settings.

BodyMedia is engaged in continued refinements to the platform and the development of new body-monitoring capabilities. These include the integration of new sensors and the ongoing development of data models to extract new physiological features and contextual activities. Some of the other

projects that BodyMedia is focusing on include blood-glucose estimation (Vyas et al. 2010, Rollins et al. 2009), heart-rate estimation (Al-Ahmad, Homer, and Wang 2004), critical-care parameter estimation (Convertino et al. 2010), fine-grained sleep detail, and estimation of total calories consumed. All of these projects make extensive use of data-driven methods and sophisticated machine-learning techniques.

Acknowledgements

We thank the reviewers for useful feedback and suggestions. We thank our colleagues at BodyMedia, especially Rachel Jackson, Scott Boehmke, and Ray Pelletier, for their continuous efforts in data collection study design and management, hardware design, and assistance in modeling, respectively. We would like to thank Dr. Astro Teller for the conceptualization and work on early designs of the armbands. We would also like to thank Dr. Greg Welk, associate professor at Iowa State University, for graciously accepting our request to share and present some of the results on studies conducted by him at Iowa State University.

References

Ainsworth, B. E.; Haskell, W. L.; Whitt, M. C.; Irwin, M. L.; Schwartz, A. M.; Strath, S. J.; O'Brien, W. L.; Bassett, D. R., Jr.; Schmitz, K. H.; Emplincourt, P. O.; Jacobs, D.R.,

- Jr.; and Leon, A. S. 2000. Compendium of Physical Activities: An Update of Activity Codes and MET Intensities. *Medicine and Science in Sports and Exercise* 32(9 Suppl): S498–504.
- Al-Ahmad, A.; Homer, M.; and Wang, P. 2004. Accuracy and Utility of Multi-Sensor Armband ECG Signal Compared with Holter Monitoring. Paper presented at the New Arrhythmia Technologies Retreat, 11 November, Chicago.
- Barry, V.; Shuger, S.; Beets, M.; Sui, X.; Wilcox, S.; Hand, G.; McClain, A.; and Blair, S. 2010. Wearing the SenseWear Armband and Weight Loss in Sedentary Overweight and Obese Adults. Paper presented at the American College of Sports and Medicine meeting, 29 May–2 June, San Francisco.
- Berntsen, S.; Hageberg, R.; Aandstad, A.; Mowinkel, P.; Anderssen, S.; Carlsen, K.; and Andersen, L. 2010. Validity of Physical Activity Monitors in Adults Participating in Free Living Activities. *British Journal of Sports and Medicine* 44(9): 657–664.
- Beighle, A.; Pangrazi, R.; and Vincent, S. 2001. Pedometers, Physical Activity, and Accountability. *Journal of Physical Education, Recreation and Dance* 72(9): 16–19.
- Bodymedia Inc. 2011. What Is Bodymedia FIT? Pittsburgh, PA: BodyMedia Inc. (www.bodymedia.com/Shop/Learn-More/What-is-Body-Media-FIT).
- Calabro, M. A.; Lee, J.; St-Maurice, P.; and Welk, G. 2011. Validation of Pattern-Recognition Monitors in Children Using the Doubly Labeled Water Method. *Medicine and Science in Sports and Exercise*. 43(5): 132.
- Convertino, V.; Hurst, G.; Ryan, K.; Vyas, N.; Ward, K.; and Rickards, C. 2010. Bleeding or Active? Validation of a Machine-Learning Algorithm for Remote Determination of Blood Volume Status. *American College of Sports and Medicine* 42(5): 285.
- Crouter, S.; Schneider, P.; Karbult, K.; and Bassett, D. 2003. Validity of 10 Electronic Pedometers for Measuring Steps, Distance, and Energy Cost. *Medicine and Science in Sports and Exercise* 35(8): 1455–1460.
- Freedson, P., and Miller, K. 2000. Objective Monitoring of Physical Activity Using Motion Sensors and Heart Rate. *Research Quarterly for Exercise and Sport* 71(2 Suppl): S21–9.
- Holdy, K. 2004. Monitoring Energy Metabolism with Indirect Calorimetry: Instruments, Interpretation, and Clinical Application. *Nutrition in Clinical Practice* 19(5): 447–454.
- Jakicic, J.; Marcus, M.; Gallagher, K.; Randall, C.; Thomas, E.; Goss, F. L.; and Robertson, R. J. 2004. Evaluation of the SenseWear Pro Armband to Assess the Energy Expenditure during Exercise. *Medicine and Science in Sports and Exercise* 36(5)(May): 897–904.
- Johannsen, D. L.; Calabro, M. A.; Stewart, J.; Franke, W.; Rood, J. C.; and Welk, G. J. 2010. Accuracy of Armband Monitors for Measuring Daily Energy Expenditure in Healthy Adults. *Medicine and Science in Sports and Exercise* 42(11): 2134–2140.
- Malavolti, M.; Pietrobelli, A.; Dugoni, M.; Poli, M.; Romagnoli, E.; DeCristofaro, P.; and Battistini, N. C. 2007. A New Device for Measuring Resting Energy Expenditure (REE) in Healthy Subjects. *Nutrition, Metabolism and Cardiovascular Diseases* 17(5): 338–343.
- Pellegrini, C. A. 2010. The Comparison of a Technology-Based System and an In-Person Behavioral Weight Loss Intervention in Overweight and Obese Adults. Ph.D. Dissertation, University of Pittsburgh, Pittsburgh, PA.
- Rollins, D.; Bhandari, N.; Kleinedler, J.; Kotz, K.; Strohbehn, A.; Boland, L.; Murphy, M.; Andre, D.; Vyas, N.; Welk, G.; and Franke, W. 2010. Free-Living Inferential Modeling of Blood Glucose Level Using Only Noninvasive Inputs. *Journal of Process Control* 20(1): 95–107.
- Schoeller, D.; Ravussin, E.; Schutz, Y.; Acheson, K.; Baertschi, P.; Jequier, E. 1986. Energy Expenditure by Doubly Labeled Water: Validation in Humans and Proposed Calculations. *American Journal of Physiology* 250(5 Pt 2): R823–30.
- St-Onge, M.; Mignault, D.; Allison, D.; and Lhoret-Rabasa, R. 2007. Evaluation of a Portable Device to Measure Daily Energy Expenditure in Free-Living Adults. *American Journal of Clinical Nutrition* 85(742): 9.
- Sui, X.; Meriwether, R.; Hand, G.; Wilcox, S.; Dowda, M.; and Blair, S. 2010. Electronic Feedback in a Diet and Physical Activity-Based Lifestyle Intervention for Weight Loss: Randomized Controlled Trial. Presented at the American Heart Association 50th Annual Joint Conference: Nutrition, Physical Activity and Metabolism and Cardiovascular Disease Epidemiology and Prevention. San Francisco.
- Vyas, N.; Andre, D.; Rollins, D.; Stivorc, J.; and Jackson, R. 2009. Development of a Personalized Non-Invasive Glucose Monitoring System for Free-Living Environments. Paper presented at the Annual Meeting of Diabetes Technology, November 5–7, San Francisco.
- Welk, G. 2002. *Physical Activity Assessments for Health Research*. Champaign, IL: Human Kinetics.
- Welk, G.; McClain, J.; Eisenmann, J.; and Wickel, E. 2007. Field Validation of the MTI Actigraph and Bodymedia Armband Monitor Using the IDEEA Monitor. *Obesity* 15(4): 918–928.
- Wells, J., and Fuller, N. 1998. Precision and Accuracy in a Metabolic Monitor for Indirect Calorimetry. *European Journal of Clinical Nutrition*. 52(7): 536–540.
- Yates, J.; Cullum, M.; and Pittsley, J. 2004. Validation of a Portable, Indirect Calorimeter for the Measurement of Resting Metabolic Rate. *Medicine and Science in Sports and Exercise*. 36(5): S247.

Nisarg Vyas is a consulting research scientist for Bodymedia Inc.

Jonathan Farrington is director of informatics at BodyMedia Inc.

David Andre is the chief executive officer at Cerebellum Capital, Inc., and a research advisor at BodyMedia, Inc.

John (Ivo) Stivorc is the CTO and vice president of new products for BodyMedia, Inc.