Using Multi-modal Sensing for Human Activity Modeling in the Real World

Beverly L. Harrison, Sunny Consolvo, and Tanzeem Choudhury

Abstract This chapter describes our experiences over a five-year period of building and deploying a wearable system for automatically sensing, inferring, and logging a variety of physical activity. In this paper, we highlight some of the key findings resulting from the deployment of our system in 3-week and 3-month real world field trials, which are framed in terms of system usability, adaptability, and credibility.

1 Introduction

Traditionally smart environments have been understood to represent those (often physical) spaces where computation is embedded into the users' surrounding infrastructure, buildings, homes, and workplaces. Users of this "smartness" move in and out of these spaces. Ambient intelligence assumes that users are automatically and seamlessly provided with context-aware, adaptive information, applications and even sensing – though this remains a significant challenge even when limited to

Beverly Harrison and Sunny Consolvo
Intel Research, 1100 NE 45th Street, Seattle, WA 98103, USA, e-mail: beverly.harrison, sunny.consolvo@intel.com

Tanzeem Choudhury
Dartmouth College, Department of Computer Science, 6211 Sudikoff Lab, Hanover, NH 03755, USA, e-mail: tanzeem.choudhury@dartmouth.edu

these specialized, instrumented locales. Since not all environments are "smart" the experience is not a pervasive one; rather, users move between these intelligent islands of computationally enhanced space while we still aspire to achieve a more ideal anytime, anywhere experience. Two key technological trends are helping to bridge the gap between these smart environments and make the associated experience more persistent and pervasive. Smaller and more computationally sophisticated mobile devices allow sensing, communication, and services to be more directly and continuously experienced by user. Improved infrastructure and the availability of uninterrupted data streams, for instance location-based data, enable new services and applications to persist across environments.

Previous research from our labs have investigated location-awareness [8, 20, 21] and instrumented objects and environments [6, 19, 11]. In this chapter, we focus on wearable technologies that sense user behavior, applications that leverage such sensing, and the challenges in deploying these types of intelligent systems in real world environments. In particular, we discuss technology and applications that continuously sense and infer physical activities.

In this chapter, we discuss our efforts over five years to implement, iterate, and deploy a wearable mobile sensing platform for human activity recognition. The goal of this system was to encourage individuals to be physically active. We use technology both to automatically infer physical activities as well as employ persuasive strategies [7] to motivate individuals to be more active. This general area is of growing interest to the human-computer interaction and ubiquitous computing research communities, as well as the commercial marketplace. We highlight some of the key issues that must be considered if such systems are to become successfully integrated into real world human activity detection based on our experiences.

2 Technologies for Tracking Human Actvities

We are interested in building mobile platforms to reliably sense real world human actions and in developing machine learning algorithms to automatically infer high-level human behaviors from low-level sensor data. In 2.2, we briefly discuss several common systems and approaches to collect such data. In 2.3, we outline our particular implementation of the Mobile Sensing Platform (MSP) and an application built using this MSP – UbiFit Garden.

2.1 Methods for Logging Physical Activity

Several technologies used to sense human physical activity employ a usage model where the technology is used only while performing the target activity. These technologies include Dance Dance Revolution, the Nintendo Wii Fit, the Nike+ system, Garmin's Forerunner, Bones in Motion's Active Mobile & Active Online, bike computers, heart rate monitors, MPTrain, Jogging over a distance, shadowboxing over a distance, and mixed- and virtual-reality sports games [17, 16, 14, 15, 10, 12].

Perhaps the most common commercial device that detects physical activity throughout the day is the pedometer–an on-body sensing device that detects the number of "steps" the user takes. The usage model of the pedometer is that the user clips the device to his or her waistband above the hip, where the pedometer's simple "inference model" counts alternating ascending and descending accelerations as steps. This means that any manipulation of the device that activates the sensor is interpreted as a step, which often leads to errors. Another more sophisticated commercial on-body sensing device that infers physical activity is BodyMedia's SenseWear Weight Management Solution. The SenseWear system's armband monitor senses skin temperature, galvanic skin response, heat flux, and a 2-d accelerometer to infer energy expenditure (i.e., calories burned), physical activity duration and intensity, step count, sleep duration, and sleep efficiency. SenseWear's inference model calculates calories burned and exercise intensity; aside from step count, it does not infer specific physical activities (e.g., running or cycling).

However, as we have learned from our own prior research [3], a common problem when designing systems based on commercial equipment is that the systems are often closed. That is, in our prior work that used pedometers, users had to manually enter the step count readings from the pedometer into our system, as our system could not automatically read the pedometer's step count. Several researchers have recognized this problem, which has led to a new generation of experimental sensing and inference systems.

One approach that is being used is to infer physical activity from devices the user already carries/wears, such as Sohn et al.'s [20] software for GSM smart phones that uses the rate of change in cell tower observations to approximate the user's daily step count. Shakra [12] also uses the mobile phone's travels to infer total "active" minutes per day and states of stationary, walking, and driving.

A different approach that is being used to detect a wider range of physical activities such as walking, running, and resistance training is to wear multiple accelerometers simultaneously on different parts of the body (e.g., wrist, ankle, thigh, elbow, *and* hip) [1]. While this approach has yielded high accuracy rates, it is not a practical form factor when considering all-day, everyday use. Another approach uses multiple types of sensors (e.g., accelerometer, barometer, etc.) worn at a single location on the body (e.g., hip, shoulder, *or* wrist) [9, 2]. Such multi-sensor devices are more practical for daily use, while still capable of detecting a range of activities, and are thus the approach that we have chosen to use in our work.

2.2 The Mobile Sensing Platform (MSP) and UbiFit Garden Application

Unlike many of the above technologies, our work focuses on detecting physical activities throughout the day, rather than during single, planned workout sessions. This requires that the sensing technology be something that the individual can wear throughout the day and that disambiguates the target activity(ies) from other activities that are performed throughout daily life.

The UbiFit Garden system uses the *Mobile Sensing Platform (MSP)* [2] to automatically infer and communicate information about particular types of physical activities (e.g., walking, running, cycling, using the elliptical trainer, and using the stair machine) in real-time to a glanceable display and interactive application on the phone. The MSP is a pager-sized, battery powered computer with sensors chosen to facilitate a wide range of mobile sensing applications 1.

The sensors on the MSP can sense: motion (accelerometer), barometric pressure, humidity, visible and infrared light, temperature, sound(microphone), and direction (digital compass). It includes a 416MHz XScale microprocessor, 32MB RAM, and 2GB of flash memory (which is bound by the storage size of a removable miniSD card) for storing programs and logging data 2. The MSP's Bluetooth networking allows it to communicate with Bluetooth-enabled devices such as mobile phones.

The MSP runs a set of boosted decision stump classifiers [18, 22] that have been trained to infer walking, running, cycling, using an elliptical trainer, and using a stair machine. The individual does not have to do anything to alert the MSP or interactive application that she is starting or stopping an activity, provided that the MSP is powered on, communicating with the phone, and being worn on the individual's waistband above her right hip (the common usage model for pedometers). The MSP automatically distinguishes those five, trained activities from the other activities that the individual performs throughout her day 3.

The inferences for those five types of physical activities are derived from only two of the MSP's sensors: the 3-axis accelerometer (measuring motion) and the barometer (measuring air pressure). The barometric pressure sensor can be used to detect elevation changes and helps to differentiate between activities such as walking and walking up or down. The sensor data is processed and the activity classification is performed on the MSP itself, the inference results are then communicated via Bluetooth to a mobile phone that runs the interactive application and glanceable display 4. The MSP communicates a list of activities and their predicted likelihood values to the phone four times per second. This means that the MSP and phone must

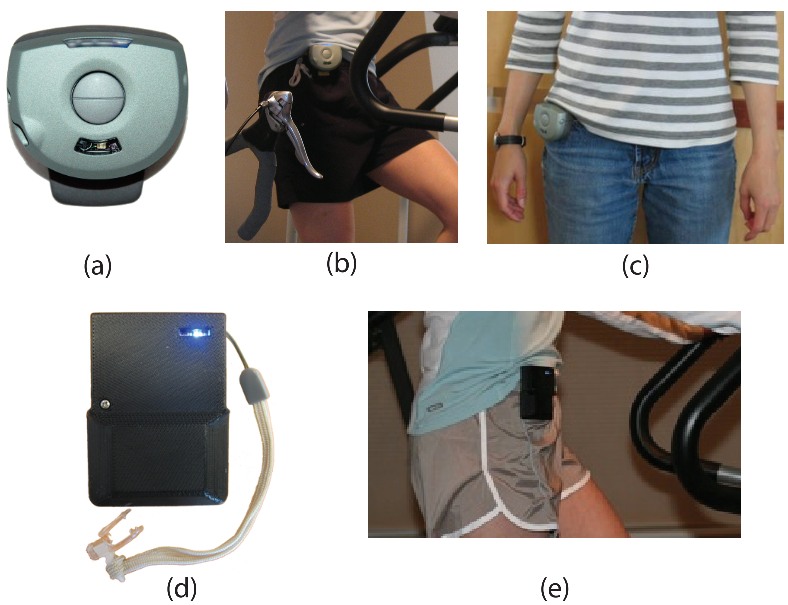

Fig. 1 The Mobile Sensing Platform (MSP) prototypes.(a) The MSP in its original gray case; (b) the gray case MSP as worn by a woman while using the elliptical trainer; (c) the gray case MSP worn by the same woman in casual attire; (d) the MSP in its redesigned black case; and (e) the black case MSP as worn by the same woman while using the elliptical trainer.

be within Bluetooth range at all times during the performance of the activity. The interactive application then aggregates and “smoothes” these fine-grain, noisy data resulting in “human scale” activities such as an 18-minute walk or 45-minute run. The activity’s duration and start time appear in the interactive application about six to eight minutes after the individual has completed the activity 4. “Tolerances” have been built into the definitions of activities to allow the individual to take short breaks during the activity (e.g., to stop at a traffic light before crossing the street during a run, walk, or bike ride).

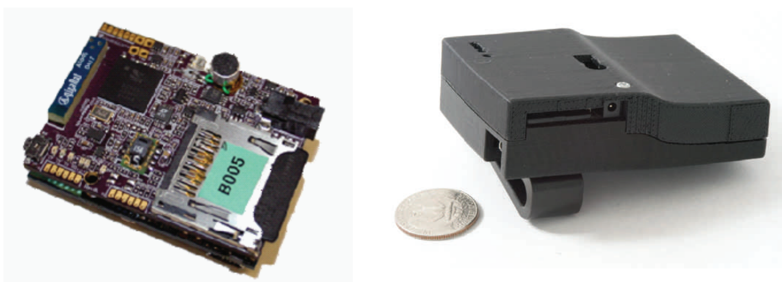

Fig. 2 The Mobile Sensing Platform (MSP) consists of seven types of sensors, Xscale processor, Bluetooth radio and Flash memory in pager size casing.

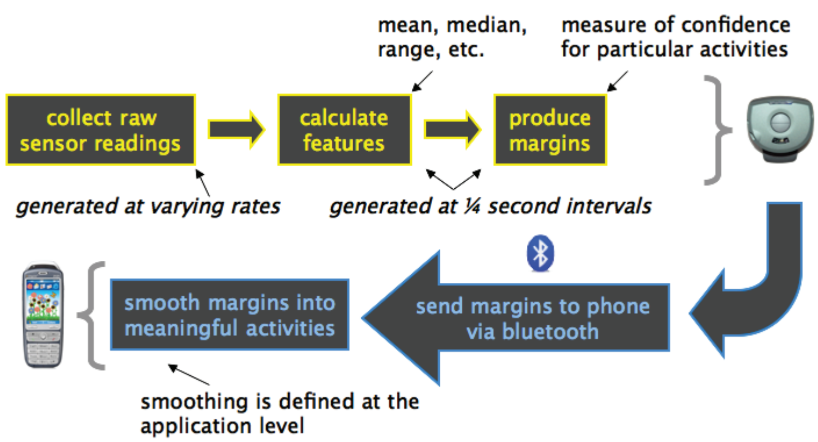

Fig. 3 Inferring activities from sensor readings

UbiFit defines the minimum durations and tolerances for activities to appear in the interactive application's list. Walks of five or more minutes automatically post to the interactive application, as do cardio activities of eight or more minutes. However, to receive a reward for the activities performed (i.e., a new flower appearing on the glanceable display), each instance of the activities must be at least 10 minutes in duration. This means that activities may appear in the individual's list that do not map to flowers on the glanceable display (e.g., an inferred 8-minute walk would appear in the interactive application, but would not have a corresponding flower in the glanceable display). If the MSP is reasonably confident that the individual performed *an activity*, but cannot determine *which activity,* the interactive application will pop up a question asking the individual if she performed an activity that should be added to her journal.

Fig. 4 (a) Screen shot of UbiFit Glanceable Display; (b) UbiFit Garden's interactive application showing two automatically inferred walks.

2.3 User Studies

Two field trials of UbiFit Garden (with 12 and 28 participants respectively) helped illustrate how activity monitoring technologies [5, 4] fit into everyday experiences. Participants recruited were from the Seattle metropolitan area and were regular mobile phone users who wanted to increase their physical activity. They received the phone, fitness device, and instructions on how to use the equipment. The participants agreed to put their SIM cards in the study phone and use it as their personal phone throughout the study. They set a weekly physical activity goal of their own choosing which had to consist of at least one session per week of cardio, walking, resistance training, *or* flexibility training. The participants were interviewed about their experiences in the study and also given the opportunity to revise their weekly physical activity goal.

The intent of the second experiment was to get beyond potential novelty effects that may have been present in the three-week field trial and to systematically explore the effectiveness of the glanceable display and fitness device components through the use of experimental conditions. Additionally, whereas our three week study focused on reactions to activity inference and the overall concept of UbiFit Garden, the second field experiment specifically investigated the effectiveness of the glanceable display and fitness device as a means of encouraging awareness and behavior of physical activity.

3 Usability, Adaptability, and Credibility

Based on the results of our field trials, we present in this section three fundamental aspects of intelligent systems that we believe critically impacts the success of mobile activity monitoring systems. The overall *usability* of the system from the end user perspective includes not only the user interface itself through which the data is communicated, modified, or monitored; usability of such systems also encompasses the underlying infrastructure to support data communications and power management (in particular, when users need to diagnose or detect and correct errors or failures). The *adaptability* of the system includes learning particular user patterns and adapting across usage contexts (noisy environments, outdoor vs. indoor, etc.). Finally, the system performance, usability, adaptability, and accuracy combine to directly impact the overall *credibility*. Credibility means that end users build a mental model of the system that enables them to feel it is reliable, produces expected and accurate data, and when errors do occur, users can mentally “explain” the causes for these errors and take appropriate corrective actions. There is an assumption that such errors (and corrective actions) improve the system over time. A lack of credibility is typically reflected in abandoning technology. We discuss specific examples of each of these from our experiences with the MSP and UbiFit Garden deployments.

3.1 Usability of Mobile Inference Technology

When creating smart technologies such as the one described above, a number of significant basic challenges must be addressed. The system must provide value to motivate its use and this value must outweigh the costs of using such a system. Fundamental usability issues include not only how wearable the technology is, but how to handle any variations in connectivity smoothly, how long the system can run between charges, how accurate the underlying inference is, the extent of personalization or user-specified training data required, and the design of the user interface that provides these intelligent services or applications to consumers. We briefly discuss our experiences with these issues below.

3.1.1 Form Factor and Design

Despite many design iterations over two years and pilot testing within our research lab, participants in the two field studies of UbiFit Garden complained about the form factor of the MSP prototypes. This result was not surprising given the early nature of the prototypes and the fact that several members of the research team had used the full UbiFit Garden system for up to two years and were well aware of the MSPs limitations (too well informed on technical constraints perhaps). Both the original gray case and redesigned black case versions of the MSP were large, still slightly too heavy at 115g (though lighter than many other comparable devices), bulky, somewhat uncomfortable (e.g., the prototypes poked several participants in the hip), drew attention from others (particularly the bright LED), occasionally pulled participants' waistbands down (particularly females during high-intensity activities), and did not last long enough on a single charge (the battery life was approximately 11.5 hours for the gray case MSPs and 16 hours for the black case MSPs) (see Figure 1). (The main difference between these two cases was that the 2nd version contained two batteries instead of a single battery thereby increasing time between charges.) A participant explained:

"It was a lot of machinery to carry around, to be perfectly honest. Like it's just trying to remember to plug in the one [MSP] when I also had to remember to plug in my phone and I mean, I'm just never on top of that stuff and so then after I plugged it in and I'd grab my phone, I'd forget, you know, some days forget the device [the MSP]. And it [the MSP] would pull my pants down when I would be running or doing something more vigorous like that."

Some participants in the 3-month field experiment joked that when they transitioned from the gray case to the black case MSP, they upgraded from looking as if they were wearing an industrial tape measure on their waist to looking like a medical doctor from the 1990's. In spite of its "dated pager" appearance, all participants preferred the black case, and despite the complaints, most participants liked the idea of the fitness device and understood that they were using an early-stage prototype, not a commercial-quality product.

3.1.2 Power and Connectivity Issues

Many intelligent applications or services assume users have continuous network connectivity. While our deployments ran in urban environments with extensive GSM/cell phone, WiFi and WiMax network coverage and we continuously tested Bluetooth connectivity between sensor platform and communication platform (i.e., MSP to cell phone), nevertheless there were times when some portion of the *user-perceived communications* failed. Note that most end users are not experienced at diagnosing and troubleshooting such problems –the system simply appears broken.

If network connectivity was temporarily lost, our system continued running in a local mode and would re-synch when able without user intervention. Similarly, if the Bluetooth connection failed, we could again continue running continuous (in this case sensing motion data and inferring activity) and we stored the data locally on the MSP storage card until Bluetooth connectivity was restored (often by prompting the user on the cell phone to power on/off the devices or to verify that they were both in range). If this connectivity break occurred while a participant was performing an activity, the activity often either did not appear on the phone at all or appeared with errors (e.g., appearing as two activities instead of one, having an error in start time, and/or having an error in duration). At the time this work started in 2003, we did anticipate that cell phones would eventually have integrated sensors and thus having a separate sensor platform was a means of prototyping and testing such systems until sensor-equipped cell phones are more prevalent. Clearly such integration would solve one communication issue.

When designing systems that rely upon several battery powered components (in our case the MSP and the cell phone), the usability of the overall system depends upon its weakest link (or in this case most-power hungry link). Cell phone design has evolved to optimize battery life and power consumption but typically this assumes far less data-intensive applications and less data communication. Early versions of this system sent raw data from the sensor platform to the phone (a power greedy use of Bluetooth). After migrating the inference algorithms to the embedded MSP platform, we could send inferred activity data and timestamps and we did some data packet optimization to reduce the amount of Bluetooth communication required thereby improving the power performance to about 8 hours. However, for smart services and applications that are expected to run in real world scenarios, this 8 hour limit results in an almost unusable system. We initially gave participants in our studies two re-chargers (one for work and one for home) to facilitate mid-day recharging but this is not a realistic expectation. While this works well for pilot testing and research, real world deployments cannot rely on 8-hour recharging strategies. Realistically these technologies must run for between 12 and 16 hours (i.e., during waking hours) with overnight recharges. We went through a number of code optimizations, altered data packet sizes, and ultimately built a second version of the MSP that could hold two cell phone batteries to attain a 12 hour power life.

In our experiences, even with integrated sensor platforms we still believe that the data intensive nature of the communications will have significant impact on usability since current devices have not been optimized for this. This is particularly an issue when the types of intelligent systems require real time inference and continuously sensed data (even at low data rates).

These are the types of problems that are inherent in field studies of early stage, novel technologies. Despite the problems with the MSP prototypes, the two field studies provided invaluable insights into our UbiFit Garden system and on-body sensing and activity inference in general. See [5, 4] for more details.

3.1.3 Accuracy and Generalizability

We want our system to be based on general enough models that it would work for most people "out-of-the-box" without users having to supply personalized training data.

In previous work, the MSP has been shown to detect physical activities with about 85% accuracy [9]. For the field trials mentioned in the previous section, we re-trained and tuned the activity inference models to increase the accuracy for detecting walking and sitting using labeled data from 12 individuals of varying ages, heights, and gender (none were study participants). A wide range of participant types were chosen to ensure that our model parameters did not overfit to a specific sub-type of user. A Naïve Bayes model took several samples of the output from the boosted decision stump classifier to produce the final classification output. This resulted in a smoother temporal classification and was simpler to implement on the embedded device with results comparable to an HMM-based model (e.g., [9]).

In trying to iterate and improve system accuracy, we found it helpful to take each target activity (for instance walking) and create a list of as many sub-classes of the target activity as we could. This enabled us to more precisely label our training data by sub-class first (e.g., walking uphill, walking down hill, walking fast, walking leisurely, walking in high heels, etc.) and then these sub-classes were merged into the higher level activity of walking. Even with variations in personal interpretations of what "walking fast" actually means, we found this strategy produced better models with much shorter training data samples needed.

The value of these cascading 2-level models is best illustrated by issues we had classifying cycling when we used a single-level model. Our initial training data samples were of “bike rides” (one label) where this comprised very different sub-classes (i.e., coasting without pedaling, stopping and standing at a street light for a moment, standing in pedals out of seat to pedal uphill, etc.). We did not separate out these sub-classes and therefore some instances of biking appeared to be confused with sitting or standing. We believe using a 2-level model with sub-classes of training data more precisely labeled would have reduced this ambiguity – this is not yet implemented in our system.

3.2 Adaptability

Progressively more applications are reflecting some degree of end user adaptation over time as devices and systems become more personalized and more intelligent about user context. Recent menu item selections may appear immediately while seldom used menu items remain temporarily hidden unless the user does a more deliberate and prolonged open menu request. Recent documents, links, phone numbers called, or locations visited can be short-listed for convenience. In this section, we briefly discuss some of the issues we encountered in trying to create a flexible and adaptive system based on user context and real world needs.

3.2.1 Facilitating a Flexible Activity Log

Our system was designed to automatically track six different physical activities, however, people’s daily routines often include other forms of activities which they would like to also account for but for which we have no models (e.g., swimming). While we could continue building and adding new models for each new type of activity, it is more immediately practical to allow users to manually journal activities that are out of scope of what we could automatically detect. This enables us to create a single time stamped database for any physical activity regardless of whether or not we have current automated models. When users enter these new activities, they can incorporate them into the glanceable display and "get credit" for them as part of a single user experience. This combination of manual data entry combined with the automatic journaling seemed to give users the feeling that the overall system was more adaptable to their needs.

In addition, because this was a wearable system and the feedback was incorporated into the users' own cell phone background screen, the system was operational on both planned physical workouts (e.g., going for a 30 minute run) and unplanned physical activities (e.g., walking to a lunch restaurant). This flexibility meant that users were more motivated to carry the device all the time in order to get credit for all the walking and incidental physical activity they did.

Finally, we included a deliberately ambiguous "active" item for cases when the system detected some form of physical activity but confidence margins were too low to speculate upon an exact activity type. In this way, the system could provide a time stamped indicator to the user that something had been detected (rather than leaving out the item entirely). For our 3-week field trial, half of the participants received a generic "active" flower for these events; they had the option to either edit these items and provide more precise labels or they could leave them as generic "active" flowers. The other half received a questionnaire on their cell phones asking them if they had just performed an activity that should be added to their journal. Based on experiences in the 3-week field trial, we employed the latter strategy (i.e., the questionnaire) only in the 3-month field trial. Again, this reinforced the notion that the system was adaptable to varying degrees of uncertainty and communicated this uncertainty to the end user in a reasonable and interpretable way.

3.2.2 Improving Accuracy During Deployment

Semi-supervised and active-learning techniques can be particularly useful for personalizing activity models to a specific user's behavior. The system can adapt the model parameters according to the user's unlabeled data, even after the system is deployed. We have seen some significant performance improvement using semi-supervised methods [13]. (For the deployment system described above, this research was still underway and thus these methods have not yet been incorporated into the embedded system). Based on the results thus far, we believe that active learning methods can be used to query the users for new labels that could further improve the performance of the recognition module.

3.2.3 Flexible Application Specific Heuristics

As we described earlier, we used cascading models to predict the moment-to-moment physical activity of the user wearing the MSP device. However, for most intelligent systems, the final application goal will determine the precision and data types and sampling rates necessary to infer meaningful contexts.

We created a number of software configurable parameters for application developers that would allow us to improve accuracy while preserving application meaningful activity distinctions. Even though more sophisticated modeling techniques can be used to group the activity episodes we found from our studies, the notion of episodes and how much gaps in activity should be tolerated is very subjective. So, we developed a more easily configurable mechanism to facilitate this. In fact the parameters are read in from a simple text file so, in theory, a knowledgeable end user might be able to set these.

While we sample and infer activity four times per second, clearly for detecting physical activities the second-to-second variations over the course of a day do not tend to meaningfully map to what users might consider “episodes”. In fact, if the system indicates a 2 second bike ride preceded and followed by 5 minutes of walking it is almost certain that this resulted from noise in the data rather than an actual 2 second bike ride. The duration or number of confidence margins to be considered as a group can be set. In an application such as the one described earlier, we typically group the activity inferences and their associated confidence margins. Within this group, we provide the option to set a threshold that indicates the number of items that must be above a specified confidence margin in order to successfully categorize that activity. For instance, 18 of 20 activities must have the same label and be above a confidence margin of 80%. In this way, we can use one general platform for a variety of applications and easily configure it in a way to minimize both noise in the data and the amount of precision required.

Finally, we have a “human scale” activity smoothing parameter which lets us easily specify how long an activity needs to be sustained for a valid resulting “episode” and what permissible gaps can occur within this episode. For instance, if the application goal is to track sustained physical activity, we might want a minimum duration of 10 minutes per activity type but we may allow up to 60 seconds gap to accommodate things like jogging and pausing at a street light before resuming the same jogging activity.

This layered approach to data smoothing and the ability for either application developers or even end users to tune and configure these parameters is a critical element if the system is to be flexible enough to support more than one potential target application.

3.3 Credibility

A user's level of confidence in the system is influenced by the reliability of technology and by the accuracy of the inferred activities. This is especially noticeable to end users if the activity inference results are visible in real time. At the heart of most inference-based systems, the inferred activities (or locations or contexts or identities of users) are probabilistic in nature. This uncertainty is often not reflected in the user interface. Also such systems can introduce false positives (where events of interest did not really occur) or false negatives (where events did occur but were not reported). The amount of such errors, the ability to correct them, and the ability of users to diagnose or explain these errors are critical in establishing a credible system. We propose that intelligent systems may inherently require a different style of user interface design that reflects these characteristic ambiguities.

3.3.1 Correcting or Cheating

One of the current debates in automatically logging physical exercise data is whether or not users should be allowed to edit the resulting data files. Since these data sets are often shared as part of a larger multi-user system, a number of commercial systems do not allow users to add or edit items presumably in the interests of preserving data integrity and reducing the ability to cheat. However, our experiences indicate that users are more often frustrated when they cannot correct erroneous data or add missing data.

We made a deliberate decision in our system to allow users to add new items, delete items or change the labels on automatically detected items (in the event of misclassification). However, we limited the time span during which users could alter data to the last 2 days of history only. This seemed a reasonable compromise to both allow users to manually improve the data accuracy and completeness (thereby improving their perceptions of system credibility) while limiting this to a time span where they would have more accurate recall of their actual activities. Additionally, the system was not used in the context of a multi-user application (i.e., they did not have workout “buddies” or a competition/collaboration mechanism built in) so the users had less of an inclination to cheat.

3.3.2 Use of Ambiguity of User Interfaces

Using ambiguity in either the graphic representations and/or in “labels” as presented to users can help maintain system credibility when algorithm confidence margins for any one activity are low and hence uncertainty is high. For instance, we provided users with feedback using a generic “active” label (and a special flower) when some form of physical activity was detected but the algorithm could not determine specifically which type of activity it was. Rather than incorrectly selecting a lower probability activity or presenting no feedback at all, this generic but higher level activity class (active versus inactive or sedentary) indicated that the system had detected some user movements. As a consequence, users reporting feeling better that the system detected something at least and they could later edit this vague activity to be more precise (or not). This was preferred over not indicating any activity.

We have additionally considered an option to display the top *n* activities (based on confidence margins) where users can pick from the list or else indicate “other” and specify the activity if it is not in the list. This feature was not implemented in the system used in our reported field trials. Previous experience with mobile phone user interfaces [e.g., Froehlich et al, 2006] suggests that such lists would need to fit within one screen (i.e., less than 10 items). This remains an area to be explored to weigh off user effort against data accuracy.

3.3.3 Learning from User Corrections

We anticipate that using an initial model and then learning from user correction will impact overall system credibility. Users experience frustration if they must repeatedly correct the same types of error. By applying active learning techniques we can take advantage of the additional data supplied by these user corrections to uniquely personalize the models of physical activity to better match those of a particular user. Over time, such corrections should become less frequent thereby improving perceived accuracy and credibility of the system. Such personalization assumes these systems are not multi-user or shared devices unless the models can be distinguished per user based on some other criterion, for instance through login information or biometrics.

Our experiences suggest that when systems appear to be “good” at automatically categorizing some types of activities, users might assume the system has abilities to learn to be “smarter” in other situations or for other categories. Thus, there is an expectation that error corrections will improve system accuracy over time even when users are told there is no learning capability.

4 Conclusions

The task of recognizing activities from wearable sensors has received a lot of attention in recent years. There is also a growing demand for activity recognition in a wide variety of health care applications. This chapter provides an overview of a few key design principles that we believe are important in the successful adoption of mobile activity recognition systems. Two real-world field trials allowed us to systematically evaluate the effectiveness of the different design decisions. Our findings suggest that the main design challenges are related to the usability, adaptability, and credibility of the system being deployed. Furthermore, we believe that some of the design considerations are applicable to scenarios beyond mobile activity tracking and will be relevant for designing any context-aware system that provides real-time user feedback and relies on continuous sensing and classification.

Acknowledgements Thanks to Gaetano Borriello, Mike Chen, Cherie Collins, Kieran Del Pasqua, Jon Froehlich, Pedja Klasnja, Anthony LaMarca, James Landay, Louis LeGrand, Jonathan Lester, Ryan Libby, David McDonald, Keith Mosher, Adam Rea, Ian Smith, Tammy Toscos, David Wetherall, Alex Wilkie, and many other family, friends, colleagues, and study participants who have contributed to this work.

References

1. Bao, L. & Intille, S.S., "Activity Recognition from User-Annotated Acceleration Data," In *Proceedings of Pervasive '04*, (2004), 1-17.
2. Choudhury, T., Borriello, G., Consolvo, S., Haehnel, D., Harrison, B., Hemingway, B., Hightower, J., Klasnja, P., Koscher, K., LaMarca, A., Landay, J.A., LeGrand, L., Lester, J., Rahimi, A., Rea, A., & Wyatt, D. (Apr-Jun 2008). "The Mobile Sensing Platform: An Embedded Activity Recognition System," *IEEE Pervasive Computing Magazine Special Issue on Activity-Based Computing, 7(2),* 32-41.
3. Consolvo, S., Everitt, K., Smith, I., & Landay, J.A. (2006). "Design Requirements for Technologies that Encourage Physical Activity," *Proceedings of the Conference on Human Factors and Computing Systems: CHI 2006*, *(Montreal, Canada)*, New York, NY, USA: ACM Press, 457-66.
4. Consolvo, S., Klasnja, P., McDonald, D., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher, K., & Landay, J.A. (Sept 2008). "Flowers or a Robot Army? Encouraging Awareness & Activity with Personal, Mobile Displays," *Proceedings of the 10th International Conference on Ubiquitous Computing: UbiComp 2008, (Seoul, Korea)*, New York, NY, USA: ACM Press
5. Consolvo, S., McDonald, D.W., Toscos, T., Chen, M.Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., Smith, I., & Landay, J.A. (Apr 2008). "Activity Sensing in the Wild: A Field Trial of UbiFit Garden," *Proceedings of the Conference on*

Human Factors and Computing Systems: CHI 2008, (Florence, Italy), New York, NY, USA: ACM Press, 1797-806.

6. Fishkin, K. P., Jiang, B., Philipose, M. and Roy, S. I Sense a Disturbance in the Force: Long-range Detection of Interactions with RFID-tagged Objects. *Proceedings of the 6th International Conference on Ubiquitous Computing: Ubicomp 2004*, pp. 268-282.
7. Fogg, B.J. (2003). *Persuasive Technology: Using Computers to Change What We Think and Do*. San Francisco, CA, USA: Morgan Kaufmann Publishers.
8. Hightower, J., Consolvo, S., LaMarca, A., Smith, I. E., Hughes, J. Learning and Recognizing the Places We Go. *Proceedings of the 7th International Conference on Ubiquitous Computing: Ubicomp 2005*: 159-176.
9. Lester, J. Choudhury, T., & Borriello, G. (2006). "A Practical Approach to Recognizing Physical Activities," In *Proceedings of the 4th International Conference on Pervasive Computing: Pervasive '06*, Dublin, Ireland, 1-16.
10. Lin, J.J., Mamykina, L., Lindtner, S., Delajoux, G., & Strub, H.B. (2006). "Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game," In *Proceedings of the 8th International Conference on Ubiquitous Computing: UbiComp '06*, Orange County, CA, USA, 261-78.
11. Logan, B., Healey, J., Philipose, M., Munguia-Tapia, E., Intille. S. Long-Term Evaluation of Sensing Modalities for Activity Recognition. *In Proceedings of Ubicomp 2007*. Innsbruck, Austria, September 2007
12. Maitland, J., Sherwood, S., Barkhuus, L., Anderson, I., Hall, M., Brown, B., Chalmers, M., & Muller, H. "Increasing the Awareness of Daily Activity Levels with Pervasive Computing," *Proceedings of the 1st International Conference on Pervasive Computing Technologies for Healthcare 2006: Pervasive Health '06*, Innsbruck, Austria.
13. Mahdaviani. M. and Choudhury, T. Fast and Scalable Training of Semi-Supervised CRFs with Application to Activity Recognition. Appears in the *Proceedings of NIPS 2007*. December 2007.
14. Mueller, F., Agamanolis, S., Gibbs, M.R., & Vetere, F. (Apr 2008). "Remote impact: shadowboxing over a distance," *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, Florence, Italy, 2291-6.
15. Mueller, F., Agamanolis, S., & Picard, R. (Apr 2003). "Exertion Interfaces: Sports Over a Distance for Social Bonding and Fun," *Proceedings of CHI '03*, pp.561-8.

16. Mueller, F., O'Brien, S., & Thorogood, A. (Apr 2007). "Jogging Over a Distance," In *CHI '07 Extended Abstracts*, 1989-94.
17. Oliver, N. & Flores-Mangas, F. (Sep 2006). "MPTrain: A Mobile, Music and Physiology-Based Personal Trainer," In *Proceedings of MobileHCI '06*.
18. Schapire, R.E., "A Brief Introduction to Boosting," Proc. 16th Int'l Joint Conf. Artificial Intelligence (ijcai 99), Morgan Kaufmann, 1999, pp. 1401-1406.
19. Smith, J. R., Fishkin, K., Jiang, B., Mamishev, A., Philipose, M., Rea, A., Roy, S., Sundara-Rajan, K., RFID-Based Techniques for Human Activity Recognition. *Communications of the ACM*, v48, no. 9, Sep 2005.
20. Sohn, T., et al., "Mobility Detection Using Everyday GSM Traces," In *Proceedings of the 8th International Conference on Ubiquitous Computing: UbiComp 2006,* Orange County, CA, USA.
21. Sohn, T., Griswold, W., Scott, J., LaMarca, A., Chawathe, Y., Smith, I.E., Chen, M.,Y., Experiences with place lab: an open source toolkit for location-aware computing. *ICSE 2006*: 462-476.
22. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: *Proc. Computer Vision and Pattern Recognition* (2001).