Tools for Studying Behavior and Technology in Natural Settings

Stephen S. Intille, Emmanuel Munguia Tapia, John Rondoni,
Jennifer Beaudin, Chuck Kukla, Sitij Agarwal, Ling Bao, and Kent Larson

Massachusetts Institute of Technology
1 Cambridge Center, 4FL
Cambridge, MA 02142 USA
`intille@mit.edu`

Abstract. Three tools for acquiring data about people, their behavior, and their use of technology in natural settings are described: (1) a context-aware experience sampling tool, (2) a ubiquitous sensing system that detects environmental changes, and (3) an image-based experience sampling system. We discuss how these tools provide researchers with a flexible toolkit for collecting data on activity in homes and workplaces, particularly when used in combination. We outline several ongoing studies to illustrate the versatility of these tools. Two of the tools are currently available to other researchers to use.

1 Introduction

Products developed and tested in laboratories often fail when introduced into natural settings such as homes and workplaces. Human behavior in these natural settings is strongly tailored to the settings themselves and to the behavior of the people nearby. Differences between expectations of how people will behave and how they actually do behave in the complexity of real settings contribute to product failures. Developers often make erroneous assumptions about the need for, use of, and reaction to new technologies.

Ubiquitous computing technologies are more intimately tied to their environments and the people within those environments than desktop applications. Therefore, these ubicomp applications must respect, detect, and respond to the interaction between the person and the natural setting. The common desktop model of software evaluation with usability labs will fail to capture the influence of setting on behavior. Simulation of realistic natural behavior in the laboratory is difficult, because to do so requires reconstructing the environments themselves.

Ubicomp developers, therefore, face a challenge: how can models of behavior in natural environments be obtained? Once applications are created, how can they be evaluated *in situ* and over long stretches of time? Effective development and deployment of persistent, context-aware interfaces will require tools and strategies for using those tools that can help researchers establish good answers to such questions. Longitudinal study of interface use in natural settings is particularly challenging. Today, many ubicomp interfaces are proposed that would

run for months or years at a time when deployed. Yet these systems are typically only tested for hours at a time in labs. Both the *impact of the environment* and the *impact of time* on the behavior of the users of these applications must be considered.

In this paper we describe three tools we have created, each in different stages of development. The first tool, context-aware experience sampling, extends electronic experience sampling to include a set of sensors that both collect data and proactively trigger data collection. The second tool is a data collection system of small, simple state-change sensors that can be quickly installed throughout nearly any environment to collect information about patterns of activity. These two tools, both of which are available to the research community, are most powerful when used in combination. We describe how we are deploying them in multiple studies to gather data on activity in natural settings. Finally, we describe a laboratory prototype of a third tool that further extends the observational capabilities of researchers and developers: image-based experience sampling.

2   Studying People in Natural Settings

Developing meaningful ubiquitous computing applications should begin with a thorough understanding of how people behave in their environments, how they perform the tasks to be undertaken, and how this behavior may be influenced by the introduction of new technologies. In short, first we must understand the behavior of people and *then* develop the technology. For instance, despite little evidence that non-gadget-oriented consumers want or need automated home lighting and HVAC control or remote control of appliances, many popular visions of ubiquitous computing technology advocate the use of complex sensing to achieve these goals (e.g. [22,8]). Often, proposed applications would strip control from end users and "simplify" life by acting as autonomous, controlling agents. Many demonstration applications are built upon unrealistic assumptions about user tolerance for applications that make erroneous control decisions.

2.1   Motivation: Empowerment vs. Control

An alternative approach is to develop ubiquitous computing environments that use technology not to automatically control the environment but instead to help users learn how to control the environments themselves. In this vision, many ubiquitous computing applications will not make a decision for the users but instead present information to users. The task of interpreting a suggestion or information in context rests with the user. Studies in preventive medicine and energy management have shown that simple, passive, but context-dependent and relevant indicators can have a dramatic influence on behavior (see [15]).

This approach has several advantages over a controlling approach:

– Information can be presented so that the occupant can react to without interrupting ongoing activity in potentially irritating ways; this is especially true

if information can be digitally "augmented" onto the physical environment itself using projection.

– Leaving the user in control of making decisions without confusing the occupant; the occupant will naturally consider contexts that the ubicomp system has not and adjust his or her actions accordingly.
– Algorithms that make suggestions can degrade gracefully; algorithms that make decisions typically do not.
– Lack of control over aspects of life has been shown to diminish health [24]; this strategy empowers the occupant.

A shift in focus from creating automatic ("smart") environments to environments that help the occupant learn how to take self control impacts not only the type of technology one might design but also one's outlook on how to conduct research to evaluate the work. A "Jetsonian" model of a future environment, for example, where the computer simplifies life with automated control, can be evaluated by designing demonstration environments in a traditional laboratory with little user involvement. However, environments that are designed to help occupants learn cannot be evaluated independently of the people using them. We need to study the *people using the technology* in realistic, non-laboratory settings for long periods of time and then measure whether our interventions are leading to learning and behavior change. This is the primary motivation for the development of the three tools described below for naturalistic observation.

2.2   Standard Approaches to Naturalistic Observation

Developers of ubiquitous and mobile computing applications for the home and workplace currently lack a powerful and economical assessment tool set. There are five classes of methods used to elicit user needs:

Interviews. Interviews are performed individually or using focus groups and are particularly effective for critique of an existing idea or (with an effective interviewer) gathering general information about user tasks. Often, however, users know more than they say in a single or even several interviews [25] and will tend to have difficulty understanding and recalling how context impacts their behavior (i.e. exhibiting selective recall and selective reporting biases [26]). One special form of interviewing consists of participatory design games that are conducted in the setting where the technology will ultimately be used.

Direct observation. Direct observation with trained observers does not suffer from selective recall and is considered the "gold standard" for assessment in medical and psychological research studying behavior in natural settings. Although direct observation can provide helpful qualitative and quantitative measures, it is costly, time-consuming, and disruptive and therefore not practical for many design tasks, particularly those where researchers need to invade private settings such as the home. Recently, direct observation with photographic and video analysis has been used (e.g. [14]).

Self-report: recall surveys. Despite the widespread use of self-report surveys for assessment of behavior in naturalistic settings, these surveys are known to suffer from recall and selective reporting biases - users can often not remember what they did or do not report what they actually did. Further, they often report what they did incorrectly [26].

Self-report: time diaries. To minimize selective recall and selective reporting bias, time diaries can be used. Users write down what they do during the day as they do it or at regular intervals [23]. Diaries can provide less biased data than recall surveys but are burdensome for the user, can impact the activity itself, and often result in records with missing information. Providing users with devices such as dictaphones, cameras, and video cameras has been used to simplify self-reporting.

ESM/EMA. The experience sampling method (ESM), also known as ecological momentary assessment (EMA), uses a timing device to trigger self-report diary entries [11,26]. In electronic ESM, survey questions can be answered on a portable computing device that "samples" (e.g. via a beep) for information. Only recently has ESM been employed for interface design [16,17,13,10,14]. Sampling can occur using fixed, statistical, or user-initiated schedules. With a sufficient number of subjects and samples, a statistical model of behavior can be generated. The ESM is less susceptible to subject recall errors than other self-report feedback elicitation methods [11,26], but at high sampling rates it can interrupt activities of interest and irritate subjects, resulting in some subject-selection bias [11].

3   Challenges: Improving Observation Tools

We are interested in using and developing sensing tools to improve the assessment techniques. We have created three such tools: environmental sensors (ES), context-aware experience sampling (CAES), and image-based experience sampling (IBES). No tool will suit every need, but ideally combinations of tools can be paired to achieve the properties discussed below:

Dense measurement of activity. For many in-home studies, researchers need a dense description of activity recorded from the home environment to analyze. The collection tool, however, should not impact behavior in the environment. Therefore, while sometimes invasive cameras or microphones can be used, in other cases sensors are needed that allow a researcher to study activity without direct, invasive observation. (ES)

Fast install/removal. Most homes and workplaces do not easily accommodate even the simplest of new technologies. Therefore, researchers require tools that can be used to study natural settings that can be retrofitted into these environments easily and at low cost. (ES, CAES)

Low cost. Both low-cost manufacturing and installation and maintenance of sensors must be possible so that they can be deployed in multiple households for long time periods. (ES, CAES)

Robust longitudinal data collection. Many studies require data collection over the course of weeks from home environments. (ES, CAES, IBES).

Longitudinal acceptability. Directly querying a user is a powerful technique, but one that should be done sparingly to avoid annoying a user (e.g. perhaps when an application is first used to perform some customization to the user's routines). We are experimenting with the combination of *context-aware* sampling and other tools to minimize the burden of self-report on subjects so the technique can be used for longer time periods. (ES, CAES, IBES)

Not perceived as invasive. To deploy technologies into homes for any extended period of time, they must not be perceived as invasive. (ES, CAES)

Autonomous operation possible. Although some tools require human intervention, ideally observational tools will be useful not only for observation but also for intervention with observation, where some event is detected automatically, some information is provided, and the user's response is observed. (ES, CAES)

Sensing technology can be used to extend and compliment traditional observational techniques in a cost-effective manner. Our goal is to select a suite of sensors that permits a researcher and, eventually an algorithm, to quickly analyze the data and construct a record of activity. Exactly what the researcher is looking for depends upon the activity. Therefore, the tools should be flexible. We advocate using a combination of room-mounted, body-mounted, and object-mounted sensors in combination with a self-report mobile computer tool. In the remainder of this paper we describe the three three tools we have created and some observations resulting from their deployment.

4   Tool 1: Context-Aware Experience Sampling

The development of the first tool began when members of our research team identified a need for a robust self-report data collection tool. As we worked on the development of computational perception algorithms for automatic identification of activity and contextual information from sensor data acquired from mobile computing devices, we realized that we needed accurate annotation of the sensor readings with the target activities to both train and test pattern recognition algorithms. In the laboratory obtaining this annotated data is a straightforward process: use direct observation and label activity in real time or through observation of video sequences. In our case, however, we were interested in identifying activities as people went about their lives outside of the lab. Direct observation was too costly, time-consuming, and invasive. Therefore, we employed self-report time diaries. Subjects were asked to keep diaries of their physical activity (e.g. walking to class, climbing stairs) and when they did it.

We encountered selective reporting and forgetfulness typical of self-report (see Sect. 2.2). Therefore, we decided to use electronic ESM to ease the subject's burden, improve the accuracy of time-stamps acquired, and reduce the data entry and coding burden of the experimenter. We found that a few commercial experience sampling programs are available (see listing from [6]) as well

as one open-source program for old versions of the Palm and Windows CE operating systems [5]. However, in addition to lacking some new ESM functionality we desired and only operating on outdated hardware, the available software did not permit both the acquisition of user self-report data and the simultaneous acquisition of other data streams such as position (from GPS) and heart rate (from a wireless monitor). We therefore have developed new experience sampling software. The basic tool offers options not currently available in any other open-source or commercial sampling package (e.g. leaving an audio or photo annotation). In addition, the software can collect readings from sensor devices attached to the PDA.

4.1   Using Context to Trigger Self-report

In addition to extended data collection capabilities, the tool also provides one fundamentally new type of functionality: *context-aware* experience sampling. This feature permits researchers to acquire feedback from users only in particular situations that are detected by sensors connected to a mobile computing device. The context (location, time, event, biosensor data) can trigger the sampling.

Typically a researcher using experience sampling has three options: (1) sampling on a fixed interval schedule, such as every 30 minutes, (2) sampling on a random interval schedule, such as on average once every 30 minutes or sometime randomly within every 2 hour window, and (3) sampling in response to user initiative, where the user is told to make a data entry whenever performing a particular activity [26].

Our software has been developed in a modular fashion that allows new context-sensing sensors and software to be plugged in. These sensors permit researchers to use context-sensitive sampling where specific questions are asked only when a user does a specific thing (e.g. is near the store, which is determined by a GPS and map module, or has a change in heart rate, which is determined by a wireless heart rate monitor). Context-awareness modules permit a researcher to acquire more information about the behavior or situation of interest by sampling only during or just after the activities of interest. This minimizes the interruption annoyance of the ESM technique without compromising the quality of data acquired about the target phenomena.

4.2   Implementation

We have used a participatory design process to create the interaction model for the ESM tool so that it suits both researcher and subject needs. Our goal has been to create software that permits a device to be handed to a subject at the start of a week with only a few minutes of instruction and then to be returned a week later with question and sensor data. The interface has been designed so that it is self-instructing and easily understood by non-computer users.

The first release of the software includes capabilities for standard multiple choice question experience sampling using a time-sampled protocol. The software, written in C++, runs on the PocketPC platform and has been developed

(a)   (b)   (c)
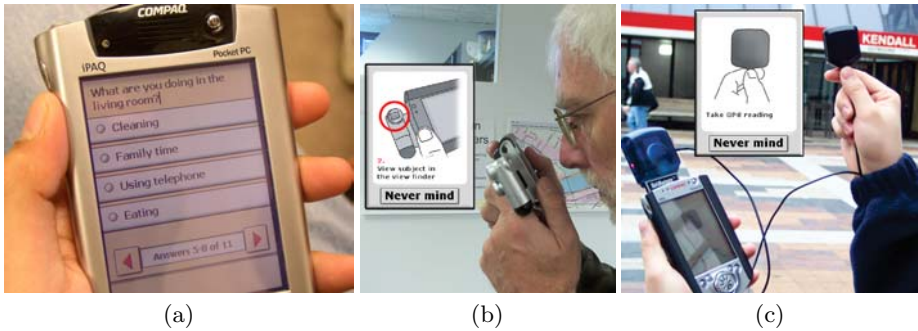
Fig. 1. (a) The context-aware experience sampling software runs on standard PDAs and offers a variety of options for acquiring self-report data from users or subjects in experiments. This image shows one screen from sampling protocol being used by a study collecting training data on activities in the home setting that is being used to develop algorithms to detect everyday activities from sensor data automatically. (b) The software permits the use of data collection sensors such as cameras, microphones, and wireless heart rate monitor. This image shows a user taking a picture with the camera plug-in and one of the built in tutorial screens. (c) The software can also either prompt for sensor data such as taking a GPS reading or use continuously-acquired sensor data to trigger a set of self-report questions to be asked. Here a user is acquiring a GPS sample, which is done by holding up a small antenna plugged into the GPS device.

primarily for iPAQ devices due to the large number of hardware connection options these devices provide and the bright, high-contrast screen that makes the questions and instructions easier to read.

Researchers load a new protocol by modifying a comma-delimited text file. Our software also includes protocol development flexibility not found in other open-source ESM software [5]. These include options for (1) chaining complex sequences of questions based upon particular question responses, (2) aggregation of questions to minimize user disruption in some situations, (3) study suppression during some events, (4) multiple choice and multiple response questions, (5) manual specification of precise query times for particular questions or question sequences, (6) flexible question recurrence patterns (by weeks, days, hours, minutes), and (7) bounded randomization (min/max time to next query). Further, researchers can allow users to leave answers via audio recording or, if the PDA has a camera plug-in, answer a question by taking a picture. The tool can therefore be used to combine the powerful techniques of ESM and photographic analysis. The device includes just-in-time tutorials to aid subjects with some of the advanced functionality. Figure 1 shows screen shots of users answering questions, including questions that ask for data samples (picture, GPS) taken using plug-in hardware.

The context-aware functionality of the first release includes the ability to sample based not only upon the standard time-based protocol but also upon a subject's location, as obtained by a GPS plug-in. Therefore, researchers can

design studies that sample only when near a location of known interest. Similarly, a wireless Polar heart rate monitor can be used to trigger samples based upon significant changes in heart rate.

Additional context-aware extensions are currently being added to the software, including the ability to sample based upon *particular activities* that have been detected using accelerometers, the GPS, and the heart rate monitors (e.g. "walking"). We are building an open source development community and invite researchers interested in using experience sampling or developing extensions to the context-aware experience sampling tool to visit the project pages found on SourceForge [6] and join the effort.

The primary benefit of context-aware experience sampling versus traditional experience sampling is that specific activities can be targeted. This means that intensity of sampling can be increased around those moments or activities being studied and reduced at other times. Further, the type of questions answered can be changed based upon the context detected by the software. Finally, context-aware experience sampling can be setup so that samples are delayed until after an activity of interest may have just finished rather than sampling at a random time during the activity, which may disrupt the activity itself.

5   Tool 2: Ubiquitous Environment State-Change Sensor System

We have developed a second observational tool that *passively* collects data via measurement of objects in the environment and compliments the self-report data collected by the context-aware experience sampling device.

We have created the software and hardware for a system of simple, robust sensors that can be ubiquitously installed into complex environments such as real homes and workplaces to collect data about activity. Analysis of the data, either by hand or by machine, may enable better understanding of activity in naturalistic settings and create new opportunities for development of context-aware applications.

In prior work where sensors have been ubiquitously installed into home or workplace environments, typically only a small number of sensors have been used or the studies have been conducted in relatively controlled research settings such as homes of the researchers themselves or close affiliates (e.g. [22,3,1]). The sensor installation itself is often difficult and time consuming to accomplish and troublesome to maintain.

Our system uses "tape on" sensors that can be quickly installed in a complex natural environment to measure just a few or several hundred object states, depending upon how many are used. A small team of researchers can install the system in a small, one-bedroom apartment of typical clutter and complexity in only about 3 hours. The devices have been used to continuously and passively collected data for two-week blocks in multiple subject homes. These subjects have had no affiliation with our research project.

Fig. 2. The second tool is a system of environment state change sensors that can be installed ubiquitously throughout an environment such as a home. Each device consists of a data collection board and a small sensor, and the components are literally taped to objects and surfaces for the duration of the data collection period. These images show the data collection board outside of the protective case and 11 of the 85 sensors installed throughout one subject's home.

5.1   Implementation

Our design goals were to permit several hundred low-cost sensors to be installed in an environment for at least two weeks, left unattended, and then recovered with synchronized data. Figure 2 shows a sensor device, which consists of the sensor itself connected by a thin wire to a 27mm x 38mm x 12mm data collection board. The devices are robust and easy to work with. Each fits snugly in a plastic case of dimensions 37mm x 44mm x 14mm. They use either reed switches, which are activated when brought into contact with a small magnet, or piezoelectric switches, which detect movement of a small plastic strip. Use of temperature, vibration, and load sensors is also possible. The plastic cases are literally taped onto surfaces using a non-damaging adhesive selected based upon the material of the application surface. The sensor components (e.g. reed and magnet) and wire are then taped to the surface so that contact is measured. Figure 2 shows 11 of 85 sensors that were installed in the home of one subject. They can be attached to many devices in the home, including light switches, containers, and furniture. A trained researcher can typically install and test a sensor in less than 3 minutes. The sensors in this subject's home operated for 16 days.

To save memory and cut cost, the boards save data with time stamps that have 1s resolution. To achieve well-synchronized measurements a temperature-compensated crystal oscillator is used to drive the real time clock of the data collection board. This achieves a time accuracy of +/- 1 minute per year if operated from 0 to +40C. To further improve the synchronization, prior to in-

stallation all of the boards are synchronized in a single session with a reference PC clock. When the data collection boards are recovered, the signals from each board are linearly interpolated to better match the reference clock. In boards installed in our laboratory, we have measured the synchronization after this correction to be +/- 2 seconds. The boards can record up to 3 activations and deactivations per second and can record a total of 2666 events in memory.

The total cost for parts and fabrication (in quantities of 150) for each data collection board as of February, 2003 is $24.66, where an additional $2 is required for each sensor (e.g. magnetic reed). When sensors are installed, each data collection board (which has a unique ID) is marked on a plan-view of the environment so that when the sensor data is collected the location and function of each sensor is known. The sensors are in continuous use at this time. We are sequentially installing and then removing them in different subject homes. The complete hardware and software specifications for the devices are available on request.

Although other portable sensing systems for ubiquitous computing applications have been developed [12,18,19,21], we are unaware of work where 100+ of these devices have been rapidly deployed in multiple homes as "tape on and forget" devices and collected data with non-researcher occupants in normal homes for weeks at a time. Several groups have hard-wired cabinets and mats [1] and a some kitchen appliances [3], but these systems have not been deployed throughout multiple, entire homes because of the difficulty of installation and maintenance. The simplicity and small size of our data collection boards make it possible to cost-effectively deploy large numbers in non-laboratory settings. Although our boards are simpler than distributed network devices such as the Berkeley motes, their parts are significantly less expensive and they can operate for substantially longer time periods.

5.2   Deployment: Combining Tools

Direct study of the sensor data may be useful to some researchers. For instance, the total number of firings or frequency of firings of particular sensors may be interest for specific inquiries (e.g. a sensor in a cabinet or drawer may offer clues about medication adherence or a light switch in the bedroom may offer clues about sleeping patterns). One of us, for example, is studying how the data may be used to help people learn about the use of their own environments (e.g. the kitchen) and to help them design new ones. Figure 3a shows one sequence of activations in a subject's home, where the arrows indicate sensors that fired after each other. This representation permits a researcher to study patterns of movement about the environment.

The sensor data collected, however, is most useful when the sensor readings can be correlated with self-report or observational data about what the person or people in an environment were actually doing at the time when objects in the environment were being manipulated. These activity labels are particularly important if the observation system is to be used to automatically collect data about behavior. With such labels, it then becomes possible to consider the cre-

ation of algorithms that recognize context automatically from the ubiquitous switch sensors using pattern classifiers customized to the individual using supervised learning (e.g. [4]).

We are exploring the use of two strategies to obtain the activity label data. The first is by using the context-aware experience sampling tool. In this approach, the environment sensors have been installed in homes of subjects who live alone. The subjects are given a PDA running the experience sampling software described in Sect. 4. As the state-change sensors record data about movement of objects, the subject uses experience sampling to record information about his or her activities. We have used a high sampling rate, where the subject is beeped once every 15 minutes for up to 16 days. At the beep, the subject receives the following series of questions. First the user is asked "What are you doing at the beep?". The subject selects the activity that best matches the one that he/she was doing at the time of the beep from a menu showing up to 33 activities. Next, we ask "For how long have you been doing this activity?" The subject can select from a list of four choices: less than 2 min., less than 5 min, less than 10 min., and more than 10 min. Then, the user is asked, "Were you doing another activity before the beep?". If the user responds positively, the user is presented with menu of 33 activities once again. For our current studies, we are using an adaptation of the activity categories used in the multi-national time-use study [27] supplemented with some activities from the compendium of physical activities [2].

The self-reported activities can be used to study the environment sensor data, or visa versa. Figure 3b shows all the environment sensor activations for one subject on a particular day at about the time a "cooking breakfast" activity was recorded by the context-aware experience sampling tool. Figure 3c shows the sensor activations for this same activity on a different day for the same subject.

In our work to date, we have found that subjects quickly acclimate to the presence of the environmental sensors. However, they have reported that the experience sampling software (sampling at roughly 15 minute intervals) becomes a significant burden after about 1-2 days. Interestingly, the interruption itself does not appear to cause the greatest annoyance (see Sect. 6. We are currently conducting tests to qualitatively measure ESM compliance and the benefits of context-aware ESM.

The second method to obtain activity labels we are exploring is direct observation. Fifty of the environment sensors have been deployed in a researcher's home with a wireless webcam that captures 30-100 frames per hour (based on a motion trigger) and saves the images to a nearby computer. Software can then correlate the timestamped imagery with the environment sensor readings. Figure 3d shows an image obtained from the webcam system. Such images can be annotated based upon the interests of the observing researchers. Alternatively, the environment sensor triggers can be used to find images that can be used for photographic analysis where a researchers asks a subject, "What were you doing here?". Annotation of image data with other sensor data is useful for a variety of applications and analysis tasks [9,7]. In our system, the ESM, state-sensor,
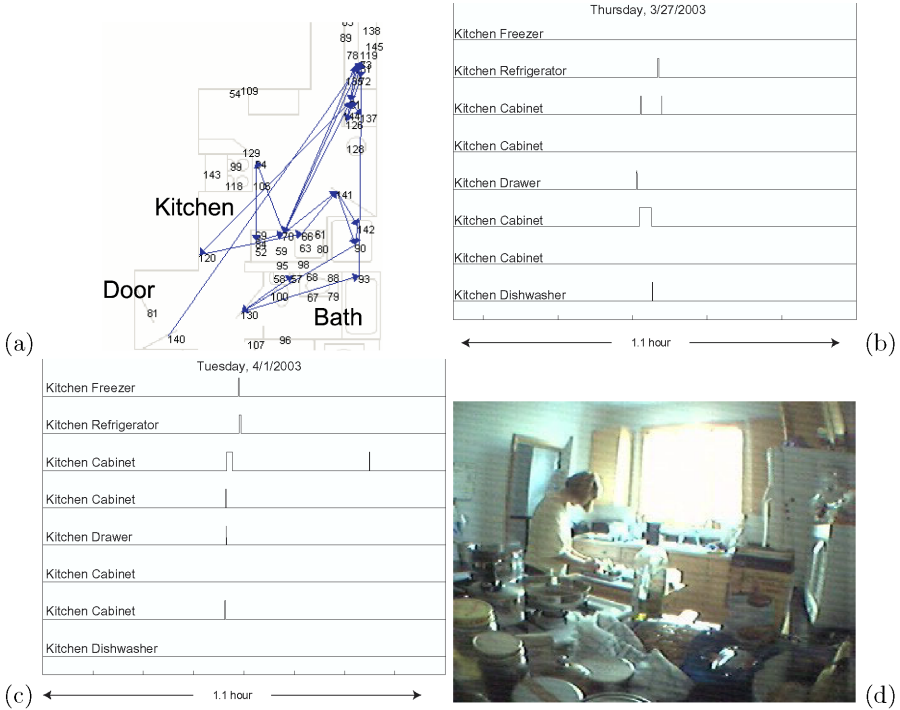
Fig. 3. (a) Data from the environment sensor system can be analyzed for clustering. This image shows the movements of one subject during a cooking event based upon sensor activation times. (b,c) The environment sensors are being used simultaneously with the context-aware experience sampling system to create datasets for pattern recognition research. These images show the environment sensors that fired around the time the subject reported a "cooking breakfast" event on two different days. (d) Time-lapse image capture is being correlated with the environment sensor data in some studies. Here a researcher works in her kitchen, which has been wired with environment sensors and a wireless camera.

and imagery data are correlated in time, which permits researchers to view only those portions of the data related to the event of interest.

5.3   Tool 3: Image-Based Experience Sampling

The third observational tool, which can be used in highly-instrumented environments, combines scene-based sensing with sampling techniques. ESM is less susceptible to subject recall errors than other self-report feedback elicitation methods, but the primary drawback to ESM is the interruption created by the sampling itself. ESM disrupts the user's activity, requiring the user to stop the current activity and answer questions. Therefore, we propose the use of context-sensitive image capture we call image-based experience sampling. Instead of disrupting the current activity with an alert, an audio-visual "image" is captured

of the current activity at each sample time. The image can be a static picture or a small video clip of activity taken from one of potentially many recording devices installed in the environment.

This tool can sample on a fixed schedule, as the system that captured Fig. 3d is doing, or it can sample based upon sensor readings. We have implemented a context-aware prototype of this tool in our laboratory using a multi-camera computer vision person detection system. Samples are taken (i.e. images are captured) only when a person is identified as being in the room. Figure 4a shows one such picture that was captured when someone was sitting in the room. The sampling itself does not interrupt the activity or require any proactive action on the part of the person.

At some later time, a researcher can study the images. However, the full power of the technique is realized when not only the researcher but the user reviews the sampled imagery. The user was not interrupted at the sample time, but the rich contextual information provided by the image or video clip triggers the user's memory of the moment when the sample was taken [9]. Therefore, specific questions of interest can be asked of the user about that situation in time. A mobile interface allows the user to view the images and answer questions whenever it is convenient, such as during "idle time" waiting in line or riding transportation.

5.4   An Example Application

Our prototype shows off how this observational technique can be used to drive a new type of application: helping users make design decisions. Determining requirements for any design project involves identifying and ranking user needs and preferences. Image-based experience sampling can be used to assist someone who is interested in learning about his or her own preferences in a way that is more personalized and less disruptive than interviews, focus groups, and standard experience sampling.

Consider this scenario. Susan feels unhappy with her current kitchen and plans to remodel it sometime in the next six months. She has a limited budget and knows that she must prioritize the changes she would like to make. Such a scenario is common. The user knows that something in his or her life needs to change but is uncertain how to evaluate the relative importance of different options. A common method for helping users evaluate their preferences in such instances is to construct interfaces that prompt the user for information that is used to determine the combination of attributes that provides the most perceived value to the user [20]. Interfaces that use such an approach are typically prompting the user about preferences outside of the context of everyday activity (i.e. kitchen redesign software might ask Susan a series of questions about her kitchen). The best time to ask the user about preferences, however, would be in the midst of the actual activity being scrutinized. Further, most desirable would be if an interface could help the user build up awareness and understanding of the user's preferences over time. Combining image-based experience sampling and conjoint analysis, a technique for measuring preferences, a non-disruptive

(a)   (b)
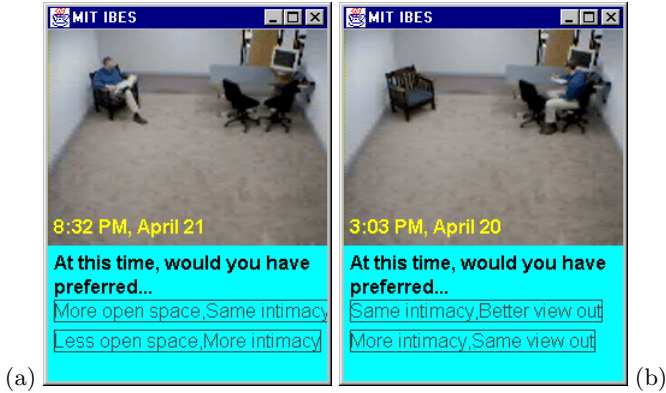
Fig. 4. In a prototype implementation of image-based experience sampling in our laboratory, images are captured when someone is in a room instrumented with cameras. At some later time, the user can view the captured images on a mobile computing device and rank preferences about the architectural space, as shown in (a) and (b). Conjoint analysis is then used to create an overall ranking of preferences, but those preferences are based upon samples taken in context of everyday life rather than using unaided retrospective recall.

but context-sensitive preference elicitation user interface mechanism can be developed.

Conjoint analysis is typically implemented via a written or online survey. A large set of features is preference ranked by asking users to compare smaller sets of features against each other. For example, Susan might be asked, "Choose a number from 1 (always) to 10 (never) indicating how often each of the following situations occur: kitchen counter too crowded, I feel tired, shoulders get sore, feel somewhat drab." If Susan answers a sufficient number of such questions, statistical analysis can be used to weight the relative importance of features and conditions.

Figure 4 illustrates the interface's operation. Using the cameras located in the environment, samples are randomly acquired whenever the user is in the space. Each day a few new images are acquired. The goal is to acquire samples of everyday activity so that the user can later refer back to remember a particular situation and comment on how the user was feeling about the environment.

At the user's convenience, subsets of images can be reviewed on a mobile computing device. Our model is that the user interacts with the interface in bursts of 30-90 seconds during idle moments throughout the day. The user can quickly scan an image or two and provide some preference information. This creates awareness and learning during short bursts of activity on a regular basis. Upon viewing a picture, the user will see the image or video clip, the time the clip was taken, and a question used to acquire information about how the person was feeling about the adequacy of the physical environment during the pictured activity.

Image-based experience sampling combines the power of three techniques: use of media for contextual recall, use of experience sampling, and use of con-

joint analysis for preference rankings. In this example we implemented context-awareness (i.e. detecting if a person is around) using optical sensing. However, if imagery is captured continuously, the environment sensors could be used to identify specific events about which to later ask the user. For instance, the user could be shown images taken just before making a change to the lighting and asked to evaluate lighting preferences in the home.

6   Applications and Next Steps

As we deploy these tools and interview our subjects, we are learning how to improve them. For example, after several environmental sensor and ESM simultaneous deployments, we decided to add new functionality to the ESM software that permits a subject to fluidly switch between prompted data collection and self report. We found users develop mental models of how the experience sampling works that influences how they use it and that can contribute to feelings of frustration. For instance, subjects have expressed that they do not mind answering questions about new actions, but they get frustrated quickly when the ESM software is asking them questions about things they feel they have already "taught it." We are working to improve the context-aware experience sampling by giving the user more control (or the perception of more control) over the device's question-asking behavior. We are also working to add new functionality to each of the tools so they can be used more effectively in combination. For instance, the ESM software can now sample by taking pictures from the PDA directly. If worn in the front shirt pocket, the researcher can then obtain a continuous stream of pictures of what a person was facing. Because the system is not taking a picture of the user but instead of what the user sees, the system may be perceived as less invasive. The images, however, can improve a researcher's ability to interpret self-reported data or the data from the environmental switches.

These three tools are relatively new and yet we are rapidly expanding the number of uses we find for them – particularly the combination of the three techniques. Further, although image-based experience sampling is in use only as a laboratory prototype, the context-aware experience sampling tool and the environment sensing system have been deployed in multiple, non-laboratory settings for data collection. The tools are being used for the following ongoing work.

Mobile activity recognition algorithm development. The context-aware experience sampling tool is being used to collect datasets on physical activity for the development of algorithms that can detect various types of everyday activities automatically (e.g. walking, going to work, climbing stairs) using body-mounted accelerometers and a GPS plug-in. The results of some of our work on algorithms that automatically identify context that uses the tool will be rolled back into the tool itself, providing future researchers with more context-specific triggers.

Environment activity recognition algorithm development. The context-aware experience sampling tool in combination with the environmental sensors are being used to collect multiple datasets from *real homes* that can

be used to study the development of new pattern recognition algorithms to detect everyday routines.

Interruptibility. We, like others [14], have found experience sampling to provide a valuable tool for studying detection of interruptibility, a key emerging problem for ubicomp acceptability. However, we are also using the context-aware capabilities of the context-aware experience sampling device to study the relationship between bio-monitored data and interruptible moments.

Kitchen design. We are using the environmental sensors with video image capture to study how these tools could be used to help people learn about their own behavior, particularly with respect to how that understanding might impact how people make future design decisions.

Other studies are also being considered, such as using the context-aware experience sampling device in a small workplace. Using a Bluetooth sensor the device could be programmed to detect moments of interaction between two people (within Bluetooth range). This particular context-aware trigger would permit the study of how casual interaction impacts work and the perception of productivity.

We have designed these tools to meet the needs of a set of ongoing and planned experiments in our laboratory, but we believe they are sufficiently robust to be of value to other researchers. Although the tools have only recently been deployed, we have been surprised at the number of new studies of people in naturalistic environments that they enable. Therefore, we have created an open source project to further develop the context-aware experience sampling tool [6], and the specifications for the hardware and software design of the environment sensors are available on request.

Finally, an apartment living laboratory is currently under construction in Cambridge, Massachusetts and will be completed by October, 2003. This lab will be fully instrumented with same sensor infrastructure discussed here. In particular, the environmental sensing system will be physically wired so that PDA-based sampling or image-based sampling protocols can be developed where particular actions (e.g. opening a cabinet, turning a faucet) can trigger sampling. The facility will be a shared scientific resource available to researchers. Studies will be possible using a combination of tools in stages: (1) first using the portable tools to study subjects in their own homes, then (2) studying the same subjects as they live temporarily in the living laboratory, and then (3) studying the same subjects as the move back into their own homes. We encourage researchers in the ubicomp community interested in using or helping to further develop these observational tools to contact us.

Acknowledgments

References

1. G.D. Abowd. Personal communication, November 2002. Director of the Aware-Home Initiative, Georgia Insitute of Technology.
2. B.E. Ainsworth, W.L. Haskell, M.C. Whitt, M.L. Irwin, A.M. Swartz, S.J. Strath, W.L. O'Brien, Jr. Bassett, D.R., K.H. Schmitz, P.O. Emplaincourt, Jr. Jacobs, D.R., and A.S. Leon. Compendium of physical activities: An update of activity codes and MET intensities. *Medicine & Science in Sports & Exercise*, 32(9 Suppl):S498–504, 2000.
3. M. Alwan. Personal communication, November 2002. Assistant Professor, University of Virginia.
4. T. Barger, M. Alwan, S. Kell, B. Turner, S. Wood, and A. Naidu. Objective remote assessment of activities of daily living: Analysis of meal preparation patterns. Poster presentation, Medical Automation Research Center, University of Virginia Health System, 2002.
5. L.F. Barrett and D.J. Barrett. An introduction to computerized experience sampling in psychology. *Social Science Computer Review*, 19(2):175–185, 2001.
6. CAES. Context-Aware Experience Sampling Website, Access date: January 15 2003. http://caes.sourceforge.net.
7. P. Castro, M. Mani, S. Mathur, and R. Muntz. Managing context for internet video conferences: The multimedia internet recorder and archive. *Multimedia and Computer Networks*, 2000.
8. M. Coen. Design principles for intelligent environments. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*. AAAI Press, 1998.
9. J. Collier and M. Collier. *Visual Anthropology: Photography as a Research Method*. University of New Mexico Press, Albuquerque, rev. and expanded edition, 1986.
10. S. Consolvo, L. Arnstein, and B.R. Franza. User study techniques in the design and evaluation of a ubicomp environment. In G. Borriello and L.E. Holmquist, editors, *Proceedings of the Fourth International Conference on Ubiquitous Computing*, pages 73–90. Springer-Verlag, Berlin, 2002.
11. M. Csikszentmihalyi and R. Larson. Validity and reliability of the Experience-Sampling Method. *The Journal of Nervous and Mental Disease*, 175(9):526–36, 1987.
12. S. Hollar. *COTS Dust*. Ph.D. thesis, University of California, Berkeley, 2001.
13. J.M. Hudson, J. Christensen, W.A. Kellogg, and T. Erickson. 'I'd Be Overwhelmed, But It's Just One More Thing to Do:' Availability and Interruption in Research Management. In *Proceedings of the Confererence on Human Factors in Computing Systems*. ACM Press, 2002.
14. S.E. Hudson, J. Fogarty, C.G. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J.C. Lee, and J. Yang. Predicting human interruptability with sensors: A Wizard of Oz feasibility study. In *Proceedings of the Conference on Human Factors and Computing*. ACM Press, 2003.
15. S.S. Intille. Designing a Home of the Future. *IEEE Pervasive Computing*, April-June:80–86, 2002.
16. S.S. Intille, C. Kukla, and X. Ma. Eliciting user preferences using image-based experience sampling and reflection. In *Proceedings of the Conference on Computer Human Interface Short Abstracts*, pages 738–739. ACM Press, 2002.
17. S.S. Intille, J. Rondoni, C. Kukla, I. Anacona, and L. Bao. A context-aware experience sampling tool. In *Proceedings of the Conference on Human Factors and Computing Systems: Extended Abstracts*. ACM Press, 2003.

18. J.M. Kahn, R.H. Katz, and K.S.J. Pister. Mobile networking for Smart Dust. In *ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom 99)*. 1999.
19. O. Kasten and M. Langheinrich. First experiences with Bluetooth in the Smart-Its distributed sensor network. In *Workshop on Ubiquitous Computing and Communications, PACT*. 2001.
20. V. Mahajan and J. Wind. New product models: Practice, shortcomings and desired improvements. *The Journal of Product Innovation Management*, 9(2):128–139, 1992.
21. A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson. Wireless sensor networks for habitat monitoring. In *Proceedings of the ACM International Workshop on Wireless Sensor Networks and Applications*. 2002.
22. M. Mozer. The Neural Network House: an environment that adapts to its inhabitants. In *Proceedings of the AAAI Spring Symposium on Intelligent Environments*, Technical Report SS-98-02, pages 110–114. AAAI Press, Menlo Park, CA, 1998.
23. J.P. Robinson. The validity and reliability of diaries versus alternative time use measures. In F.T. Juster and F.P. Stafford, editors, *Time Goods and Well-Being*, pages 33–62. Ann Arbor, MI, 1999.
24. J. Rodin and E.J. Langer. Long-term effects of a control-relevant intervention with the institutionalized aged. *Journal of Personality and Social Psychology*, 35(12):897–902, 1977.
25. D.A. Schèon. *The Reflective Practitioner: How Professionals Think in Action*. Basic Books, New York, 1983.
26. A.A. Stone and S. Shiffman. Ecological momentary assessment (EMA) in behavioral medicine. *Annals of Behavioral Medicine*, 16(3):199–202, 1994.
27. S. Szalai. *The Use of Time. Daily Activities of Urban and Suburban Populations in Twelve Countries*. Mouton, The Hague, 1973.