Activity Sensing in the Wild: A Field Trial of UbiFit Garden

Sunny Consolvo[1, 2], David W. McDonald[2], Tammy Toscos[1], Mike Y. Chen[1], Jon Froehlich[3], Beverly Harrison[1], Predrag Klasnja[1, 2], Anthony LaMarca[1], Louis LeGrand[1], Ryan Libby[3], Ian Smith[1], & James A. Landay[1, 3]

[1]Intel Research Seattle
Seattle, WA 98105 USA
[sunny.consolvo, beverly.harrison, anthony.lamarca, louis.l.legrand]@intel.com, ttoscos@indiana.edu, mike@ludic-labs.com, iansmith@acm.org

[2] The Information School
DUB Group
University of Washington
Seattle, WA 98195 USA
[consolvo, dwmc, klasnja]@u.washington.edu

[3] Computer Science & Engineering
DUB Group
University of Washington
Seattle, WA 98195 USA
[landay, jfroehli, libby]@cs.washington.edu

ABSTRACT
Recent advances in small inexpensive sensors, low-power processing, and activity modeling have enabled applications that use on-body sensing and machine learning to infer people's activities throughout everyday life. To address the growing rate of sedentary lifestyles, we have developed a system, UbiFit Garden, which uses these technologies and a personal, mobile display to encourage physical activity. We conducted a 3-week field trial in which 12 participants used the system and report findings focusing on their experiences with the sensing and activity inference. We discuss key implications for systems that use on-body sensing and activity inference to encourage physical activity.

Author Keywords
persuasive technology, sensing, activity inference, mobile phone, ambient display, fitness, activity-based applications.

ACM Classification Keywords
H.5.2 User Interfaces, H.5.m Miscellaneous.

INTRODUCTION
Recent advances in small inexpensive sensors, low-power processing, and activity modeling have enabled new classes of technologies that use on-body sensing and machine learning to automatically infer people's activities throughout the day. These emerging technologies have seen success with participants in controlled and "living" lab settings [11] and with researchers in situ [18]. The next step is to conduct in situ studies with the target user population. Such studies expose important issues, for example, how the systems are used as part of everyday experiences, where the technology is brittle, and user reactions to activity inference and the presentation of those inferences.

One application domain for on-body sensing and activity inference is addressing the growing rate of sedentary lifestyles. Regular physical activity is critical to everyone's physical and psychological health, regardless of their being normal weight, overweight, or obese [6,16]. Physical activity reduces risk of premature mortality, coronary heart disease, type II diabetes, colon cancer, and osteoporosis, and has also been shown to improve symptoms associated with mental health conditions such as depression and anxiety. Yet despite the importance of physical activity, many adults in the U.S. do not get enough exercise [1].

Technologies that apply on-body sensing and activity inference to the fitness domain are faced with a challenge regarding which physical activities should be detected. The American College of Sports Medicine (ACSM) recommends that physical activity be regular and include *cardiorespiratory training* (or "cardio") where large muscle groups are involved in dynamic activity such as running or cycling; *resistance training*, that is weight training that builds muscular strength and endurance; and *flexibility training* where muscles are slowly elongated to improve or maintain range of motion [22]. Technologies that attempt to encourage physical activity should support the *range of activities* that contribute to a physically active lifestyle, rather than focus on a single activity such as walking.

Our goal in this work is to investigate users' experiences with a system that we have developed, UbiFit Garden, which uses on-body sensing, activity inference, and a novel personal, mobile display to encourage physical activity. While our future work will focus on how the system affects awareness and sustained behavior change, at this stage, we are exploring how the system affects individuals' everyday lives, how they interpret and reflect on the data about their physical activities, and how they interact with that data. We conducted a three-week field trial (n=12) with participants who were representative of UbiFit Garden's target audience. In this paper, we discuss the types of physical activities participants performed, how those activities were recorded and manipulated, and participants' qualitative reactions to activity inference and manual journaling. We also discuss participants' general reactions to the system.

We begin with a discussion of related work that uses on-body sensing and inference to encourage physical activity. We follow with a description of the UbiFit Garden system as it was deployed during the field trial. Next, we discuss the methods we employed during the three-week field trial, followed by key results. We conclude with future work.

TECHNOLOGIES TO ENCOURAGE PHYSICAL ACTIVITY

Technology has long been used to track and encourage physical activity, from heart rate monitors and bicycle odometers to web sites that support goal setting and self-monitoring. We outline the main classes of technologies in this domain and discuss how UbiFit Garden fits within the landscape. Our discussion of related work is grouped into three categories: technologies used to detect a specific, pre-planned physical activity, commercial devices that detect physical activity throughout the day, and experimental prototypes that detect physical activity throughout the day. We discuss representative technologies from each category.

Detecting a specific pre-planned physical activity

Several technologies to encourage physical activity are used *only* while performing the target activity and are not trying to disambiguate that activity from others performed throughout everyday life. Such 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* [17], *Jogging over a distance* [15], and mixed- and virtual-reality sports games [13,14].

Physical activity detection with commercial devices

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. Pedometers are usually clipped to the user's waistband and use a simple "inference model" in which alternating ascending and descending accelerations are counted as steps. This means that any manipulation of the device that activates the sensor is interpreted as a step.

The *Nokia 5500 Sport* mobile phone has an on-board accelerometer that detects running and walking when "Sports" mode is turned on and the phone is worn on the user's waist in the accompanying holder. The "Sports" mode helps the phone distinguish physical activity from other activities. The 5500 Sport includes a diary for planning and tracking workouts as well as enabling users to add workouts that the phone does not detect.

BodyMedia's SenseWear Weight Management Solution consists of three components: armband monitor, wrist-mounted display, and web site. Skin temperature, galvanic skin response, heat flux, and a 2-d accelerometer infer energy expenditure (i.e., calories burned), physical activity duration and intensity, step count, sleep duration, and sleep efficiency. SenseWear's inference model, which calculates calories burned and exercise intensity, does not infer specific physical activities other than step count.

Four recent research projects have used pedometers to persuade individuals to take more steps each day. *Laura* [3] is an animated conversational agent, displayed on a wall-mounted touch-screen display, who acts as a virtual exercise advisor. *Fish'n'Steps* [10] uses personal goals, social influence, and a non-literal, aesthetic display. The user's step count is linked to the emotional state, growth, and activity of a virtual fish in a virtual fish tank—a tank that includes the fishes of other users. The fish tank is displayed on a public kiosk in the users' offices and on personal web sites for an individual progress view. In *Houston* [5] and *Chick Clique* [21], small groups of users share their step counts and progress toward daily step count goals with each other via their mobile phones. Houston also provides small rewards, such as a message and symbol next to the user's step count, when she reaches her daily goal.

A common problem when designing systems based on commercial equipment is that the closed nature of the devices usually prevents the capture of activity inferences that can be directly used by a prototype (e.g., in the four projects just described, the users or researchers manually entered step count readings into the system). This problem has been recognized, and a new generation of experimental sensing and inference systems is being developed.

Physical activity detection with experimental prototypes

One approach to detecting 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) [2]. While this approach is known for yielding strong 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 (e.g., hip, shoulder, *or* wrist) [9]. Such multi-sensor devices are more practical for daily use, while still capable of detecting a range of activities.

A different approach 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. Shakra employed ideas similar to Houston and Chick Clique, where groups of users exchange physical activity information via mobile phones.

UbiFit Garden draws from several of these projects. For example, encouraging and accounting for a range of physical activities and allowing the user to correct inference mistakes (per recommendations from Houston), prompting the user to engage with the system (as done by Houston), using non-literal, understandable, aesthetic representations of behavior (as done by Fish'n'Steps), providing trending information (as done by Fish'n'Steps, Houston, and Shakra), providing positive reinforcement rather than punishment (drawing from Houston's success and problems found by Fish'n'Steps), providing frequent opportunities for self-reflection (similar to Houston, Chick Clique, and Shakra), and integrating use into everyday life (all projects).

THE UBIFIT GARDEN SYSTEM

Mistress Mary, quite contrary, how does your garden grow? With silver bells, and cockle shells, and marigolds all in a row.
– Frances Hodgson Burnett

We have designed a healthy lifestyle technology, *UbiFit Garden*, which uses on-body sensing, real-time statistical modeling, and a personal, mobile display to encourage regular physical activity. UbiFit Garden is designed for individuals who have recognized the need to incorporate regular physical activity into their everyday lives but have not yet done so, at least not consistently[1].

The UbiFit Garden system consists of three components: (1) fitness device, (2) interactive application, and (3) glanceable display. The *fitness device* automatically infers and communicates information about several types of physical activities to the glanceable display and interactive application. The *interactive application* includes detailed information about the individual's physical activities and a journal where activities can be added, edited, and deleted. The *glanceable display* uses a non-literal, aesthetic representation of physical activities and goal attainment to motivate behavior (Fig 1). It resides on the background screen, or "wallpaper," of a mobile phone to provide a subtle reminder whenever and wherever the phone is used.

We are using an iterative process to design UbiFit Garden. In addition to drawing from prior work (including our own [5]), we conducted a survey (n=75) with respondents from 13 states across the U.S. that covered a range of attitudes and behaviors with mobile devices and physical activity. This survey tested assumptions about the glanceable display and elicited general feedback. Overall, respondents were very positive about the concept and confirmed that the display was understandable. A majority of respondents could imagine using UbiFit Garden. Common concerns had to do with assuming that all exercise data would have to be manually entered into the phone or that the phone would have to be carried during exercise (the fitness device and interactive application were not addressed in the survey).

UbiFit Garden's Fitness Device

UbiFit Garden is part of a larger research program to explore how sensing and activity inference technologies can be applied to real world problems, such as encouraging people

a) b) c)

Figure 1. UbiFit Garden's glanceable display. a) at the beginning of the week—small butterflies indicate recent goal attainments; the absence of flowers means no activity this week; b) a garden with workout variety; c) the display on a mobile phone—the large butterfly indicates this week's goal was met.

to engage in healthy activities. UbiFit Garden relies on the *Mobile Sensing Platform (MSP)* [4,9], a research platform for mobile sensing and inference applications.

The MSP is a pager-sized, battery powered computer with sensors chosen to facilitate a wide range of mobile sensing applications (Fig 2). The MSP's sensors include: 3-d accelerometer, barometer, humidity, visible and infrared light, temperature, microphone, and compass. It includes a 400MHz XScale microprocessor, 32MB RAM, 2GB of flash memory for storing programs and logging data, and a rechargeable lithium ion battery. The MSP's Bluetooth networking allows it to communicate with other Bluetooth-enabled devices such as mobile phones.

UbiFit Garden uses the MSP to automatically infer physical activities in real time. The MSP runs a set of boosted decision stump classifiers that have been trained to infer walking, running, cycling, using an elliptical trainer, and using a stair machine. These inferences are derived from two sensors: the 3-d accelerometer and barometer. The sensor data is processed and the activity is inferred on the MSP, then the inferences are communicated via Bluetooth to a mobile phone that runs the interactive application and glanceable display (Fig 3). The MSP communicates a list of activities and their predicted likelihoods to the phone four times per second. The phone application aggregates and "smoothes" these fine-grain, noisy data resulting in "human scale"

Figure 2. UbiFit Garden's fitness device–the MSP. At left is the MSP worn by a woman; at center is the MSP in its waist-mount case; and at right are the sensor boards inside the casing.

[1] UbiFit Garden targets the contemplation, preparation, and action stages of change of the *Transtheoretical Model*, which describes the stages through which individuals progress to intentionally modify addictive and other problematic behaviors [19].

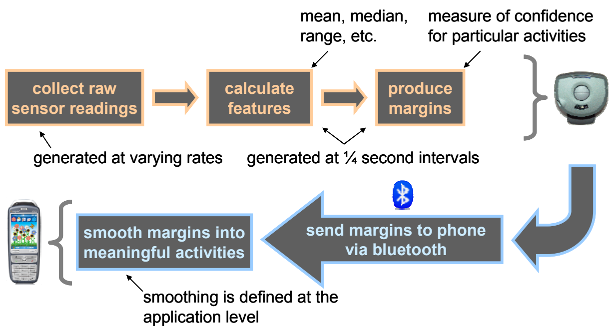

Figure 3. Inferring activities from sensor readings—from raw sensor readings, features are extracted and used to produce likelihood estimates (margins) which are sent to the phone for activity smoothing and use by the application.

activities such as a 22-minute walk or 35-minute run. Given the frequent Bluetooth communication between the phone and MSP, the MSP's battery lasts about 11.25 hours.

While the size of the MSP research prototype is a little large, it could be reduced from pager-sized to pedometer-sized. Alternatively, we anticipate that future mobile phones will contain the required sensors and have sufficient processing power to duplicate the MSP's activity inference functionality, eliminating the need for a separate device. Additional details about the MSP can be found in [4].

UbiFit Garden's Interactive Application

UbiFit Garden includes an interactive application that runs on mobile phones. The application includes details about inferred activities and a journal to add, edit, or delete information about activities, including those not inferred by the fitness device. If nothing has been manually journaled for about two days, a prompt asks if the user has anything to add. Through the application, the user can:

1. View a daily activity list of physical activities performed today and any prior day (Fig 4a);
2. Add, edit, or delete physical activities for today and yesterday including:
   - 2a. Add activities that the fitness device is *not* trained to infer (e.g., yoga or swimming),
   - 2b. Add activities that the fitness device *is* trained to infer but were performed while the user was not wearing the device or if the device experienced a problem (e.g., dead battery), and
   - 2c. Correct mistakes made by the fitness device (e.g., by (i) correcting the type, duration, or start time of an inferred activity, or (ii) deleting an activity that the device inferred, but was not actually performed);
3. View progress toward the weekly goal (Fig 4b); and
4. Add a comment to the daily activity list (e.g., "sick," "hiked Mt. Rainier," "CHI paper deadline," etc.).

The interactive application is built using the MyExperience framework [8], a scripting environment for mobile phones. MyExperience is used by UbiFit Garden to manually journal activities, communicate with the MSP, store activity data,

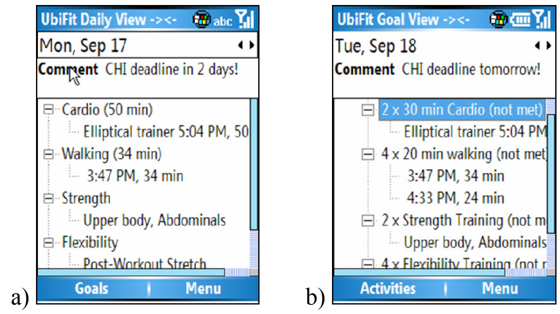

Figure 4. Screen shots from the interactive application. The user's a) daily activity list and b) goal view, showing progress.

update the glanceable display (discussed next), and synchronize study data to a remote server.

Additional features were included based on pilot testing. A set of troubleshooting surveys were added which trigger upon particular technology failures. For example, if the Bluetooth connection between the phone and MSP drops, a dialog alerts the user to the problem and helps them fix it. During the field trial, the research team was automatically notified via text message if the problem was not remedied. Finally, care was taken to avoid disrupting normal phone usage. Prompts and other actions are delayed when the user is on the phone, and incoming calls always receive priority.

UbiFit Garden's Glanceable Display

UbiFit Garden's glanceable display is a non-literal, aesthetic image that presents key information about the user's physical activities and goal attainment. The display resides on the background screen, or "wallpaper," of the user's mobile phone, providing a subtle reminder whenever and wherever the phone is used. Given the high frequency of mobile phone use, the user should see the display often. The glanceable display provides weekly goal attainment status, physical activity behavior, and a subtle but persistent reminder of commitment to physical activity.

The display uses the metaphor of a garden that blooms throughout the week as the user performs physical activities. Upon meeting the weekly goal, a large, yellow butterfly appears near the upper right corner of the display. Smaller, white butterflies represent recent goal attainments, serving to reward and remind the user of past successes (Fig 1). The different types of flowers represent the types of physical activities that the ACSM suggests are important to a well-balanced physical activity routine[2]: cardio, resistance training, flexibility, and walking. With UbiFit Garden's display, a healthy garden represents healthy behavior.

UbiFit Garden uses positive reinforcement—the user is *not* punished for inactivity, for example, with wilting flowers,

[2] The ACSM specifies three types of activities: cardio, resistance, and flexibility. Based on results from [5], UbiFit Garden maintains walking as a separate, fourth category to distinguish it from more vigorous types of cardio activities such as running and cycling.

weeds, or a stormy sky. Instead, the garden will simply be sparse (or empty) with a blue sky and healthy grass (e.g., Fig 1a). If the user does not meet the weekly goal, the garden will simply not have the large butterfly (e.g., Figs 1a & 1b). Each flower represents an individual event (*e.g.,* a 40-minute run and a 3-hour bicycle ride are each represented by one pink cardio flower; a 22-minute walk by one sunflower; a yoga class by one white flexibility flower; a weight lifting session by one blue resistance training flower, etc.). Walking and cardio activities must be at least 10 minutes in duration to receive a flower. A flower's height has no relation to the activity's duration and varies simply for aesthetics.

3-WEEK FIELD TRIAL OF UBIFIT GARDEN

A three-week field trial of UbiFit Garden with 12 participants from the Seattle metropolitan area illustrated how these types of persuasive technologies [7] fit into everyday experiences.

Participants & Method

Six women and six men, aged 25-35, volunteered to participate in the field trial. Eleven participants were recruited by a market research agency and the twelfth by the research team. Participants were regular mobile phone users who wanted to increase their physical activity. They agreed to put their SIM cards in the study phone and use the study phone as their personal phone throughout the study which ran for 21-25 days depending on participants' schedules. None of the participants knew each other.

Study participation included three in-person sessions and at least 21 days of using UbiFit Garden. During the first session (conducted individually with each participant at the start of week 1), participants were interviewed about their attitudes and behaviors as they related to physical activity, and they completed several questionnaires about their barriers to physical activity, current physical activity routines, stage of change [19], demographics, and familiarity with technology. Participants also 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. Each participant had their height and weight measurements taken. They also received the study phone, fitness device, and instructions on how to use the equipment.

In the second and third in-person sessions (conducted individually with participants approximately 7 and 21 days after their first in-person session), participants were interviewed about their experiences in the study. In the second session, participants were also given the opportunity to revise their weekly physical activity goal.

In-person interviews were audio recorded and transcribed. Participants were compensated for their participation; they were encouraged to wear the fitness device, carry the phone, and had to validate a daily activity list. Participants had up to one day later to validate their daily list (e.g., Monday's list could be validated on Monday or Tuesday). Compensation was *not* based on performing activities, wearing the fitness device, or meeting weekly goals.

ID	cardio	walk.	resis.	flex.	other	*totals*
p1	none	6 / 18	19 / 0	11 / 1	0 / 5	*60*
p2	2 / 0	6 / 4	7 / 0	5 / 0	none	*24*
p3	1 / 0	13 / 0	1 / 0	none	1 / 1	*17*
p4	5 / 3	16 / 28	7 / 0	7 / 0	0 / 3	*69*
p5	22 / 4	19 / 12	8 / 0	14 / 0	0 / 5	*84*
p6	3 / 0	1 / 8	23 / 0	none	none	*35*
p7	2 / 11	4 / 18	6 / 0	none	2 / 1	*52*
p8	2 / 5	9 / 10	3 / 0	4 / 0	1 / 0	*34*
p9	none	5 / 20	none	8 / 0	4 / 5	*42*
p10	1 / 5	6 / 7	none	5 / 0	3 / 1	*28*
p11	0 / 1	10 / 7	none	none	1 / 0	*19*
p12	4 / 4	20 / 12	5 / 0	21 / 0	2 / 0	*68*
totals	*42 / 34* *55% / 45%*	*115 / 143* *45% / 55%*	*79 / 0* *100% / 0%*	*75 / 1* *99% / 1%*	*14 / 29* *33% / 67%*	*325 / 207* *61% / 39%*

Table 1. Total number & type of activities performed. On the left of each '/' is the number of manually journaled activities and on the right is the number of inferred activities (the type is the final type after any editing by the participant).

The participants represented a broad range of occupations including marketing specialist, receptionist, elementary school employee, musician, copywriter, director of external affairs for a non-profit agency, professional actor/dancer, film maker/videographer, and software implementer. Eight were employed full-time (one was also a student), two were homemakers, one was employed part-time, and one was self-employed. Six participants were classified as *overweight,* five as *normal*, and one as *obese* according to Body Mass Index calculations performed on their height and weight measurements taken during the first session.

Results

In this section, we focus on results from the three-week field trial as they pertain to the activities participants performed. We discuss the types of activities performed to see if UbiFit Garden supported variety. We also discuss how participants interacted with and reacted to activity inference and manual journaling, and how that affected perceptions about the system, including general reactions.

Types of physical activities performed

In the relatively short period of the three-week field trial, participants performed a wide range of activities, suggesting that UbiFit Garden supports the type of variety recommended by the ACSM. Five participants did each type of physical activity—cardio, walking, resistance, and flexibility—at least once during the field trial, five did three types, and two did only two types. All participants walked, 10 did cardio training, nine resistance, and eight flexibility. The number of activities recorded by each participant ranged from 17 to 84 activities (mean: 44, median: 39), for a total of 532 activities recorded during the field trial. Table 1 shows the number and types of activities performed by each participant, including how activities were recorded.

UbiFit Garden listed a fifth activity type, "Other," in the interactive application. Ten participants recorded at least one activity as "Other." "Housework" or "gardening" were commonly listed. Some participants recorded activities such as rowing, tennis, and dancing. One participant recorded that she was in the pool with her kids, but was not swimming and thus did not consider it to be cardio. Even though these "other" activities did not count toward their weekly goal, most participants liked including the activities in their daily activity list (and they wanted a flower in their gardens—perhaps a smaller, more modest one—to represent these less vigorous physical activities).

Of the five activities that the fitness device infers—walking, running, bicycling, elliptical trainer, and stair machine[3]—participants performed all but stair machine during the field trial. The types of cardio performed that the fitness device was *not* trained to infer (as determined by manually journaled cardio activities), in order of frequency, were: dancing, swimming, hiking, basketball, bouncing, gardening, kayaking, and Ultimate Frisbee. Unlike results from our prior work [5], participants did not feel that UbiFit Garden discouraged them from performing a particular type of activity, such as one not inferred by the fitness device.

Inferences, additions, edits, & deletions

Activities were inferred and journaled throughout the field trial. The number of days on which activities were performed ranged from 10 days (for p3) to 20 days (for p1, p4, p9 and p12), with p5 and p7 closely following at 19 days each. Figure 5 shows on which days during the trial participants performed activities and how those activities were recorded. 61% (325) of activities were manually journaled by participants and 39% (207) were inferred by the fitness device. Table 1 shows the 532 activities broken down by type and how they were recorded.

From the interview data, we know that 35 of the 42 journaled cardio activities were activities for which the device was not trained to infer or was not worn. We do not have data to know whether the remaining seven events or the 115 journaled walks were not inferred because the algorithms or the Bluetooth connection between the phone and device failed, the device was not charged, or if the device was not worn. However, the manual journaling feature enabled participants to record those events.

Inferred activities. Of the 207 inferred activities, most (77%) were left unchanged and the rest were edited for type, duration, and/or start time. Most of the unchanged activities were *walks* (71%). Four of the activities that were

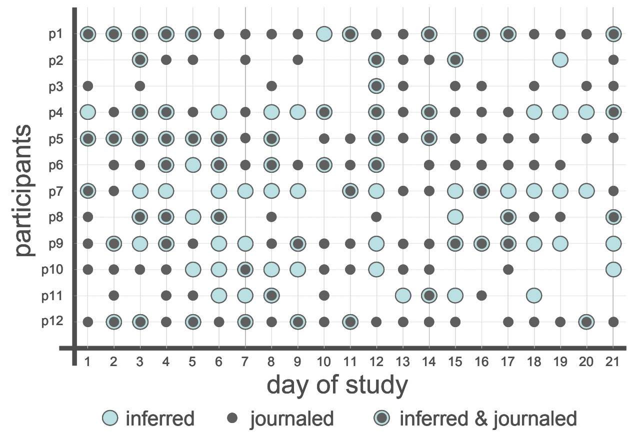

Figure 5. Frequency of performed activities & how recorded for each participant—"performed activities" include inferred (unchanged and edited) and manually journaled activities.

left unchanged were inferred as *stair machine* events—they were actually hilly walks, all for the same participant, who clarified what they were in the comment field of the interactive application. She explained in an interview that she did not edit those activities as she felt they more closely resembled stair machine sessions than walking, as they were hilly and involved lifting her children.

For a majority of the edited activities (80%), the *type* of inferred activity was changed (by eight participants). Of the activities edited for type, 45% were incorrectly detected as *bicycling* events. The incorrect bicycling inferences were actually housework, walking, shopping, an elliptical trainer session, and a run. 26% of activities edited for type were actually walks that were confused as *stair machine* sessions. *"Active"* events accounted for 21%, and were changed to walks, elliptical trainer, housework, shopping, and yoga. 8% of activities edited for type were actually walks that were confused as *elliptical trainer* sessions.

While 207 performed activities were inferred by the fitness device, the device inferred 230 activities. The additional 23 inferred activities were deleted by six participants. Of the 23 deleted activities, 70% were inferred as *bicycling* events, 9% as *walks*, 9% as *stair machine* events, and 4% each as a *run*, an *elliptical trainer* event, and an *"active"* event.

Manually journaled activities. Participants manually journaled activities they performed that the fitness device was not trained to infer (e.g., swimming, lifting weights, etc.) or was trained to but failed to infer (e.g., inference mistake, dead battery, connection problem between the fitness device and phone, broken sensor, device not worn, etc.). Activities were manually journaled throughout the study by all participants. Each participant journaled 11 to 63 activities (mean: 27, median: 20), for a total of 325 manually journaled activities. Table 1 shows how many of which types of activities were journaled by each participant. The frequency with which each participant performed manually journaled activities is shown in Figure 5.

[3] If the fitness device was reasonably certain that the participant did an activity, but was unsure of which activity, it labeled the activity as "active" and classified it as type "other."

These data show that cycling, which accounted for a majority of inferred activities that were deleted and edited for type, needs to be retrained. While the fitness device often correctly detects cycling when cycling is performed, it also confuses other activities as cycling, particularly for a subset of participants; p4 and p9 were most affected by the incorrect cycling inferences. These data also show how participants interacted with activity inference and manual journaling over the three weeks.

Reactions to recording physical activities

In this section, we present participants' reactions to the different types of perceived errors the fitness device made as well as how they felt about manually journaling with UbiFit Garden's interactive application.

Perceived fitness device errors. Participants perceived seven types of errors made by the fitness device. It could: (i) make an error in the start time, (ii) make an error in the duration, (iii) confuse an activity it *was* trained to infer with another it *was* trained to infer, (iv) confuse an activity it *was not* trained to infer with one it *was* trained to infer, (v) fail to detect an activity it *was* trained to infer, (vi) fail to detect an activity it *was not* trained to infer, and (vii) detect an activity when none occurred. Representative quotes are used to explain reactions in participants' own words to six error types; (i) rarely occurred and thus is omitted here.

Error in duration. Seventeen inferred activities were edited for duration. Participants did not seem to mind minor discrepancies except when an underestimate meant the difference between an activity counting toward their goal or not. As p1 explained, "*The only time I ever edited it* [duration] *was when…it would've made the difference between hitting a goal or not,"* for example, when her 30-minute walk was detected at 28 minutes. Several participants commented that they did not closely watch the time for walks, but simply trusted the fitness device.

Confusing an activity it was trained to infer with another. In general, participants thought it was more important for the start time and duration to be correct than the activity, as it was easier to remember *what* they did rather than *when* they did it and *for how long*. P1 explained:

> "[Sometimes] it [the fitness device] didn't quite know what to make of what I was doing. Um, which is walking…once it detected it, I didn't really mind that it didn't detect the right thing you know. I just wanted it to detect the time." {p1}

Confusing an activity it was not trained to infer with one it was trained to infer. When activities were performed that the device was not trained to infer but it inferred something anyway, participants were amused and often appreciated the recognition, even though misattributed. P9 explained:

> "what was really funny was, um, I did, I did some, um a bunch of housework one night. And then boom, boom, boom, I'm getting all these little pink flowers. I'm like ooh, that was very satisfying to get those." {p9}

Failing to detect an activity it was trained to infer. Not surprisingly, participants were frustrated and disappointed when the fitness device failed to detect any activity when they performed an activity that the device was trained to infer. Of the perceived errors, this is among the top two that most affected participants' perceived credibility of the device. As p4 explained, *"Ah…Why am I wearing you* [the fitness device] *if you don't pick it* [the activity] *up?"* This frustration often led to participants questioning if the device was malfunctioning or if they had accidentally broken it. P2 wondered if he was using the system properly:

> "…it's not the end of the world, [but] it's a little disappointing when you do an activity and it [the fitness device] doesn't log it [the activity]…and then I think, 'am I doing something wrong?'" {p2}

Unfortunately, p11's fitness device malfunctioned shortly after his first session, yet we did not realize it for almost a week. He commented on what it felt like when he went for a walk that the device did not detect, "*I wanted the flower, you know?"* He was pretty frustrated by what turned out to be a broken accelerometer, *"*…[I would] *make a point to go for walks and wear the big honkin' thing and then nothing so, you know…Disappointment."*

P1 experienced an interesting side effect of manually journaling a walk that the fitness device failed to detect:

> "The [missed] walks I did add manually. But it—cheating is not the right word, but I couldn't think of the right word. But it kind of felt like cheating when I put those manually in even though there was nothing cheating about it. I did do it [the walks] and it was this long, but it just like, no, it [the fitness device] should have figured that out and told me." {p1}

Failing to detect an activity it was not trained to infer. Most participants did not consider it to be an error when the fitness device did not detect an activity that it was not trained to infer. However, two participants were surprised and a little disappointed that it did not detect *something*. For example, p2 was surprised when the fitness device did not detect something for an exercise DVD that he did. Similarly, p9 thought it was odd that her housework was detected as cycling, yet when she worked for hours digging, shoveling, and lifting things at a farm, nothing was inferred.

Detecting an activity when none occurred. Though it was infrequent, this was the other type of perceived error that had a noticeable impact on the device's credibility, that is, when the device detected an activity when none occurred. One participant occasionally received cycling events while she was in the car, and another got cycling while he was typing email at his coffee table and got a stair machine event while relaxing in a recliner. While participants could understand the device confusing shopping or doing housework for other physical activities, reclining in a chair, typing email, or driving was not so understandable.

Journaling. Participants found UbiFit Garden's journal to be quick and simple. They appreciated the occasional prompts to journal and the minimal text entry it required. Unlike typical paper or electronic journals, the glanceable display itself served as a frequent but subtle reminder to journal (e.g., "Where's my flower from this morning's yoga?") and indirectly provided a reward (a flower) for journaling. As p11 claimed, *"It's easy…the whole journaling thing, it's pretty painless and simple."* P12 explained *"…it* [the journaling] *goes really quick really. It's just like responding to a text message or something."* P1 speculated that she could keep a UbiFit Garden-style journal, as it was fairly quick to use. However, when considering keeping a traditional notebook-style journal, she explained *"if I have a notebook and I have to write it out, for some reason it just does not happen."*

Despite descriptions of the journal being simple, quick, and something they could imagine using long-term, most participants wanted a fitness device to augment the journal by automatically logging activities. As many participants noted, while it is not that difficult to remember that they performed an activity, it can be difficult to remember when or for how long—particularly for walks. Further, the study was conducted over a relatively short time period when considering journaling compliance—a longer-term study is needed (and planned) to investigate reactions over time.

General reactions to the UbiFit Garden system

Reactions to UbiFit Garden were very positive. Most participants mentioned that the garden was motivating, often surprisingly so. In the participants' own words:

> "The silly flowers work, you know?...It's right there on your wallpaper so every time you pick up your phone you are seeing it and you're like, 'Oh, look at this. I have all those flowers. I want more flowers.' It's remarkable, for me it was remarkably like, 'Oh well, if I walk there it's just 10 minutes. I might get another flower.' So, sure, I'll just walk." {p5}

> "I think it's a great idea really…so you can physically see how much you're really moving around with the flowers growing and everything…You kind of want to see more flowers grow or whatever opposed to working out or walking around and not seeing any results…I mean it's going to take several months, but I feel like on the phone you can actually see that you're achieving something…You're not going to lose like five pounds or anything, but you'll have a couple flowers on your phone to kind of encourage you along the way." {p12}

For some, UbiFit Garden helped them focus on *planning* or simply finding time for physical activity:

> "…I want to accomplish these things [activities in her weekly goal] now, whereas before I did but I was a lot more likely to go, 'Eh, let's make up for it tomorrow' or something like that." {p4}

> "[UbiFit Garden is] motivating me to think about exercise like a plan, think more as how to plan it as part of my day, rather than haphazardly, you know, cram it in there when I can." {p8}

> "…what I liked the most is that it helped me to take more advantage of the times where I could do something that was uh, a physical activity…[UbiFit Garden] gave me a little more motivation to take more advantage of opportunities where I could get on a walk…I would think…right now my little girl is on the floor. I could be stretching. I could be doing some things.' So it kind of helped me just you know take more…opportunities." {p2}

However, having to remember to wear the fitness device or simply remembering to bring the phone along on workouts was less than ideal for some, and not surprisingly, participants were concerned about the fitness device's size and weight. Fortunately, participants understood that the fitness device was an early stage prototype and that it would eventually be smaller or incorporated into the phone, and UbiFit Garden's manual journal accommodated the ability to leave all equipment behind yet still include the activity in the system—a feature that participants found important. A longer-term study, which is planned, is needed to explore if the motivating effects of the system persist over time.

DISCUSSION

Given that the three-week field trial is the first time that the MSP was deployed with "real users" in daily life for the purposes of inferring physical activities in real time, the system performed relatively well. However, some activities must clearly be retrained, in particular cycling, as it was responsible for the majority of edited and deleted activities.

Our results suggest that UbiFit Garden supports the type of variety recommended by the ACSM, as shown by the types of activities performed. Our results also point to challenges and opportunities for on-body activity sensing. First, given the range of activities performed in only a three-week trial, there is a challenge in inferring an even wider range of activities. Second, the form factor of on-body sensors, even if incorporated into a mobile phone, is an issue in some common situations. For activities such as basketball, Ultimate Frisbee, and dancing, it may not be practical for even a very small fitness device to be worn during the activity as users are likely to bump into other players or even the ground and do not want to hurt themselves or the device. Similarly, on-body sensors may not be practical for inferring activities such as swimming and kayaking unless their casing is waterproof. Several resistance and flexibility training activities pose challenges for the placement of on-body sensors as well, as the device should not interfere with the proper form needed to perform the activities correctly. Providing a manual journal, as is done by UbiFit Garden and the Nokia 5500 Sport, is essential to enabling individuals to keep track of such healthy physical activities.

We want to highlight two key implications of the field trial's results on systems that use on-body sensing and activity inference to encourage physical activity: (1) traditional error metrics used to describe the effectiveness of activity inference systems do not capture important subtleties of how users perceive inference errors, and (2) the usefulness and credibility of such systems is improved by allowing users to manipulate and add to inferred data.

Traditional error metrics versus user-perceived errors

Traditionally, the effectiveness of inference systems is described in terms of four error metrics: *true positive* (TP), *false positive* (FP) or *type I error*, *false negative* (FN) or *type II error*, and *true negative* (TN). Table 2 shows the error metric breakdown for each of the following six examples from the field trial: (i) the system detected that p4 went for a bike ride when she actually drove her car, (ii) the system detected that p9 went for a bike ride when she actually did housework, (iii) the system detected that p1 used the stair machine when she actually went for a walk, (iv) the system did not detect anything when p2 did a boot camp exercise DVD, (v) the system did not detect anything when p12 went for a walk, and (vi) the system detected that p7 went for a bike ride when he went for a bike ride.

As mentioned earlier, participants perceived important subtleties to the errors—subtleties that are not apparent with the traditional metrics. In the case of the false positives, participants reacted quite differently depending on which of the perceived errors it was—a "bike ride" inference for a ride in the car was intolerable (i), whereas a "stair machine" inference for a walk was tolerable (iii), and a "bike ride" inference for housework was appreciated (ii). Note how the intolerable (i) and appreciated (ii) inferences are assessed identically according to the traditional metrics. Similarly, when considering the false negatives, recall that the walk that was detected as the stair machine (iii) was tolerable, but the walk where nothing was detected (v) was quite disappointing. From the traditional metrics, it is not clear that (i) and (v) were the biggest problems for participants, most negatively impacting the credibility of the system. These results suggest that traditional error metrics are not the most helpful way to describe the effectiveness of such activity inference systems, but rather a new terminology is needed that considers the subtleties of the user's perspective and how they react to the different types of errors. Our results, which describe the seven types of errors perceived by participants in this study, are a step in that direction.

User manipulation of inferred data

Our results emphasize the importance of allowing users to add to, edit, and delete inferred data in systems that use on-body sensing and inference to encourage physical activity. While this may seem obvious, it is not permitted by the majority of commercial products and research projects. Instead, some focus on preventing "cheating" by not allowing users to add to or manipulate inferred data [10].

By allowing users to add to, edit, and delete inferred data, the user can have an accurate record of the physical activities performed, despite flaws with the system's activity inference (e.g., mistaking one activity for another, missing an activity, or detecting an activity that did not occur) or usage model (e.g., if the user did not wear the device or forgot to charge the device). Such data could also be used in a more sophisticated system to adapt or customize activity models for individual users, thus improving the system as users interact with it over time.

Over the three weeks of the field trial, all participants journaled multiple activities, and 10 participants edited and/or deleted data about inferred activities. Most participants thought that all three components were essential to such a system. All found the glanceable display and interactive application to be critical, and most found the fitness device to be important as well. When p8 speculated on using a system that did not permit her to add to, edit, and delete inferred data, she claimed,

Ex.	Inferred Activity	Actual Activity	Traditional Error Metrics True/False Positive/Negative	
i	bike ride	car ride	TP: none	FP: B
			FN: none	TN: W, R, E, S
ii	bike ride	house-work	TP: none	FP: B
			FN: none	TN: W, R, E, S
iii	stair machine	walk	TP: none	FP: S
			FN: W	TN: R, E, B
iv	NONE	boot camp DVD	TP: none	FP: none
			FN: none	TN: W, R, E, S, B
v	NONE	walk	TP: none	FP: none
			FN: W	TN: R, E, S, B
vi	bike ride	bike ride	TP: B	FP: none
			FN: none	TN: W, R, E, S

Table 2. Traditional error metrics describe the inference's effectiveness for six examples from the field trial. W=walk, R=run, E=elliptical trainer, S=stair machine, B=bike ride.

> "[that would be] supremely annoying, because it would've been wrong. You know if I couldn't add the things that it missed when it wasn't on I would've felt jipped, like, 'Ah! I *did* meet that goal, damn it! There should have been two more flowers there.'" {p8}

However, she did not want a system without automatic recording. If UbiFit Garden had been manual entry only,

> "[it would have been] somewhat annoying…that I would've had to track everything, the duration…as well as the fact that I did it…When I think I just want it [the system] to pick it up by osmosis so I really don't want to have to think about documenting. I just want it to be documented." {p8}

Our results suggest that allowing users to add to, edit, and delete inferred data improves the credibility of an imperfect system (we are not sure there is such a thing as a "perfect system"). This is particularly important as most participants prefer a system with activity inference to one that relies on manual entry only. We caution that we did not explore manipulating data in a system that shares data with others (e.g., personal trainer or friend) or promotes competition.

FUTURE WORK & CONCLUSION

In this paper, we described UbiFit Garden, a system we developed that uses on-body sensing, activity inference, and a novel personal, mobile display to encourage individuals to be physically active. We reported on a three-week field trial of UbiFit Garden (n=12), focusing on findings related to participants' experiences with the sensing and activity inference capabilities as well as general reactions to the system. The system was well received, in particular the glanceable display. Some participants were surprised at how motivating the garden was. Most found all three components of UbiFit Garden–the glanceable display, interactive application, and fitness device–to be essential.

Based on results from our field trial, we discussed two key implications for systems that use on-body sensing and activity inference to encourage physical activity: (1) traditional error metrics used to describe the effectiveness of activity inference systems do not capture important subtleties of how users perceive inference errors, and (2) the usefulness and credibility of such systems is improved by allowing users to manipulate and add to inferred data.

We are revising our system based on results from the field trial, including retraining activities such as cycling and adding features that we believe will help with discretionary use over longer periods of time. We are improving the accuracy of the activity inference, focusing on minimizing the two perceived errors that most affected the system's credibility: failing to detect an activity that it is trained to infer and detecting an activity when none occurred.

We are also preparing to run a multi-month study of UbiFit Garden to continue to investigate issues for systems that use on-body sensing and activity inference to encourage individuals to be physically active. In that study, we plan to run three conditions, one with the entire system, one without the glanceable display, and one without the fitness device, to help us investigate the impact of the individual components on users' experiences, as well as explore how attitudes about the system and behaviors change over time.

ACKNOWLEDGMENTS

Thanks to Gaetano Borriello, Tanzeem Choudhury, Cherie Collins, Kieran Del Pasqua, Dirk Haehnel, Jonathan Lester, Jean Moran, Matthai Philipose, Adam Rea, David Wetherall, Alex Wilkie, and many other family, friends, colleagues, study participants, and paper reviewers who have contributed to this work. The 3-week field trial was covered by University of Washington Human Subjects Division application #06-1385-E/C 02.

REFERENCES

1. Anderson, L.H. et al., "Health Care Changes Associated with Physical Inactivity, Overweight, and Obesity," *Preventing Chronic Disease, 2(4)*, (Oct 2005).
2. Bao, L. & Intille, S.S., "Activity Recognition from User-Annotated Acceleration Data," In *Proceedings of Pervasive '04*, (2004), 1-17.
3. Bickmore, T.W., Caruso, L., & Clough-Gorr, K., "Acceptance and Usability of a Relational Agent Interface by Urban Older Adults," In *CHI '05 Extended Abstracts*, (Apr 2005), 1212-5.
4. Choudhury, T., et al., "The Mobile Sensing Platform: An Embedded System for Capturing and Recognizing Human Activities," To appear in *IEEE Pervasive Computing Special Issue on Activity-Based Computing* (Apr-Jun 2008).
5. Consolvo, S., Everitt, K., Smith, I., Landay, J.A. "Design Requirements for Technologies that Encourage Physical Activity," *Proceedings of CHI '06*, (2006), 457-66.
6. DeJong, G., Sheppard, L., Lieber, M., & Chenoweth, D., "The Economic Cost of Physical Inactivity in Michigan," *Michigan Fitness Foundation*, (2003).
7. Fogg, B.J. *Persuasive Technology: Using Computers to Change What We Think and Do*, San Francisco, CA, USA: Morgan Kaufmann Publishers, (2003).
8. Froehlich, J., et al., "MyExperience: A System for In Situ Tracing and Capturing of User Feedback on Mobile Phones," *Proc of MobiSys '07*, (Jun 2007), pp.57-70.
9. Lester, J. Choudhury, T., & Borriello, G., "A Practical Approach to Recognizing Physical Activities," *Proceedings of Pervasive '06*, (2006), 1-16.
10. Lin, J.J., et al. "Fish'n'Steps: Encouraging Physical Activity with an Interactive Computer Game," *Proceedings of UbiComp '06*, (Sep 2006), 261-78.
11. Logan, B., et al., "A Long-Term Evaluation of Sensing Modalities for Activity Recognition," *Proceedings of UbiComp '07*, Innsbruck, Austria (Sep 2007).
12. Maitland, J., et al., "Increasing the Awareness of Daily Activity Levels with Pervasive Computing," *Proceedings of Pervasive Health '06*, (Nov/Dec 2006).
13. Mokka, S., Väätänen, A., Heinilä, J., & Välkkynen, P., "Fitness Computer Game with a Bodily User Interface," *Proc of ICEC '03,* Pittsburgh, PA, (May 2003), pp. 1-3.
14. Mueller, F., Agamanolis, S., & Picard, R, "Exertion Interfaces: Sports Over a Distance for Social Bonding and Fun," *Proceedings of CHI '03*, (Apr 2003), pp.561-8.
15. Mueller, F., O'Brien, S., & Thorogood, A., "Jogging Over a Distance," In *CHI '07 Extended Abstracts* (Apr 2007), 1989-94.
16. National Center for Health Stats, "Health, United States, 2005 with Chartbook on Trends in the Health of Americans," Hyattsville, MD, USA, (2005).
17. Oliver, N. & Flores-Mangas, F., "MPTrain: A Mobile, Music and Physiology-Based Personal Trainer," In *Proceedings of MobileHCI '06*, (Sep 2006).
18. Pentney, W., et al., "Sensor-Based Understanding of Daily Life via Large-Scale Use of Common Sense," *Proceedings of AAAI '06*, Boston, MA, USA (Jul 2006).
19. Prochaska, J.O., DiClemente, C.C., Norcross, J.C. "In Search of How People Change: Applications to Addictive Behaviors," *American Psychologist, 47(9)*, (Sep 1992), 1102-14.
20. Sohn, T., et al., "Mobility Detection Using Everyday GSM Traces," In *Proc of UbiComp '06*, (2006), 212-24.
21. Toscos, T., Faber, A., An, S., & Gandhi, M.P. "Chick Clique: Persuasive Technology to Motivate Teenage Girls to Exercise." In *CHI '06 Ext Abstracts* (2006), 1873-8.
22. Whaley, M.H., Brubaker, P.H., & Otto, R.M. (Eds). "General Principles of Exercise Prescription," *ACSM's Guidelines for Exercise Testing and Prescription*, 7th Ed, Baltimore, MD: Lippincott Williams & Wilkins, (2006).