Sensor Data Streams

Stephen Voida, Donald J. Patterson, and Shwetak N. Patel

A Short Description of the Method

The increased availability of inexpensive sensors, tremendous processing capabilities (even in mobile devices), high-bandwidth wireless networks, and vast quantities of data storage have made it much more practical to continuously collect streams of low-level data about people and their environments. This approach enables researchers to compile detailed records of various contextual factors surrounding people's interactions with their world. Their locations, physiological states, contact with other people, situated uses of devices, and other digital traces can potentially be recorded and analyzed. Each of these kinds of data can be collected at virtually any frequency, with or without participant knowledge or intervention, and for extended periods of time. Due to the automated nature of the method, a large number of samples can be gathered quickly and with relatively low overhead by the researchers during sessions in the field. However, the degree of automation involved in this method requires a number of pragmatic and analytic considerations, beginning with careful experimental design to ensure appropriate sensor design, and including how the study is deployed, participant training, privacy safeguards, and data storage requirements.

S. Voida (✉)
Indiana University School of Informatics and Computing, Indianapolis (IUPUI),
535 W. Michigan St., IT 475, Indianapolis, IN 46205, USA
e-mail: svoida@iupui.edu

D.J. Patterson
Donald Bren School of Information and Computer Sciences, University of California,
Irvine, 5084 Donald Bren Hall, Irvine, CA 92797-3440, USA
e-mail: djp3@ics.uci.edu

S.N. Patel
Department of Computer Science and Engineering, University of Washington,
Box 352350, Seattle, WA 98195-2350, USA
e-mail: shwetak@cs.washington.edu

While the large amounts of recorded data captured using this technique lend themselves to quantitative analysis—for example, counting the frequency, duration, or variance of signals within an event stream—analyzing and interpreting the data often requires the application of additional qualitative methods.

Its History, Intellectual Tradition, Evolution (in Brief)

While sensors of various kinds have been used to collect data about people and their activities for a long time across a wide variety of scientific domains, the combination of data collection and analysis techniques that we present here primarily evolved from various quantitative data collection methods employed within the domain of experimental psychology. There is a parallel here with psychological research. Psychologists frequently define objective measures for various human or group behaviors and develop various technical or nontechnical instruments to collect data to characterize, quantify, or investigate variance in measurable phenomena based on a specific set of research questions. Similarly, streams of data are a modern equivalent frequently employed in the computer and information sciences. In order to study people, groups, or environments using streams of data, one or more sensing devices are used to automatically collect data on researchers' behalf. The decreasing cost and increasing availability of powerful computational devices and sensors (Weiser, 1991) make it possible to bring a wide array of various sensors to bear on observing a single phenomenon, environment, or individual. Innovations in data fusion, user modeling, and inferencing are then used to aggregate, filter, and interpret the data to answer research questions.

Automated capture of sensed participants' behavior also complements in situ survey techniques like the *experience sampling method* (*ESM*), pioneered by Larson and Csikszentmihalyi (1983). ESM is also known as *ecological momentary assessment or EMA* (Stone, Shiffman, & DeVries, 1999). Like sensor data streams, ESM involves collecting data from a participant while they go about their everyday business; typically, they are asked to answer a question or complete a very short survey when they receive an alarm, alert, or telephone notification. ESM is a sociological data collection technique that elicits information about participants' actions and beliefs *as they happen* and *in the context* of everyday activity, overcoming the recall/recollection limitations of post-interviews or diary studies. Although technology-oriented adaptations of ESM (e.g., Consolvo & Walker, 2003, Intille, Rondoni, Kukla, Ancona, & Bao, 2003) have been utilized for some time within the ubiquitous computing research community, this approach still relies on active responses from research participants during data collection, making the approach less than ideal for longitudinal studies or for research in which introduction of interruptions (e.g., studies of multitasking) or foregrounding of the research study would influence or alter the behavior(s) of interest. Sensor data streams can provide an alternative, or at least complementary, means of collecting contextually situated data as well, but typically through the capture of implicit participant actions or behaviors.

What Questions the Method Can Answer

Sensor data stream collection may be used to understand people's activities, behaviors, or practices by instrumenting:

- a person (Bao & Intille, 2004; Choudhury et al., 2008; Choudhury & Pentland, 2003; Consolvo, Klasnja et al., 2008; Consolvo, McDonald et al., 2008; Liao, Patterson, Fox, & Kautz, 2006; Marmasse & Schmandt, 2000; Olguín & Pentland, 2008; Olguín et al., 2009; Patterson, Fox, Kautz, & Philipose, 2005; Patterson, Liao, Fox, & Kautz, 2003; Philipose et al., 2004),
- an environment, such as an office (Begole & Tang, 2007; Fogarty et al., 2005; MacIntyre et al., 2001; Mark, Voida, & Cardello, 2012), or
- the home (Brumitt, Meyers, Krumm, Kern, & Shafer, 2000; Cohn, Gupta, Froehlich, Larson, & Patel, 2010; Cohn, Stuntebeck et al. 2010; Froehlich et al., 2009; Froehlich et al., 2011; Gupta, Reynolds, & Patel, 2010; Intille, 2002; Intille et al., 2006; Kidd et al., 1999; Kientz et al., 2008; Orr & Abowd, 1999; Patel, Robertson, Kientz, Reynolds, & Abowd, 2007, Tapia, Intille, & Larson, 2004).

The resulting streams of sensor data can be used to help answer a wide variety of questions about the people instrumented with sensors or occupying the instrumented spaces, for example:

- Where do people travel over the course of a day?
- With whom do they normally communicate or collaborate?
- What tools or information resources do they use at various points during the day? When, where, and with whom?
- What routines help to define a "typical" or "atypical" day?
- How healthy are a person's daily behaviors? Is he or she making good health choices?

HCI researchers often use information gathered in this way to learn about people's current behaviors—for example, assessing the duration or frequency of an established technology's use throughout the day. The same data can also form a baseline in studies of how people's behaviors change over time, such the changes affected by the deployment of a new technology among a representative set of end users (e.g., Hutchinson et al., 2003). The ability to make such a comparison is important and enables researchers to draw conclusions about the impact of changing people's tools or environment.

Sensor data often begins as a stream of very fine-grained events. These events include timestamps and durations that are much more accurate than most human observers could possibly record in real time (without having to code a video-taped record of the session after the fact). The kinds of actions that can be recognized depends greatly on how tightly coupled the action is to a specific sensor. For example, a pressure sensor on a chair might suggest an individual's presence in the office, but is less definitive as an indicator of, say, task focus or interruptibility (after Horvitz, Koch, & Apacible, 2004; Mark et al., 2012). The more abstract the action

or the more the action veers from being physically descriptive to being descriptive of intent, the more uncertainty is introduced in the interpretation, even with the availability of relevant sensing. Nonetheless, these kinds of data can answer questions about what people do, relevant stimuli, and physical or social constraints [e.g., the frequency with which members of a social group check in on one another's social network status (Miluzzo et al., 2008)]. Currently, inexpensive and widely available sensors like accelerometers and barometers can quantify things as disparate as motion signatures and changes in air–pressure, providing insight into questions about participants' activity levels, gestures, and other bodily movement. Capacitive, pressure, and acoustic sensors can detect the intensity and number of physical contacts made between people and objects or among objects. Video- and depth-based camera systems can count the number of people or objects in a space, their orientation with respect to one another and gestures or motion taking place, enabling a range of studies about participants' interpersonal interactions and proxemics.

The low-level data collected by various sensing devices to answer these questions can be analyzed to determine higher-level behavior such as activity (Bao & Intille, 2004; Tapia et al., 2004), routing and place-visiting behavior (Patterson et al., 2003), gesture performance (Fan et al., 2012; Westyn, Brashear, Atrash, & Starner, 2003), or communication roles (Wyatt, Choudhury, Bilmes, & Kitts, 2011). Statistical summaries can then be created which may help with the triangulation of other methods around phenomena of interest or to generate predictive models. (See Looking Back: Retrospective Study Methods for HCI.)

Historically, questions of *where*?—that is, questions that demanded location-awareness capabilities—were some of the earliest research questions thoroughly explored by the ubiquitous computing community. Many projects have been extremely successful in answering questions about the locations or places that a participant has visited (Abowd et al., 1997; Liao et al., 2006, Patel, Truong, & Abowd, 2006, Patterson, 2009), other people with whom they have been in close proximity (Want, Hopper, Falcão, & Gibbons, 1992; Choudhury & Pentland, 2003; Choudhury et al., 2008), and the participant's routines and patterns of mobility (Consolvo, McDonald et al., 2008; Liao et al., 2006). The level of fidelity of these streams of information varies greatly, depending both on the geo-location technology being used to collect the data and the level of fidelity at which the data are represented and stored; Hightower and Borriello explore many of these considerations in their overview article in *IEEE Computer* (2001b) and Varshavsky and Patel provide an updated perspective (Ubiquitous Computing Fundamentals 2009).

Research Questions with Various Units of Analysis

One advantage of using streams of data as a means of understanding people's activities and behaviors is that the technique can be used to answer research questions across a range of units of analysis. The main trade-off in realizing this analytic

flexibility is that the "right" number, type, and combination of sensor data sources used to collect the data do vary based on the specific scope of research and the research questions that are being asked.

Egocentric Sensor Data Streams

Sensors focused on monitoring the movements, activities, and interactions of a single individual can answer questions at an *egocentric* unit of analysis. Studies that collect log data from an individual's use of a computer, a smartphone, or a different piece of technology; measure a person's physiological state; monitor their movements, location, or interactions using the sensors on a mobile phone; or collect information about activity within a private or semiprivate space could all be described as *egocentric* in this way. This is a typical study design as derived from the experimental and engineering psychology tradition and can be replicated over a population sample to look at patterns or trends. Egocentric data streams are often used to answer research questions like *How do people allocate their time or attention*? *How is a person's mental state or mood affected by real-world stimuli*? *How do electronic communications or mobile computing interactions affect daily routines*? *How do people's own understanding or interpretations of their activities, colleagues, or environment differ from what a ubiquitous computing application or tool is able to automatically sense*?

Group-Centric Sensor Data Streams

Sensors can also be used to instrument—and answer research questions about— groups of people at the same time. This *group-centric* approach can involve simply capturing the same signals as for a single person, but across a group over the same window of time, or it might involve deploying a broader set of environmental or infrastructural sensors in a shared/community space or collecting data about more interpersonal types of interactions (e.g., sociometer-style data, after Choudhury & Pentland, 2003). This style of sensor data stream-based research is often utilized for understanding research questions about organizational and group dynamics, such as *How often do members of this group interact with one another*? *What do these interactions entail*? *How do power relations manifest in different kinds of work environments or work teams*?

Space-Centric Sensor Data Streams

Finally, researchers can use streams of sensor data to answer questions about how spaces are used, irrespective of their particular inhabitants, given appropriate instrumentation of a space. This kind of *space-centric* research is often carried out in semipublic or public spaces, such as museums (Hornecker & Nicol, 2012; Sparacino, 2003). A standard approach to *space-centric* sensing is to employ a wide variety of

sensors to detect signals from a place of interest. Commonly used devices include high information density sensors, such as cameras and microphones, as well as low density sensors such as pressure sensitive floor tiles (Orr & Abowd, 1999), passive infrared (PIR) motion detectors, or RFID readers. If an experiment requires a large number of sensors to be distributed throughout the environment, special networking support (either wired or wireless) may additionally need to be installed in the space to move the data from the site of collection to the storage and analysis servers. Maximizing these sensors' area of coverage often necessitates that they be mounted on a room's walls, ceiling, or doors so that the sensors have unobstructed views of a wide area; the aesthetic impact of these placements can lead to reduced adoption by homeowners or altered behavior on the part of the participants because they are reminded that they are being observed. The use of high information density sensors in certain spaces often raises concerns about the balance between the value of the proposed use of the data stream and impact of the resulting surveillance, particularly in home settings.

The alternative, which is low density sensing, includes the use of many simple, low-cost sensors, such as motion detectors, pressure mats, break beam sensors, and contact switches, to determine activity and movement. The principal advantages of the low-density approach are lower bandwidth and processing needs, and potentially reduced privacy concerns. In some cases, limitations imposed by the need to take into account participants' privacy risks can prohibit the use of sensors that might be beneficial for answering specific research questions, such as putting cameras in bathrooms. In such a case, low-density sensors may be the only alternative. Notable examples of space-centric sensing that were designed for multiple deployments include Georgia Tech's Aware Home (Kidd et al., 1999; Kientz et al., 2008) and MIT's House *n* (Intille, 2002; Intille et al., 2006; Tapia et al., 2004).

Researchers have addressed some of the limitations of *space-centric* research, by inventing a technique called "Infrastructure Mediated Sensing" or IMS. IMS involves using existing home infrastructure to detect activity within a home (Patel et al., 2007; Patel, Reynolds, & Abowd, 2008; Froehlich et al., 2009; Gupta et al., 2010). Electrical, plumbing, and HVAC systems, as well as natural gas piping and computer networks, are already widely deployed in much of the world. This existing infrastructure can be instrumented to sense when the home occupants engage in activities that utilize those infrastructures (e.g., monitoring the flow of electricity throughout a home to detect when individual appliances are being used). In addition, those infrastructures can be used to communicate the detection signals through a home. This approach reduces the need for researchers to install many sensors throughout a space, and enables researchers to answer questions like *How are the occupants of a home spending their time throughout the day and night*? *Is a senior adult living by herself continuing to maintain healthy levels of physical activity*? *What is the impact of ambient feedback promoting environmental awareness on cooking*, *cleaning*, *and hygiene activities within different types of families*?

There is a parallel between the ways that researchers use IMS to conduct infrastructure-based sensing with the ways that egocentric sensing is carried out

using smartphones. In both cases, researchers have piggybacked sensing onto infrastructures that were already deployed but that were not originally meant for data collection. In a hybrid example, Isaacman and colleagues (2012) developed a technique that bridged between egocentric data streams and IMS by using cell phone billing records to analyze large-scale commuting patterns.

The attractiveness of IMS is that millions of locations and people, in effect, have been instrumented by highly refined and well-understood technologies. Additionally, this instrumentation occurs across a wide (but not exhaustive) variety of places and people. The amount of additional effort that must be done to leverage these infrastructures, however, varies greatly. Newer approaches, such as IMS, require more effort than location-based services, for example. Depending on the effort required to enable data collection through these infrastructures, these techniques can enable scaling that would not be possible with traditional sensor deployment approaches.

What Data Are Captured by Sensor Data Streams

Sensor data streams are similar to other kinds of approaches that are commonly used to capture and analyze user behavior, but are distinguished by their source (the physical world) and their level of interpretive fidelity. Sensor data streams are time-stamped data that flow from the physical environment surrounding and permeating user interactions. The range of sensors that might be employed in these kinds of studies are virtually limitless; sensing toolkits often facilitate data collection using a variety electrical switches, motion sensors, pressure sensors, voltmeters, photometers, thermometers, moisture sensors, proximity sensors, RFID tags and beacons, microphones, or cameras (e.g., Greenberg & Fitchett, 2001; Villar, Scott, & Hodges, 2011). New sensors are continually being developed and incorporated into consumer devices and research-oriented toolkits all the time.

Data streams might also come from log files or usage statistics from instrumented user interfaces (see also Looking Back: Retrospective Study Methods for HCI). These logs can be thought of as "virtual sensor" data streams, as these data *are* being collected in situ during an interaction but are the result of observations of the *digital* world as opposed to the physical world. Another source of "virtual sensor" data streams can come from the nonautomated human observation of interactions in which time stamped annotations of events are also produced—albeit with considerably more overhead and somewhat less precision (e.g., Mark et al., 2012). This approach can include observations of recorded audio or video data and differs from sensor data streams in that the quantity of information is typically much smaller, although the kinds of data collected can be much richer. For example, a log file analysis might present a user clicking a mouse on a "save" button as a "virtual sensor" event. A human observer might record very similar data from a video recording by annotating that the person "saved a document" at a particular moment in the video. A typical sensor data stream, however, would record the sound of a mouse click, some evidence that the person and the computer were collocated and perhaps

some fluctuation in the power used by the computer at the same time. The boundaries between log file analysis, human observation, and sensor data stream-based data collection are not clearly delineated, however, and many of the same data storage, processing, and analytic techniques could be used to make sense of the information gathered using any of these data sources without much difficulty. Clearly, these data sources could strongly reinforce and inform one another, as well.

This style of quantitative data collection and analysis to understand human activity has been utilized within the HCI community for some time, most commonly in the analysis of software interactions by treating existing data logging capabilities as virtual sensors or by explicitly adding the ability to log new events, such as those that occur in desktop computing systems (Brdiczka, Su, & Begole, 2010; Hutchings, Smith, Meyers, Czerwinski, & Robertson, 2004; Kaptelinin, 2003; MacIntyre et al., 2001; Nair, Voida, & Mynatt, 2005; Stumpf et al., 2005), in awareness and social networking tools (Begole, Tang, Smith, & Yankelovich, 2002; Monibi & Patterson, 2009; Patterson, Ding, Kaufman, Liu, & Zaldivar, 2009; Tang & Patterson, 2010), or on the World Wide Web (e.g., Perkowitz, Philipose, Fishkin, & Patterson, 2004).

It is often helpful to augment automatically collected physical and virtual sensor data streams with other data collection techniques. This can be done to measure phenomena that span both the physical and virtual worlds, such as workplace interruptions (Bailey & Iqbal, 2008; Bailey & Konstan, 2006; Bailey, Konstan, & Carlis, 2001; Horvitz et al., 2004). It can also be done to provide clearly indexed points in the mountains of collected data that can be used for subsequent qualitative interviews, somewhat reducing the burden on an interviewee of recalling what was happening as the data was being collected (see Chapter on Looking Back: Retrospective Study Methods for HCI).

For the balance of this chapter, we will focus on the use of physical sensors as a way of studying people and their behaviors. While many of the techniques that we discuss generalize to the user of virtual sensors or log analysis to a greater or lesser extent, we refer the readers to other chapters in this volume for a more in-depth treatment of these specific topics.

Sensor Data Streams and Context-Aware Computing

The use of data streams as a research method also shares a number of similarities with *context-aware computing* research efforts, including the data sources and analytic techniques. While a complete overview of context-aware computing is beyond the scope of this chapter, several good surveys have been published, including Baldauf, Dustdar, and Rosenberg (2007); Bolchini and colleagues (2007); and Hong, Suh, and Kim (2009). Generally speaking, context-aware computing is a form of interactive computing in which a user's implicit behavior—that is, their location, their physical activity, or their interactions with other people—or the environment in which a system is being used can both serve as alternative or auxiliary inputs to the system (Dey & Abowd, 2001; Salber et al., 1999). The central differences between

collecting sensor data streams to understand user behavior and leveraging streams of data as inputs for a context-aware computing system are as follows:

- the research goals (studying user behaviors versus developing interactive systems),
- when the collected data are processed and analyzed (as part of the analysis versus in real time), and
- whether the supporting technologies are primarily intended to *generate* a user model versus *predict* user behavior based on a preexisting model.

Generally, when collecting sensor data streams, there are no real-time processing or inference-generating requirements—such processing can usually be handled offline once collection is complete. This decouples the analytic concerns from data collection, relaxing constraints that are often imposed on most context-aware computing systems. In context-aware computing, data collection and analysis often need to be tightly coupled so that the results of the data stream analysis are available as soon as possible following the physical interactions that generated them. Interactive systems require that researchers minimize latency in analyzing sensed data, which often necessitates aggressive data pruning, matching against coarse, preexisting user models, or the use of heuristic approximations; data collection for understanding users does not necessarily need to pursue these kinds of optimizations. Even given this distinction, many of the inferencing techniques that have been developed within the ubiquitous computing and context-aware computing research communities may still be of interest to researchers who are more interested in using streams of data to learn about their participants, their actions, and their surroundings, since they represent a useful set of tools for transforming raw data into higher-level constructs.

Limitations of the Approach

While sensor data streams are useful for gathering large volumes of continuous, high-fidelity data about the way that people behave and interact in the real world, there are a number of important considerations and limitations associated with capturing, processing, storing, and analyzing sensor data streams:

The primary weakness of using streams of sensor data as a technique for learning about participants is that **sensor data generally does a poor job of answering questions of *why*** things have happened in the real world. This limitation goes above and beyond the self-evident challenges in detecting *what* is happening, since a single sensor reading can often be interpreted in many different ways. Even accurately sensed participant actions (or interactions or contexts) can only record the physical impact of what happened; these kinds of readings provide little insight about the intention behind the actions or the broader aims, goals, or internal, mental states of a participant. One way that this limitation can be minimized is by *triangulation* or *data fusion*, that is, combining many data streams and methodologies together.

The **phenomena to be measured or observed must be well understood** at the outset of the study, and an appropriate sensor or sensors must be acquired (or, in many cases, built from scratch) and then deployed to capture the "right" set of data to measure or observe those phenomena. Even with a well-designed sensor deployment, additional techniques may be required to automatically process the raw data so that the phenomena of interest can be isolated, contextualized, and measured. This gap between what can be captured in a sensor data stream and what actually took place in the real world may necessitate a probabilistic interpretation of whether or not an event has occurred and may impact the knowledge generated (e.g., understanding whether or not a gap in logged computer activity is due to an external interruption or simply because a participant is rereading a prompt or looking for additional information). Because the data collection instruments (the sensors) are such an integral part of the research design, when using this approach, it is often more difficult to make changes to a sensor-based data collection protocol than it is for a corresponding, human-administered study (e.g., fielding interviews or surveys).

Sensors (and their associated processing and recording technologies) **have limitations in the quality of data they can collect** during a study. They can sometimes be expensive (limiting the amount of sensing redundancy that can be employed), are sometimes unreliable, are often prone to generating noisy output, and are often limited in the scope and range of actions they can record. Sensor-based research protocols should employ a combination or suite of sensing technologies to balance these drawbacks; however, this does increase the overhead of setting up, storing, maintaining and analyzing the data that are produced, in addition to the design and implementation work necessary when sensors need to be created or modified. However, one of the central tenets of using sensor data streams to conduct user research is that the results of a sensor data stream study can be no better than the sensed data.

If the goal is to capture *everyday* activity or behavior, **care must be taken to select sensors that can effectively capture the right kind of data at the right fidelity and minimize intrusiveness and discomfort** for the study participants. Creating a study design that requires participants to dramatize, enlarge, or vary behaviors that are part of their normal routine in order to detect the activity with sufficient accuracy (e.g., walking through a particular area of a room to indicate presence) undermines the ecological validity gained by using this method. Furthermore, there is a balance to be struck between collecting data unobtrusively and collecting data without a participant's awareness or consent. Care must be taken to clearly communicate at least a high-level overview of the sensors' data collection capabilities to participants and to be up front about *what* data are being collected; *when* or *how often* they are collected; *how* they will be stored, shared with researchers, and analyzed; *whether data can be excluded or removed* from the study; and *what confidentiality or anonymity protections* will be in place to minimize risk to participants' privacy.

Many sensing technologies will produce very large streams of (typically low-level) **data** over the course of a moderate-length deployment. Although this can be a useful characteristic of this style of research in that it quickly produces a large

corpus of data, managing these large data sets can be challenging. Transporting the data across networks can take significant battery power and time and can expose the data to privacy threats. Streams of sensor data may need to be aggressively filtered, aggregated, and timestamp-synchronized before any significant analysis of the output can be carried out.

Sensors also introduce a level of technical complexity to a research project that is not always present when other methodologies are used. These kinds of studies often require a moderate level of technical expertise in order to select appropriate sensors; to configure their recording frequency, fidelity, and output representations; to manage data storage, either locally to the sensor or on a network-connected server; to provide technical support when sensors or recording media inevitably fail; to parse or convert the recorded data into a format suitable for analysis; to protect the data from inadvertent release; and, ultimately, to analyze the results.

How to Do It: What Constitutes Good Work

Several steps are necessary for designing a sensor data stream collection study to ensure that it will effectively answer the desired research question(s). Issues specifically related to data sensor streams include:

- Generating the research questions and planning how to analyze the data streams
- Building, acquiring, or provisioning the sensors
- Determining how frequently and at what level of fidelity to collect data samples
- Installing the sensors
- Storing the data representation
- Making sense of the collected corpus of data

At each of these steps, researchers must make specific decisions about their research design, based on the kinds of research questions that they wish to answer and the limitations inherent in using automated systems to conduct data collection.

Generating the Research Questions and Identifying the Data to Collect

Research based on sensor data streams shares many characteristics with other kinds of in situ empirical studies. Depending on the research questions that the study is designed to answer, which unit of analysis (egocentric, group-centric, or space-centric) is chosen, the recording capabilities of the specific sensor devices used, and the frequency and fidelity at which samples are collected, there may be more or less significant risks to participants' privacy and concerns raised about the intrusiveness of the research.

One major difference between data collection by sensor versus in-person observation or "shadowing" is that with a human observer, regardless of their skill level or subtlety, it is almost always obvious to the study participants when they are being observed and (roughly) how much detail can be perceived about their activities by the researcher. With sensors, particularly ones that are small in size, worn for long periods of time, or "invisibly" embedded into a space or its supporting infrastructure, it is very possible for both consented participants and incidental passersby to have their actions recorded without their knowledge. Having sensors "disappear" into the background does have some advantages in terms of reducing the chance of collecting data that are distorted or amplified as part of an intentional or unintentional performance for the researchers by the study participants. However, there are also very real ethical issues with capturing data from participants without their knowledge.

As with any other observational study, it is essential to be up front with participants about the scope, duration, frequency, and fidelity of the data that will be captured during the study. It might be advisable to have sample data sets available to share with the participants during solicitation of informed consent, so that participants will have a grounded sense about what the researchers will be able to "see." It may also be advantageous to provide participants some mechanism for revoking their consent to be recorded—for example, a button that will suspend data collection for some period of time or delete the data collected by the sensor for some number of minutes immediately preceding the button press.

Many institutional review boards have guidelines about getting participants' informed consent when automated sensing or recording devices are used for research purposes. Learning what expectations your institution has for carrying out this style of research at the outset of the study design can be an invaluable asset and prevent administrative delays later in the process of data collection and analysis (see also Chapter on Research and Ethics in HCI).

Building, Selecting and Acquiring, or Provisioning Sensors

Selecting the data sensor(s) used to acquire the data can be a function of cost, availability, technological capability, intrusiveness, or methodological needs. Good research requires that researchers either choose a sensor(s) that can reliably and accurately sense the desired phenomena or they must construct (and validate) their own custom sensor for this purpose. Broadly speaking, there are two categories of sensors that can be used in this type of research: sensors that are worn or carried by the participants, and sensors that are deployed in a particular space and record the activities of that space's occupants (Figs. 1 and 2).

The downside in instrumenting a space, as opposed to instrumenting a participant, is that data can only be collected within the instrumented environment. When the research questions require egocentric data collection, wearable sensors might be

Fig. 1 A pair of sensors that are worn/carried to collect data about a person's stress level throughout the day. The smartphone captures audio of a person's voice to look for indications of increased stress levels; the wristwatch-style sensor is a commercial electrodermal analysis device used to identify episodes of stress and arousal based on variations in a person's skin conductivity (Poh, Swenson, & Picard, 2010)
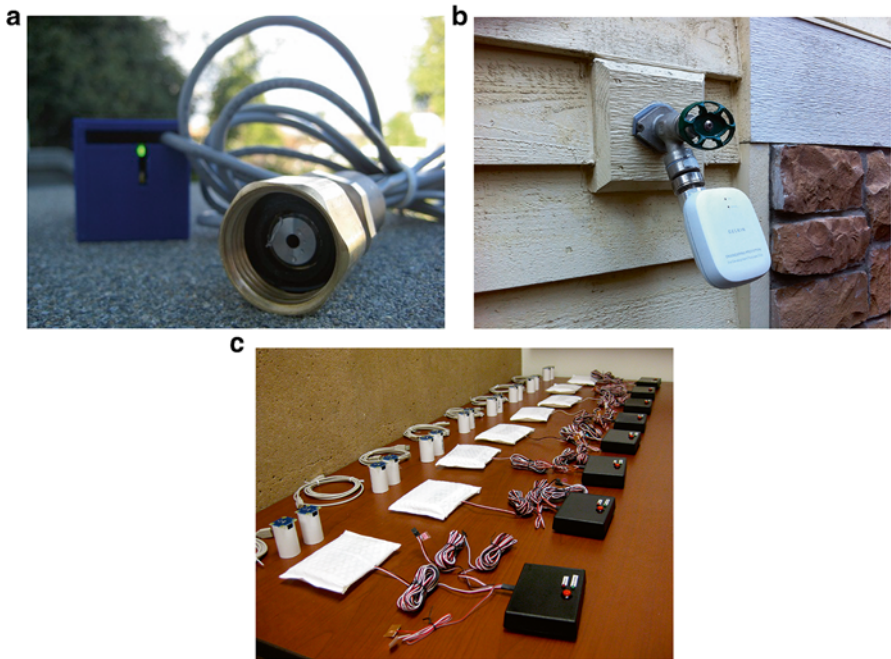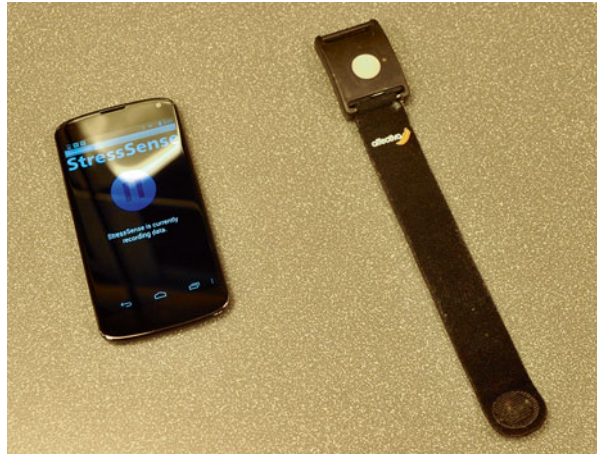

Fig. 2 Examples of sensors that have been deployed in physical spaces to collect data about the spaces' occupants. (**a**) and (**c**), two different versions of an infrastructure-based system for identifying activities related to water use in a home. (**b**), Sensing suites that collect data about multitasking activity in an office, including interruptions at the door, activity with physical artifacts on a desk, use of the telephone, and presence of a person (and their posture) seated in the office chair

required in order to collect data samples over the full range of activities, behaviors, and interactions that a person experiences throughout a day (e.g., at home, in the car, at the café, and at work), regardless of the other drawbacks to this approach.

Some technologies collect data more effectively in some environments than in others, so another consideration in the selection or design of sensing technologies are the contexts in which data collection is envisioned to take place. For example, GPS can provide good outdoor positioning information in open areas, but becomes much less useful when location is needed in situations with no power, in "urban canyons," indoors, or when outdoor episodes are too short to acquire satellite lock. In these situations, it might be more appropriate to utilize alternate positioning technologies (e.g., Wi-Fi or Bluetooth signal strength triangulation, deployment of infrared or ultrasonic location beacons) in order to capture desired movement or location information (Hightower & Borriello, 2001a; Hightower & Borriello, 2001b). In some cases, context-aware computing middleware exists that can select from the best data source given the sensor hardware limitations and current conditions. Intel's Place Lab is one such platform that was developed for capturing location information (LaMarca et al., 2005); more recently, this kind of location-sensing data fusion functionality has been integrated into mobile computing operating systems, including Google's Android and Apple's iOS.

The degree of intrusiveness of the sensing devices can also have a substantive effect on the success of the study or the comfort level of the participants (Klasnja, Consolvo, Choudhury, Beckwith, & Hightower, 2009). Some physiological sensors (e.g., heart rate monitors, galvanic skin response meters, pupil dilation detectors) can be uncomfortable for participants to wear for extended periods of time, or they may make it more difficult for the participants to carry out their everyday tasks due to physical limitations or social concerns. On the other hand, installation of a few, relatively "invisible" sensors in the participants' home or work environments, such as the kinds of infrastructure-mediated sensing developed by Patel and colleagues (Cohn, Gupta et al., 2010; Cohn, Stuntebeck et al. 2010; Froehlich et al., 2009; Froehlich et al., 2011; Gupta et al., 2010; Patel et al., 2008, Patel et al., 2007) or the pervasive sensors used in the MIT House *n* project (Tapia et al., 2004), can help to improve the ecological validity of the study, as the sensors' presence is less likely to influence or affect the participants' behavior.

Determining How Frequently and at What Level of Fidelity to Collect Data Samples

Once sensors have been acquired, constructed, or provisioned, there are a number of key considerations about how those sensors are configured to collect data during a study, namely how a balance is maintained between the sampling rate of the data streams and storage/bandwidth, processing, and power requirements. These

concerns are pushed to the forefront when using sensors incorporated into platforms like mobile phones. Aggressive sensing can:

- quickly exhaust the phones' limited on-board storage;
- lead to excessive data plan usage if the collected data is continuously transmitted to a server via the phone's cellular radios; and
- exhaust the phone's battery rapidly, leading to a loss of potential data and annoyance for an participant who expects to be able to use to phone for other purposes throughout the day.

The question of how much detail is captured by a sensor is bound up in the notions of data collection fidelity. In many cases, a sensor may be able to collect very accurate readings at very fast sampling rates, but this may lead to the collection of more data than is desired or necessary, given a particular research question(s). Usually, it is possible to generalize, abstract, or blur the data to reduce the processing and storage requirements or to protect the privacy of participants while still achieving a study's goals. Protecting participants' privacy may require a more nuanced evaluation, however, as behaviors may be able to be completely reconstructed even from low-fidelity data. Even if data collection with reduced fidelity is utilized, participants may be uncomfortable because they know that the sensor *can* collect higher fidelity data and researchers' assurances that this is not being done may be met with skepticism. This can be the case, for example, if video cameras are deployed as surrogates for motion detectors. These issues may create a negative reaction to being a subject of surveillance (Klasnja et al., 2009) and impact the naturalistic collection of data. Choudhury and colleagues explicitly discuss this tradeoff with regard to their collection of audio data for analysis of speech prosody using their *sociometric badges* (2003), a device worn around the neck to automatically activity—specifically, face-to-face conversations—during the work day.

Furthermore, a single unit of contextual data can have multiple interpretations, and the researchers' choice of which interpretation is modeled can have a significant impact on how participants perceive their privacy to be protected, as well as the kinds of analyses that can be carried out with the data. Location data, for example, can be represented in multiple possible ways (Liao et al., 2006): it could be represented as a latitude/longitude/altitude tuple, it could be represented as a geo-located address (e.g., "Donald Bren Hall"), it could be represented as being inside or outside of a particular municipal locality (e.g., "in the city of Irvine" or "on the UC Irvine campus"), or it could be represented as a semantically defined location (e.g., "at work"). Fine-grained location data can provide valuable insights about a participant's behavior, including their daily routines, paths of travel, or whether or not they cross paths with or are close enough to communicate with other individuals. But these data might also serve to identify an individual participant when analyzed or presented. The data may also reveal details about a person's activities that the participant would rather not—or does not intend to—disclose. Hightower and colleagues provide a good introduction to this problem (Hightower, 2003), and some data about user concerns in this regard has been collected, as well (Patterson et al., 2008).

Generally, however, the problem of interpreting contextual data applies to many different domains, including activity recognition, gesture recognition, etc. One solution that has been developed to navigate these concerns is to vary the fidelity of data collection based on input from other sensors. For example, by only collecting GPS location readings in an egocentric study when accelerometer readings indicate participant motion, the amount of data collected can be reduced with little to no impact and battery life can be extended. Another common solution is to save data on smartphones until a participant is near a Wi-Fi access point at which point the data is transferred to a storage server. This eliminates the need to use expensive and battery draining cellular connections, and provided there is sufficient on board storage to hold data between participants' encounters with Wi-Fi, has little impact on the data collection.

Installing the Sensors

The number of sensors required for coverage of a large space presents an inherent complexity hurdle. Installation and maintenance of (typically) tens of sensors in a home, or hundreds to thousands of sensors in a larger building such as a hotel, hospital, or assisted-living facility, results in high labor costs during installation, and an ongoing sensor network management challenge during routine operation. Furthermore, these sensors will require some type of power source, such as batteries, which create an additional maintenance schedule and costs. It is also difficult to balance the value of in-home sensing and the complexity of the sensing infrastructure. One example that illustrates this is the Digital Family Portrait (Mynatt, Rowan, Craighill, & Jacobs, 2001; Rowan & Mynatt, 2005), which communicates activity information from an elderly person's home to a remote caregiver. In the system's study, movement data was gathered from a collection of strain sensors attached to the underside of the first floor of an elder's home. The installation of these sensors was difficult, time-consuming, and required direct access to the underside of the floor. Though the value of the application was proven, the sensors' complexity limited the number of homes in which the system could be easily deployed. If the value of the proposed sensor deployment is not clear in advance, a Wizard-of-Oz method (Dahlbäck, Jönsson, & Ahrenberg, 1993) can be used to mitigate the deployment risk. This approach tests the outcome of a perfect sensor system by using a researcher to simulate the sensing infrastructure—that is, to visit the field site in person and manually record data of a corresponding type and level of fidelity as would be expected from an automated sensor. If the results are promising, then actual sensor deployments can proceed in earnest.

Storing the Collected Data

One of the more pragmatic concerns about conducting a study using streams of data are decisions about how—and where—the collected streams are stored between the

time at which they are collected and when they are aggregated and analyzed by the research team. The options available to researchers here are often directly impacted by the decisions made about how frequently and at what level of fidelity data are collected.

The two most common options for capturing a data stream are (1) to record the sequence of observed events on a device connected to (or nearby) the sensor data source and to collect a copy of the stream at various points of time during and/or at the conclusion of the study or (2) to continually transmit collected data back to a central server via a network connection. Both of these options have pros and cons that must be considered. Although creating a local cache of study data is often more technically straightforward, this approach can be problematic when large volumes of data (e.g., video or audio) need to be stored, and relying on multiple storage media to capture study data can increase the complexity of data collection. Transmitting data to a research fileserver that is connected to the Internet can make it easier for researchers to monitor the volume of data in the stream, dynamically add storage capacity, and maintain continuous backups of the collected data. However, this approach requires that each of the sensors be able to connect to the network and necessitates additional technical support to ensure that gaps in connectivity do not result in skipped data and that participants' data are reasonably secured while in transit. This technique also requires that researchers consider whether/when data might be aggregated, filtered, or pruned to protect participants' privacy, to weigh demands on sensors' batteries, and to think about whether data transmission costs might come into play (e.g., if large volumes of data need to be uploaded over a cellular data connection).

Given the falling cost and rising capacity of digital storage, as well as the increased processing capabilities of tiny mobile computing devices, storing as much data as possible at the point of sensing (e.g., on a smartphone) would, in many cases, be beneficial to researchers. However, this is not the only option. Especially when using a large number or wide variety of sensor data sources, aggregating data from multiple sources into a single "stream" at the time of collection can dramatically simplify the subsequent analysis. Creating a single event stream eliminates (or at least minimizes) the overhead involved in temporally synchronizing observations from different sources. (Synchronization can also be accomplished by planting various markers in the data, similar to the way that a movie production clapboard leaves a visual trace in the film and an audible sound in the audio recording. However, these independent data streams still need to be aligned using these markers before they can be manipulated; this process can be significantly burdensome when many streams need to be synchronized.)

Making Sense of the Collected Corpus of Data

Researchers can conduct various post hoc analyses of data streams using standard statistical software packages. A common example would be comparing the

frequency or distribution of sensed events under different conditions, revealing correlations between these conditions and the observed behaviors. However, due to the volume of sensor data that are collected by automated systems and the noise inherent in sensor data sources, it is often more valuable to use machine-learning techniques to classify sensor data into higher-level representations of participants' behaviors or to perform hypothesis testing. Examples of such classification include identifying when particular events occurred, classifying the mood of participants at any given time, or figuring out where a participant is. Some of the machine-learning techniques that are valuable for making sense of sensor data streams include data filtering and smoothing [e.g., particle or Kalman filtering (Krumm, 2010)], activity detection (Philipose et al., 2004), and sensor fusion [e.g., using Hidden Markov Models (Patterson et al., 2005), naïve or dynamic Bayesian Networks (Fox, Hightower, Kautz, Liao, & Patterson, 2003), or time series data analysis (Liao, Patterson, Fox, & Kautz, 2007)].

Oftentimes, the first step in analyzing the data collected during a sensor data stream study is to classify the data stream into segments or to identify particular events in the data (a special case of classification). This would be true of activity recognition and gesture recognition studies, for example. These classifications, in turn, may become the focus of analysis by conducting studies of their frequency or purpose. In the case of context-aware computing, they may serve as input to another system component, for example, presenting search results or map directions filtered by the user's location. In either case, it is important to carefully evaluate the classification algorithm based on machine learning techniques before advocating its broader use.

User Modeling and Event Detection

The first step in making sense of streams of sensor data is to clearly identify which data will serve as inputs to the classifier. This is equivalent to identifying the dependent variables or the "features" of a machine-learning task. Then, it is necessary to establish the classification categories or, alternatively, the metric(s) that is going to be detected; these become the independent variables or output of the classifier. A classification task would be one in which a reading of dependent variables is mapped to one of N categories (e.g., mapping accelerometer readings on a smartphone to one of {sitting, standing, lying down}). A continuous variant of classification, regression, would map the reading to a real-valued metric (e.g., mapping a skin galvanic response sensor and a heart rate monitor to a number from −1.0 to 1.0, representing emotional valence).

If the researcher has access to several (and, hopefully, many) examples of validated, true mappings from dependent variables to the independent variable—sometimes called a "gold standard"—then it is possible to use supervised machine learning to construct a classifier. In this case, the mapped values form a training set from which a classifier can be automatically trained. If the researcher does not have

any examples from which to train a classifier, unsupervised machine learning, or clustering, is an option.

In both cases, it is important that the researcher consider how the training data generalizes, and if that is the right generalization for the study. For example, a gesture classifier that is trained on one person's gold standard data should do a good job identifying gestures in un-annotated data from the same person. However, because the training data only came from one person, it may not work well when applied to a different person's unannotated data. This would be true if gesture performance varied greatly among people even for the same gesture. Another example of when generalization might fail is if accelerometer data collected from one type of hardware is used to train a classifier. That classifier may be ineffective if it is given unannotated data from different hardware as an input. This would be true if due to calibration, sensitivity or something else, different devices produced different readings for the same phenomenon. A closely related research concern is to be sure that the classifier is being trained on data that is generated by the correct phenomena; although this may seem obvious, this principle can be violated in subtle ways. For example, if a person-tracking application is trained to classify motion detector activation as being indicative of a person's presence in a particular room, this classifier will not result in correct behavior if the training data contains examples of motion detector activation that was caused, instead, by the family pet.

Unfortunately, gold-standard data is often available in limited quantities; it is time-consuming and can be expensive to capture and accurately label sensor data streams, especially across a diverse or distributed population. Nonetheless, the goal of user modeling is to create an algorithm that can both be shown to classify the available gold-standard data correctly and that will work just as well with real-world sensor data. One problem that can arise in user modeling is known as "overfitting." That is, it is possible to create or tune a classifier that works very well on the gold-standard data, but will not work for as-yet-unseen data. Overfitting is typically addressed through a process of cross-validation, described below.

Validating User Models

Cross-validation refers to the process of evaluating a classifier by dividing a single set of gold-standard data into a number of subsets and then using different permutations of these subsets as training and testing data across a set of validation experiments. Each permutation is called a "fold" and corresponds to a single sample experiment; multiple experiments are carried out, with the number of experiments matching the number of folds used. The folds are often created to exhaustively cover all available annotated data—testing each data point one time as part of the process and using it as training data the rest of the time. An example would be tenfold cross-validation. In tenfold cross-validation, the gold-standard data is divided into ten subsets of equal size. A total of ten experiments are run, such that in any one experiment, nine of the subsets are used to train the classifier. Once the training phase is complete, the tenth, "held-out" subset is classified, but with the preexisting

annotations on this part of the data hidden from the classifier. Finally, the annotations that are computed by the classifier for the training subset are compared to the actual, hidden annotations, providing a way to assess the degree to which the classifier was correct. Typically, the ten experiments will produce accuracy metrics, which can then be statistically analyzed for mean, variance, etc.

Generally speaking, it is possible to choose any number of folds from two through the number of data samples that have been collected. Broadly speaking, this is referred to as "n-fold" cross-validation. Regardless of the value of n that is chosen, each data sample is only tested (classified) once. A special case is "Leave One Out Cross Validation" (LOOCV). This is when n is chosen to be equal to the number of data samples, N. This requires the most computational power, as the classifier must be learned N times, and this process of training the classifier is typically far more computationally intensive than testing. Many software frameworks exist to automate the process of creating and testing using cross-fold validation, including the popular WEKA machine learning toolkit (Hall et al., 2009).

When using n-fold cross-validation to train and test classifier, it is important that no information "leaks" from the training set to the testing set during cross-validation; that is, that samples assigned to the training set are selected to be as independent as possible from the samples used for training. Even information leaks that appear to be very minor can create impressively accurate classifiers, which would completely fail during a real application of the same technique. How might such a subtle leak occur during experimentation? One example would be if sensor readings (e.g., temperature readings from an instrumented apartment space) were collected and stored as a difference from a running average of all previously collected data points. In this case, independence is broken; when the gold standard is separated into n folds, any given testing fold will be related to the training fold through the common mean implicitly used to encode the data. In this way, the training data "knows" something about the data to be tested. One way to solve this problem is to carefully consider the representations that are to be used for stored sensor data and to separate raw data samples into folds for cross-validation *before* calculating any aggregate statistics across the fold or subsets of the fold.

Some special cases require separating the data in unique ways in order to enforce the right kinds of independence. For example, in a gesture recognition problem, the most appropriate way to create folds might be one in which samples collected from different individuals are placed in their own fold. While this may create folds that vary in the number of samples that they contain, it will ensure that the classifier is learning gestures that will generalize to unseen *people* instead of to unseen samples from known people. This would help to support a claim that this technique would ultimately work when applied to a person for whom training has not been individually conducted. Another special case relates to the temporal nature of data streams: a sample collected from a sensor at a particular point in time is often explicitly or implicitly related to the samples collected just before or after. As a result, separating data into folds where the testing samples have a consistent temporal relationship to the samples in one or more folds might produce abnormally good results. Perhaps, in this case, creating folds based on day is more appropriate, or perhaps creating

training sets of the last hour's worth of data applied to the next data sample would be more appropriate. Analysis tools like WEKA offer a variety of capabilities for partitioning a data set into folds based on various criteria and can be particularly helpful when taking advantage of these techniques.

The nature of cross-validation assumes that all data is available for use in an offline context in which it can be repeatedly accessed in different sequences as required. A common application of this approach is to use extensive cross-validation testing to create one classifier whose parameters are fixed and then to apply it in an online setting in a testing-only configuration. Such a classifier might then be used for a context-aware application, for example. The machine-learning literature includes extensive variations to the above approaches, which relax, in various ways, nearly every possible dimension of the described approach to achieve different effects, maintain sample independence, or account for different kinds of input data. (A good introduction to issues related to creating classifiers in the machine-learning style appears in Langley, 2000 and Domingos, 2012).

The metrics for evaluating classifiers are also influenced by the goals of the research and the questions that the researchers are interested in answering. Accuracy in matching the gold standard may seem to be a natural evaluation, but, in fact, this may be inappropriate given the ultimate use of the system. For example, should a researcher care about whether a classifier can accurately assess whether a doctor has entered an exam room or not? Such a classifier may identify such an event correctly over 99 % of the time, but if the goal is to figure out when a doctor doesn't see a patient, then false positives may be a more important metric. Similarly, it may not be important to detect the exact *moment* when you, for example, walk outside, get angry, speak to your child or eat something, provided that you *do* detect each of those events within some reasonable timeframe; an excellent discussion of evaluation metrics was written by Ward, Lukowicz, and Gellersen (2011). Finally, with regard to metrics, it is important to consider the statistical significance of the findings in addition to the accuracy of the predictions. These significance values help researchers to understand whether an algorithmic improvement on a small data set is as important as a small improvement on a big data set, and this is an often-overlooked component of classifier-based data analysis (see also Demšar, 2006).

What to Report in a Study Using Sensors

Although collecting streams of data from participants can be a powerful technique for observing existing behaviors (or in situ adaptations to novel technologies), incorporating sensor data into an experimental protocol requires careful planning. Additionally, reporting study findings based on streams of sensor data necessitates careful quantitative analysis that takes into account noisy sensor sources, ambiguities in sensed data, or technical problems with the sensors, which can lead to missing, corrupt, or misleading data. A well-described study will report the experiment

in sufficient detail that another researcher would be able to reproduce the study and achieve the same results. This would include describing:

- Hardware: What type and quantity of hardware was utilized? What mode was the hardware placed in? In what ways were the hardware components configured? What mode were they utilized in?
- Experimental setup: Where were the sensors placed? Why were those locations chosen? Who installed the sensors? What did the environment look like with the hardware installed? Did the setup change? At what point? How did this impact the participants? How did the experiment end? What was the final disposition of the hardware?
- Participant knowledge: Exactly what did the participants know about the hardware and data collection? How were they told? Were they asked to do anything different from—or in addition to—their normal activities during the study? Did participants' interactions with the study hardware change during the course of the study? How were participants compensated?
- Experimental execution: During the course of the study, what transpired with respect to the experimental setup? Were they outages? How much and what type of maintenance was conducted?
- Software infrastructure: How was the data collected? Was there specialized software used in collection? What configuration was the software put into? How was data transmitted, protected, and integrity assured?
- Analysis: What analysis was conducted on the data? At what point in the running of the experiment was the analysis conducted? What was an example of the data that was collected? Were there any changes made during the course of the analysis? Was particular software used to conduct analysis? What were the parameters of the algorithms used? How accurate was the analysis based on the gold-standard data? What other metrics were evaluated to justify the approach used? Was the analysis statistically significant?
- Why were these choices, and not others, made?

Personal Story About How the Authors Got into Using this Method, What Attracted Them to It, What Set of Methods They Use with It

In a study of the effect that e-mail has on both individual multitasking and information flow within small groups in an information work site (Mark et al., 2012), Voida and colleagues used a combination of:

- semi-structured interviews;
- surveys;
- in-person "shadowing";
- logging of electronic activities (i.e., desktop window switches); and
- environmental, social, and physiological sensors

to collect a large corpus of data about the pace of work, pace of interpersonal communication, and communication channels used in the workplace. We compared behavior under typical circumstances to when targeted individuals within the organization were asked to shut off their e-mail for 1 week. (During the e-mail "holidays" that we instituted, our participants' managers and coworkers were told about our study, and informed that the participant was in the office and working but could not be reached via e-mail.)

Collecting streams of sensor data was particularly important in the course of this research for three reasons. First, we wanted to collect as many different kinds of data about these individuals' work practices during the study as possible. We were able to examine many different facets of work and to observe the variations in the ways that these work practices were carried out by comparing the differences among repeated measurements using the same set of sensors. Second, one of the more interesting—and less well-understood—measures that we could collect about workplace collaboration and multitasking was the level of stress experienced by our participants throughout the workday. To understand changes in stress levels, we used commercial heart rate monitors to collect readings about heart rate hundreds of times each day. Finally, the field site for this study happened to be located nearly 3,000 miles from our home institution. In order to collect the volume of data that we needed to draw conclusions about work practices in light of the per-day and per-participant variation in information work activities, we set up the sensors once and then used them to collect data continuously from each participant for multiple weeks. This approach substantially minimized the cost—both in terms of travel and time on-site—for running the study.

Based on the data streams that we collected, we were able to quantify the amount of time that our participants spent each day conducting various kinds of activities (based on a "virtual sensor" observing window changes on the desktop computer), examine the composition and strength of social ties in the workplace (based on face-to-face conversations sensed using sociometric badges), and measure changes in stress levels when e-mail was present and when it was not (based on readings provided by a commercial heart rate monitor). In order to make sense of these streams of data, we conducted manual statistical analyses of the aggregated event streams after data collection was complete. Generally speaking, we looked at daily averages of these events, using ANOVAs, t-tests, and parametric statistics to determine whether our e-mail availability intervention, individual differences, or other factors were more likely to explain any variation that we observed in event frequencies or durations during different phases of the study. Even with a relatively small group of fully instrumented participants ($n=13$), the environmental, social, and physiological sensors collected millions of data points over the course of the study.

Another key aspect of our study was that we used a combination of log data (window changes recorded on each participant's primary desktop computer) and in-person observations to help label the sensor data. These discrete (and less ambiguous) data points enabled us to triangulate events of interest and helped us to make sense of a very large, complex, and messy corpus of real-world sensor readings.

In ongoing work, we are investigating the degree to which the log and sensor data that we collected can be used to train a classifier to predict the multitasking observations that we were able to collect from in-person observation—the traditional "gold standard" for this kind of research. The outcomes of this research will help to clarify which data streams are most useful for understanding information work as well as the circumstances in which sensors can (and cannot) serve as a substitute for in-person, ethnographically inspired fieldwork.

In another study, Patel and colleagues wanted to conduct an in-depth, empirical investigation of the proximity of the mobile phone to its owner over several weeks of continual observation (Patel, Kientz, Hayes, Bhat, & Abowd, 2006). The overall aims of this study were to determine if a mobile phone is a suitable proxy for its owner (an assumption that had never been empirically studied in the literature), to understand the reasons behind separation between a user and his or her mobile phone, and to offer guidelines for building applications. This study relied on a mixed-methods approach that required collecting data from some sort of proximity sensing technology. The sensors used were small, custom-built, battery-powered Bluetooth tags that the user would wear throughout the day. These tags would transmit a beacon signal continuously and the phone would record its distance to each tag based on the amount of time that it took the radio signal to propagate. This approach allowed us to continuously record the user's distance from his or her phone and to gather quantitative data not otherwise possible with other investigational means. Because the experience sampling method (ESM) or self-report would have created artificial changes to the user's behavior with their phone (i.e., picking it up to answer when the very thing that we were measuring was proximity to the phone), automatic sensing was a necessity. Additionally, the quantitative data allowed us to explore whether it was possible to apply machine-learning techniques to predict proximity. In the end, since the user's behavior was modified only by wearing a Bluetooth tag, and because no qualitative evidence to the contrary emerged in follow-up interviews, it was reasonable to argue that there was little modification to the user's natural behavior with respect to the proximity of the phone during the investigation. The resulting quantitative proximity traces proved valuable during the mixed-method interview process and the final analysis.

Because the logging application resided on the user's phone, the researchers took care to design it so that it would not impact the user's normal phone use. We also created a tool that would produce daily visualizations of people's proximity to their phone at 1-min intervals. Because the sensing was not perfect (e.g., it could be impacted by clothing, multiple people between the tag and the phone, and noisy RF environments), we used the visualization during our interviews with the participants as a guide for them to think through the day, an example of a retrospective analysis technique (see Looking Back: Retrospective Study Methods for HCI). They were allowed to bring their calendars or schedules to help them remember reasons for being separated from the phone. It turned out that the sensing did not have to be perfect. Even high-level activity information captured by the sensors was sufficient to help participants bridge the gaps in their memory, reconstruct their day, and

articulate how they were using the phone. In many cases, we found that it was not until participants actually saw their activity data that they recalled the details. Combining automatically collected sensor data and prompted interviews that were grounded in the sensor data served as a powerful tool for studying detailed behavior and the reasons behind it. It provided a balance between quantitative and qualitative results that might be difficult to achieve with a single method. A few years later another research group replicated this study on smartphone proximity with similar success (Dey et al., 2011).

In a final study, Patterson et al. used a variety of sensors available in a laptop in order to measure, and, subsequently, to predict, when users were in a particular, semantically defined place (Patterson et al., 2009). This study required users to install a piece of software on their laptops that periodically asked them to provide information about where they were to establish an annotated training set, in the style of the ESM (Larson & Csikszentmihalyi, 1983). Since place is often a key component of personal context, this information was used to set custom Instant Messaging status messages, populated with the user's choice of place names. The focus of the work was on collecting accurate place data information along with sensor data so that, eventually, well-informed context-aware services could be developed. While the software likely impacted the participants' use of their laptop, it was unlikely to have affected their mobility, which was a key focus of the study. This study required a lot of engineering work on the software that was installed on participants' computers in order to support a wide variety of software and hardware configurations.

Conclusion and Further Reading

Sensor data streams can provide large quantities of fine-grained data for observational studies and can illuminate behaviors or phenomena that would otherwise be impossible to study at a comparable level of detail—if at all. Preparing a research study based on this technique requires careful planning to ensure that the correct hardware is used, configured, and managed. Additionally, some degree of technical sophistication is required to store, process, and analyze the data at scale and with appropriate methodological and statistical rigor. Finally, participants need to be appropriately informed about the scope and nature of the data collection, since sensors can be difficult to see (or be forgotten over the course of a long study) and since the data collected by sensors can introduce privacy-related risks. When these issues are managed thoughtfully, however, sensor data streams can form the basis for rigorous and replicable research, and they can be of great value when used as a complement to other types of data-collection methodologies.

For further reading about these topics, consider looking into the following publications, which are also emphasized with a bold typeface in the chapter reference listing. Although we have found these articles to be useful for the reasons enumerated below, they are also informative in their motivations for using sensor data

streams, their methodological and analytical approaches, and the ways in which they use this research technique to draw conclusions for the research community:

- For more information about machine learning, classification and event-detection methodology, refer to Langley (2000) and Domingos (2012). For good examples of this technique being applied to answer specific research questions, see Patterson et al. (2005), Liao et al. (2007), and Horvitz et al. (2004).
- For more information on data processing and filtering, see Krumm (2010).
- For more information on managing participant privacy, see Langheinrich (2010) and Klasnja et al. (2009).
- A good example of egocentric research can be found in Fogarty et al. (2005).
- A good example of group-centric research is Choudhury and Pentland's paper on the *sociometer* (2003).
- A good example of infrastructure-centric research appears in Cohn, Stuntebeck et al. (2010).
- A good example of sensor triangulation can be found in Mark et al. (2012).

Exercises

1. What gold standards of behavior could you use to compare your interpretation of sensor streams against? Do you have to have a gold standard?
2. Compare the kinds of behaviors you can sense with a wearable sensor vs. a fixed sensor (e.g., on a wall)? What situations are appropriate for each?

References

Abowd, G. D., Atkeson, C. G., Hong, J., Long, S., Kooper, R., & Pinkerton, M. (1997, October). CyberGuide: A mobile context-aware tour guide. *Wireless Networks, 3*(5), 421–433.

Bailey, B. P., & Iqbal, S. T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction, 14*(4), 1–28.

Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior, 22*(4), 685–708.

Bailey, B. P., Konstan, J. A., & Carlis, J. V. (2001). The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In *Proceedings of the IFIP TC.13 international conference on human-computer interaction (INTERACT '01)* (pp. 593–601). Amsterdam: Ios Press.

Baldauf, M., Dustdar, S., & Rosenberg, F. (2007). A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing, 2*(4), 263–277.

Bao, L., & Intille, S. (2004). Activity recognition from user-annotated acceleration data. In *Proceedings of the second international conference on pervasive computing (PERVASIVE 2004)* (pp. 1–17). Berlin: Springer.

Begole, J., & Tang, J. C. (2007). Incorporating human and machine interpretation of unavailability and rhythm awareness into the design of collaborative applications. *Human–Computer Interaction, 22*(1), 7–45.

Begole, J., Tang, J. C., Smith, R. B., & Yankelovich, N. (2002). Work rhythms: Analyzing visualizations of awareness histories of distributed groups. In *Proceedings of the 2002 ACM conference on computer supported cooperative work (CSCW '02)* (pp. 334–343). New Orleans, LA: ACM Press.

Bolchini, C., Curino, C., Quintarelli, E., Schreiber, F. A., & Tanca, L. (2007). A data-oriented survey of context models. *SIGMOD Record, 36*(4), 19–26.

Brdiczka, O., Su, N. M., & Begole, J. B. (2010). Temporal task footprinting: Identifying routine tasks by their temporal patterns. In *Proceedings of the 14th international conference on intelligent user interfaces (IUI '10)* (pp. 281–284). New York, NY: ACM Press.

Brumitt, B., Meyers, B., Krumm, J., Kern, A., & Shafer, S. (2000). EasyLiving: Technologies for intelligent environments. In *Proceedings of the second international symposium on handheld and ubiquitous computing (HUC '00)* (pp. 97–119). Berlin: Springer.

Choudhury, T., Consolvo, S., Harrison, B., Hightower, J., LaMarca, A., LeGrand, L., et al. (2008). The mobile sensing platform: An embedded activity recognition system. *IEEE Pervasive Computing, 7*(2), 32–41.

Choudhury, T., & Pentland, A. (2003). Sensing and modeling human networks using the sociometer. In *Proceedings of the seventh IEEE international symposium on wearable computers (ISWC '03)* (pp. 216–222). Los Alamitos, CA: IEEE Computer Society.

Cohn, G., Gupta, S., Froehlich, J., Larson, E., & Patel, S. N. (2010). GasSense: Appliance-level, single-point sensing of gas activity in the home. In *Proceedings of 8th international conference on pervasive computing (PERVASIVE 2010)* (pp. 265–282). Berlin: Springer.

Cohn, G., Stuntebeck, E., Pandey, J., Otis, B., Abowd, G. D., & Patel, S. N. (2010). SNUPI: Sensor nodes utilizing powerline infrastructure. In *Proceedings of the 12th international conference on ubiquitous computing (UbiComp 2010)* (pp. 159–168). New York, NY: ACM Press.

Consolvo, S., Klasnja, P., McDonald, D. W., Avrahami, D., Froehlich, J., LeGrand, L., et al. (2008). Flowers or a robot army?: Encouraging awareness & activity with personal, mobile displays. In *Proceedings of the 10th international conference on ubiquitous computing (UbiComp '08)* (pp. 54–63). New York, NY: ACM Press.

Consolvo, S., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., et al. (2008). Activity sensing in the wild: A field trial of ubifit garden. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '08)* (pp. 1797–1806). New York, NY: ACM Press.

Consolvo, S., & Walker, M. (2003). Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing, 2*(2), 24–31.

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies: Why and how. In *Proceedings of the 1st international conference on intelligent user interfaces (IUI '93)* (pp. 193–200). New York, NY: ACM Press.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research, 7*, 1–30.

Dey, A. K., & Abowd, G. D. (2001). A conceptual framework and a toolkit for supporting rapid prototyping of context-aware applications. *Human-Computer Interaction, 16*(2–4), 7–166.

Dey, A., Wac, K., Ferreira, D., Tassini, K., Hong, J., & Ramos, J. (2011). Getting closer: An empirical investigation of the proximity of user to their smart phones. In *Proceedings of the 13th international conference on ubiquitous computing (UbiComp '11)* (pp. 163–172). New York, NY: ACM Press.

Domingos, P. (2012, October). A few useful things to know about machine learning. *Communications of the ACM*, *55*(10), 78–87.

Fan, M., Gravem, D., Cooper, D. M., & Patterson, D. J. (2012). Augmenting gesture recognition with erlang-cox models to identify neurological disorders in premature babies. In *Proceedings*

of the 2012 ACM conference on ubiquitous computing (UbiComp '12) (pp. 411–420). New York, NY: ACM Press.

Fogarty, J., Hudson, S. E., Atkeson, C. G., Avrahami, D., Forlizzi, J., Kiesler, S., et al. (2005). Predicting human interruptibility with sensors. *ACM Transactions on Computer-Human Interaction, 12*(1), 119–146.

Fox, D., Hightower, J., Kautz, H., Liao, L., & Patterson, D. J. (2003). Bayesian techniques for location estimation. In M. Hazas, J. Scott, & J. Krumm (Eds.), Research paper presented at the *2003 Workshop on Location-Aware Computing*, held in conjunction with the *Fifth International Conference on Ubiquitous Computing (UbiComp 2003)*, Seattle, WA, USA, October 12, 2003.

Froehlich, J. E., Larson, E., Campbell, T., Haggerty, C., Fogarty, J., & Patel, S. N. (2009). HydroSense: Infrastructure-mediated single-point sensing of whole-home water activity. In *Proceedings of the 11th international conference on ubiquitous computing (Ubicomp 2009)* (pp. 235–244). New York, NY: ACM Press.

Froehlich, J., Larson, E., Saba, E., Campbell, T., Atlas, L., Fogarty, J., et al. (2011). A longitudinal study of pressure sensing to infer real-world water usage events in the home. In *Proceedings of the 9th international conference on pervasive computing (PERVASIVE 2011)* (pp. 50–69). Berlin: Springer.

Greenberg, S., & Fitchett, C. (2001). Phidgets: Each development of physical interfaces through physical widgets. In *Proceedings of the 14th annual ACM symposium on user interface software and technology (UIST 2001)* (pp. 209–218). New York, NY: ACM Press.

Gupta, S., Reynolds, M. S., & Patel, S. N. (2010). ElectriSense: Single-point sensing using EMI for electrical event detection and classification in the home. In *Proceedings of the 12th international conference on ubiquitous computing (UbiComp 2010)* (pp. 139–148). New York, NY: ACM Press.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations, 11*(1), 10–18.

Hightower, J., (2003). From position to place. Research paper presented at the 2003 Workshop on Location-Aware Computing, held in conjunction with the *Fifth International Conference on Ubiquitous Computing (UbiComp 2003)*, Seattle, WA, USA, October 12, 2003.

Hightower, J. & Borriello, G. (2001). *A survey and taxonomy of location systems for ubiquitous computing*. Technical Report UW-CSE-01-08-03, Intel Research Seattle and the University of Washington, Seattle, WA.

Hightower, J. & Borriello, G. (2001, August). Location systems for ubiquitous computing. *IEEE Computer, 34*(8), 57–66.

Hong, J.-H., Suh, E., & Kim, S.-J. (2009). Context-aware systems: A literature review and classification. *Expert Systems with Applications, 36*(4), 8509–8522.

Hornecker, E., & Nicol, E. (2012). What do lab-based user studies tell us about in-the-wild behavior? Insights from a study of museum interactives. In *Proceedings of the ACM conference on designing interactive systems (DIS '12)* (pp. 358–367). New York, NY: ACM Press.

Horvitz, E., Koch, P., & Apacible, J. (2004). BusyBody: Creating and fielding personalized models of the cost of interruption. In *Proceedings of the ACM conference on computer supported cooperative work (CSCW '04)* (pp. 507–510). New York, NY: ACM Press.

Hutchings, D. R., Smith, G., Meyers, B., Czerwinski, M., & Robertson, G. (2004). Display space usage and window management operation comparisons between single monitor and multiple monitor users. In *Proceedings of the working conference on advanced visual interfaces (AVI '04)* (pp. 32–39). New York, NY: ACM Press.

Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., et al. (2003). Technology probes: Inspiring design for and with families. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '03)* (pp. 17–24). New York, NY: ACM Press.

Intille, S. S. (2002). Designing a home of the future. *IEEE Pervasive Computing, 1*(2), 76–82.

Intille, S., Larson, K., Tapia, E., Beaudin, J., Kaushik, P., Nawyn, J., et al. (2006). Using a live-in laboratory for ubiquitous computing research. In *Proceedings of the 4th international conference on pervasive computing (PERVASIVE 2006)* (pp. 349–365). Berlin: Springer.

Intille, S. S., Rondoni, J., Kukla, C., Ancona, I., & Bao, L. (2003). A context-aware experience sampling tool. In *Extended abstracts of the SIGCHI conference on human factors in computing systems (CHI '03)* (pp. 972–973). New York, NY: ACM Press.

Isaacman, S., Becker, R., Cáceres, R., Martonosi, M., Rowland, J., Varshavsky, A., et al. (2012). Human mobility modeling at metropolitan scales. In *Proceedings of the 10th international conference on mobile systems, applications, and services (MobiSys '12)* (pp. 239–252). New York, NY: ACM Press.

Kaptelinin, V. (2003). UMEA: Translating interaction histories into project contexts. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '03)* (pp. 353–360). New York, NY: ACM Press.

Kidd, C., Orr, R., Abowd, G., Atkeson, C., Essa, I., MacIntyre, B., et al. (1999). The aware home: A living laboratory for ubiquitous computing research. In *Proceedings of the second international workshop on cooperative buildings (CoBuild '99)* (pp. 191–198). Berlin: Springer.

Kientz, J. A., Patel, S. N., Jones, B., Price, E., Mynatt, E. D., & Abowd, G. D. (2008). The Georgia Tech aware home. In *Extended abstracts of the SIGCHI conference on human factors in computing systems (CHI 2008)* (pp. 3675–3680). New York, NY: ACM Press.

Klasnja, P., Consolvo, S., Choudhury, T., Beckwith, R., & Hightower, J. (2009). Exploring privacy concerns about personal sensing. In *Proceedings of the 7th international conference on pervasive computing (PERVASIVE 2009)* (pp. 176–183). Berlin: Springer.

Krumm, J. (2010). Processing sequential sensor data. In J. Krumm (Ed.), *Ubiquitous computing fundamentals* (pp. 286–319). Boca Raton, FL: CRC Press.

LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, J., et al. (2005). Place Lab: Device positioning using radio beacons in the wild. In *Proceedings of the third international conference on pervasive computing (PERVASIVE 2005)* (pp. 301–306). Berlin: Springer.

Langheinrich, M. (2010). Privacy in ubiquitous. In J. Krumm (Ed.), *Ubiquitous computing fundamentals* (pp. 286–319). Boca Raton, FL: CRC Press.

Langley, P. (2000). Crafting papers on machine learning. In *Proceedings of the seventeenth international conference on machine learning (ML '00)* (pp. 1207–1211). Stanford, CA: Morgan Kaufmann.

Larson, R., & Csikszentmihalyi, M. (1983). The experience sampling method. *New Directions for Methodology of Social and Behavioral Science, 15*, 41–56.

Liao, L., Patterson, D. J., Fox, D., & Kautz, H. (2006, December). Building personal maps from GPS data. *Annals of the New York Academy of Sciences, 1093*, 249–265.

Liao, L., Patterson, D. J., Fox, D., & Kautz, H. (2007, January). Learning and inferring transportation routines. *Artificial Intelligence*, *171*, 311–331.

MacIntyre, B., Mynatt, E. D., Voida, S., Hansen, K. M., Tullio, J., & Corso, G. M. (2001). Support for multitasking and background awareness using interactive peripheral displays. In *Proceedings of the 14th annual ACM symposium on user interface software and technology (UIST '01)* (pp. 41–50). New York, NY: ACM Press.

Mark, G. J., Voida, S., & Cardello, A. V. (2012). "A pace not dictated by electrons": An empirical study of work without email. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI 2012)* (pp. 555–564). New York, NY: ACM Press.

Marmasse, N., & Schmandt, C. (2000). Location-aware information delivery with ComMotion. In *Proceedings of the 2nd international symposium on handheld and ubiquitous computing (HUC '00)* (pp. 157–171). London: Springer.

Miluzzo, E., Lane, N. D., Fodor, K., Peterson, R., Lu, H., Musolesi, M., et al. (2008). Sensing meets mobile social networks: The design, implementation and evaluation of the CenseMe application. In *Proceedings of the 6th ACM conference on embedded network sensor systems (SenSys '08)* (pp. 337–350). New York, NY: ACM Press.

Monibi, M., & Patterson, D. (2009). Getting places: Collaborative predictions from status. In *Proceedings of the European conference on ambient intelligence (AmI 2009)* (pp. 60–65). Berlin: Springer.

Mynatt, E. D., Rowan, J., Craighill, S., & Jacobs, A. (2001). Digital family portraits: Supporting peace of mind for extended family members. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI 2001)* (pp. 333–340). New York, NY: ACM Press.

Nair, R., Voida, S., & Mynatt, E. D. (2005). Frequency-based detection of task switches. In *Proceedings of the 19th annual conference of the British HCI group (HCI 2005)* (pp. 94–99). Berlin: Springer.

Olguín, D. O., & Pentland, A. (2008). Social sensors for automatic data collection. In *Proceedings of the 14th annual Americas conference on information systems (AMCIS 2008)* (p. Paper 171). Atlanta, GA: Association for Information Systems.

Olguín, D. O., Waber, B. N., Kim, T., Mohan, A., Ara, K., & Pentland, A. (2009). Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Trans. On Systems, Man, and Cybernetics – Part B: Cybernetics, 39*(1), 12 pages.

Orr, R. J., & Abowd, G. D. (1999). The smart floor: A mechanism for natural user identification and tracking. In *Extended abstracts of the SIGCHI conference on human factors in computing systems (CHI 1999)* (pp. 275–276). New York, NY: ACM Press.

Patel, S. N., Kientz, J. A., Hayes, G. R., Bhat, S., & Abowd, G. D. (2006). Farther than you may think: An empirical investigation of the proximity of users to their mobile phones. In *Proceedings of the 8th international conference on ubiquitous computing (UbiComp 2006)* (pp. 123–140). Berlin: Springer.

Patel, S. N., Reynolds, M. S., & Abowd, G. D. (2008). Detecting human movement by differential air pressure sensing in HVAC system ductwork: An exploration in infrastructure mediated sensing. In *Proceedings of the 7th international conference on pervasive computing (PERVASIVE 2008)* (pp. 1–18). Berlin: Springer.

Patel, S. N., Robertson, T., Kientz, J. A., Reynolds, M. S., & Abowd, G. D. (2007). At the flick of a switch: Detecting and classifying unique electrical events on the residential power line. In *Proceedings of the 9th international conference on ubiquitous computing (UbiComp 2007)* (pp. 271–288). Berlin: Springer.

Patel, S. N., Truong, K. N., & Abowd, G. D. (2006). PowerLine positioning: A practical sub-room-level indoor location system for domestic use. In *Proceedings of the 8th international conference on ubiquitous computing (UbiComp 2006)* (pp. 441–458). Berlin: Springer.

Patterson, D. J. (2009). Global priors of place and activity tags. In *Proceedings of the AAAI spring symposium on human behavior modeling* (pp. 75–79). Palo Alto, CA: AAAI Press.

Patterson, D. J., Baker, C., Ding, X., Kaufman, S. J., Liu, K., & Zaldivar, A. (2008). Online everywhere: Evolving mobile instant messaging practices. In *Proceedings of the 10th international conference on ubiquitous computing (UbiComp '08)* (pp. 64–73). New York, NY: ACM Press.

Patterson, D. J., Ding, X., Kaufman, S. J., Liu, K., & Zaldivar, A. (2009). An ecosystem for learning and using sensor-driven IM status messages. *IEEE Pervasive Computing, 8*(4), 42–49.

Patterson, D. J., Fox, D., Kautz, H., & Philipose, M. (2005). Fine-grained activity recognition by aggregating abstract object usage. In *Proceedings of the ninth IEEE international symposium on wearable computers (ISWC '05)* (pp. 44–51). Los Alamitos, CA: IEEE Computer Society.

Patterson, D., Liao, L., Fox, D., & Kautz, H. (2003). Inferring high-level behavior from low-level sensors. In *Proceedings of the 5th international conference on ubiquitous computing (UbiComp 2003)* (pp. 73–89). Berlin: Springer.

Perkowitz, M., Philipose, M., Fishkin, K., & Patterson, D. J. (2004). Mining models of human activities from the web. In *Proceedings of the 13th international conference on World Wide Web (WWW '04)* (pp. 573–582). New York, NY: ACM Press.

Philipose, M., Fishkin, K. P., Perkowitz, M., Patterson, D. J., Fox, D., Kautz, H., et al. (2004). Inferring activities from interactions with objects. *IEEE Pervasive Computing, 3*(4), 50–57.

Poh, M.-Z., Swenson, N. C., & Picard, R. W. (2010). A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE Transactions on Biomedical Engineering, 57*(5), 1243–1252.

Rowan, J., & Mynatt, E. D. (2005). Digital family portrait field trial: Support for aging in place. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI 2005)* (pp. 521–530). New York, NY: ACM Press.

Salber, D., Dey, A. K., & Abowd, G. D. (1999). The context toolkit: Aiding the development of context-aware applications. In *Proceedings of the SIGCHI conference on human factors in computing systems (CHI 1999)* (pp. 434–441). New York, NY: ACM Press.

Sparacino, F. (2003). Sto(ry)chastics: A Bayesian network architecture for user modeling and computational storytelling for interactive spaces. In *Proceedings of the 5th international conference on ubiquitous computing (UbiComp 2003)* (pp. 54–72). Berlin: Springer.

Stone, A. A., Shiffman, S. S., & DeVries, M. W. (1999). Ecological momentary assessment. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 26–39). New York, NY: Russell Sage.

Stumpf, S., Bao, X., Dragunov, A., Dietterich, T. G., Herlocker, J., Johnsrude, K., Li, L., & Shen, J. (2005). Predicting user tasks: I know what you're doing! Research paper presented at the 20th national conference on artificial intelligence (AAAI-05) workshop on human comprehensible machine learning, Pittsburgh, PA, USA, July 9–13.

Tang, J., & Patterson, D. (2010). Twitter, sensors and UI: Robust context modeling for interruption management. In *Proceedings of the 18th international conference on user modeling, adaptation, and personalization (UMAP 2010)* (pp. 123–134). Berlin: Springer.

Tapia, E., Intille, S., & Larson, K. (2004). Activity recognition in the home using simple and ubiquitous sensors. In *Proceedings of the second international conference on pervasive computing (PERVASIVE 2004)* (pp. 158–175). Berlin: Springer.

Varshavsky, A., & Patel, S. (2009). Location in ubiquitous computing. In J. Krumm (Ed.), *Ubiquitous computing fundamentals* (pp. 286–319). Boca Raton, FL: CRC Press.

Villar, N., Scott, J., & Hodges, S. (2011). Prototyping with .NET Gadgeteer. In *Proceedings of the 5th international conference on tangible, embedded, and embodied interaction (TEI 2011)* (pp. 377–380). New York, NY: ACM Press.

Want, R., Hopper, A., Falcão, V., & Gibbons, J. (1992). The active badge location system. *ACM Transactions on Information Systems, 10*(1), 91–102.

Ward, J. A., Lukowicz, P., & Gellersen, H. W. (2011). Performance metrics for activity recognition. *ACM Transactions on Intelligent Systems and Technology, 2*(1), Article 6.

Weiser, M. (1991). The computer for the 21st century. *Scientific American, 265*(3), 94–104.

Westyn, T., Brashear, H., Atrash, A., & Starner, T. (2003). Georgia tech gesture toolkit: Supporting experiments in gesture recognition. In *Proceedings of the 5th international conference on multimodal interfaces (ICMI '03)* (pp. 85–92). New York, NY: ACM Press.

Wyatt, D., Choudhury, T., Bilmes, J., & Kitts, J. A. (2011). Inferring colocation and conversation networks from privacy-sensitive audio with implications for computational social science. *ACM Transactions on Intelligent Systems and Technology, 2*(1), Article 7.