Synthetic Sensors: Towards General-Purpose Sensing

Gierad Laput  Yang Zhang  Chris Harrison
Human-Computer Interaction Institute, Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213
{gierad.laput, yang.zhang, chris.harrison}@cs.cmu.edu

ABSTRACT
The promise of smart environments and the Internet of Things (IoT) relies on robust sensing of diverse environmental facets. Traditional approaches rely on direct or distributed sensing, most often by measuring one particular aspect of an environment with special-purpose sensors. In this work, we explore the notion of general-purpose sensing, wherein a single, highly capable sensor can indirectly monitor a large context, without direct instrumentation of objects. Further, through what we call Synthetic Sensors, we can virtualize raw sensor data into actionable feeds, whilst simultaneously mitigating immediate privacy issues. We use a series of structured, formative studies to inform the development of new sensor hardware and accompanying information architecture. We deployed our system across many months and environments, the results of which show the versatility, accuracy and potential of this approach.

Author Keywords
Internet-of-Things; IoT; Smart Home; Universal Sensor.

ACM Classification Keywords

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

INTRODUCTION
Smart, sensing environments have long been studied and sought after. Today, such efforts might fall under catchphrases like the “smart home” or the “internet of things”, but the goals have remained the same over decades—to apply sensing and computation to enhance the human experience, especially as it pertains to physical contexts (e.g., home, office, workshop) and the amenities contained within. Numerous approaches have been attempted and articulated, though none have reached widespread use to date.

One option is for users to upgrade their environments with newly released “smart” devices (e.g., light switches, kitchen appliances), many of which contain sensing functionality. However, this sensing is generally limited to the appliance itself (e.g., a smart light switch knows if it is on or off) or when it serves its core function (e.g., a thermostat sensing occupancy). Likewise, few smart devices are interoperable, thus forming silos of sensed data that thwarts a holistic experience. Instead of achieving a smart home, the best one can hope for—at least in the foreseeable future—are small islands of smartness. This approach also carries a significant upgrade cost, which so far has proven unpopular with consumers, who generally upgrade appliances piecemeal.

To sidestep this issue, we are now seeing aftermarket products (e.g., [39,41,56]) and research systems (e.g., [36,45,54]) that allow users to distribute sensors around their environments to capture a variety of events and states. For ex-
ample, Sen.se’s Mother product [52] allows users to attach "universal" sensor tags to objects, from which basic states can be discerned and tracked over time (e.g., a tag on a coffee machine can track how often coffee is made). This approach offers great flexibility, but at the cost of having to instrument every object of interest in an environment.

As we will discuss, a single room can have dozens of complex environmental facets worth sensing, ranging from "is the coffee brewed?" to "is the tap dripping?" A single home might have hundreds of such facets, and an office building could have thousands. The cost of hundreds of physical sensors is significant, not including the even greater cost of deployment and maintenance. Moreover, extensively instrumenting an environment in this fashion will almost certainly carry an aesthetic and social cost [3].

A lightweight, general-purpose sensing approach could overcome many of these issues. Ideally, a handful of “super” sensors could blanket an entire environment – one per room or less. To be minimally obtrusive, these sensors should be capable of sensing environmental facets indirectly (i.e., from afar) and be plug and play – forgoing batteries by using wall power, while still offering omniscience despite potential sub-optimal placement. Further, such a system should be able to answer questions of interest to users, abstracting raw sensor data (e.g., z-axis acceleration) into actionable feeds, encapsulating human semantics (e.g., a knock on the door), all while preserving occupant privacy.

In this paper, we describe the structured exploration process we employed to identify opportunities in this problem domain, ultimately leading to the creation of a novel sensing system and architecture that achieves most of the properties described above. First, we provide a comprehensive review of sensors, both academic and commercial, that claim some level of generality in their sensing. Similarly, we conducted a probe into what environmental facets users care to know, and at what level of fidelity has acceptable privacy trade-offs. We then merged what we learned from this two-pronged effort to create a novel sensor tag (Figures 3 and 4).

We deployed our sensor tags across many months and environments to collect data and investigate ways to achieve our desired versatility, accuracy and privacy preservation. This directly informed the development of a novel, general-purpose, sensing architecture that denatures and virtualizes raw sensor data, and through a machine-learning pipeline, yields what we call Synthetic Sensors. Like conventional sensors, these can be used to power interactive applications and responsive environments. We conclude with a formal evaluation, deploying our entire sensing pipeline, followed by a digest of significant findings and implications.

**RELATED APPROACHES**

A full review of the literature on environmental sensing is beyond the scope of this paper. However, to help illustrate this application landscape, we created a Sensor Utility Taxonomy shown in Figure 1. Along the y-axis is the number of distinct sensed facets (e.g., states, events), while the x-axis is the number of sensors needed to achieve this output.

**Special-Purpose Sensors**

The most intuitive and prevalent form of sensing is to use a single sensor to monitor a single facet of an environment. For example, in UpStream [32] and WaterBot [2], a microphone is affixed to a faucet so that water consumption can be inferred (which in turn is used to power behavior-changing feedback). Similarly, efficient management of HVAC has been demonstrated through room-level temperature [30] and occupancy sensors [50]. Special-purpose sensors tend to be robust for well-defined, low-dimensional sensing problems, such as occupancy sensing and automatically opening doors. However, this relationship is inherently a one-sensor to one-sensed-facet relationship (i.e., one-to-one; Figure 1, bottom left quadrant). For example, an occupancy sensor can only detect occupancy, and a door ajar sensor can only detect when a door is open. There is no notion of generality; each desired facet is monitored by a specific and independent sensor.

**Distributed Sensing Systems**

It is also possible to deploy many sensors in an environment, which can be networked together, forming a distributed sensing system [26]. This approach can be used to enlarge the sensed area (e.g., occupancy sensing across an entire warehouse) or increase sensing fidelity through complementary readings (e.g., seismic events [5,58]). The distributed sensors can be homogeneous [14] (e.g., an array of identical infrared occupancy sensors) or heterogeneous (i.e., a mix of sensor types) [54,55,61]. Also, the array can sense one facet (e.g., fire detection) or many (e.g., appliance use).

A home security system is a canonical example of a heterogeneous distributed system, where door sensors, window sensors, noise sensors, occupancy sensors and even cameras work together for a singular classification: is there an intruder in the home? This is a many-to-one scheme, and thus occupies the bottom right of our Figure 1 taxonomy. Conversely, for example, Tapia et al. [54] use a homogenous array of 77 magnetic sensors to detect object interactions throughout an entire house, and thus is a many-to-many scheme (upper right in Figure 1). Thus, distributed systems occupy the entire right side of our Figure 1 taxonomy.

A distributed sensing system, as one might expect, is highly dependent on the quality of its sensor distribution. Achieving the necessary sensor saturation often implies a sizable deployment, perhaps dozens of sensors for even a small context, like an office. This can be costly; with sensors often costing $30 or more, even small deployments can become unpalatable for consumers. Moreover, as the number of sensors grow, there is a danger of becoming invasive in sensitive contexts such as the home [8,16,28,54].

**Infrastructure-Mediated Sensing**

To reduce deployment cost and social intrusiveness, researchers have investigated the installation of sensors at
strategic infrastructure probe points. For example, work by Abott [1], Hart [22,23] and Gupta [20] used sensors coupled to a building’s power lines to detect “events” caused by electrical appliances. Since home electrical lines are shared, a single sensor can observe activities across an entire house.

This infrastructure-mediated sensing approach has also been applied to e.g., HVAC [42], plumbing [16,17], natural gas lines [10] and electric lighting [19]. In all of these cases, a sensor was installed at a single probe point, enabling application scenarios that would otherwise require more costly distributed instrumentation of an environment. Although considerably more general purpose than the other approaches we have discussed, this approach is still constrained by the class of infrastructure it is coupled to. For example, a plumbing-attached sensor can detect sink, shower and toilet use, but not microwave use. Thus, we denote it as a one-to-few technique in our taxonomy (left-middle).

**Direct vs. Indirect Sensing**

Many of the aforementioned systems utilize direct sensing, that is, a sensor that physically couples to an object or infrastructure of interest. For example, most window sensors need to be physically attached to a window. This approach is popular as it generally yields excellent signal quality. However, powering such sensors can be problematic, as most objects do not have power outlets. Instead, such systems rely on batteries, which must be periodically recharged [10,16,17,36,54,61]. Other systems avoid this by requiring access to a power outlet [19,20], though this limits possible sensor locations or requires cords be run across the environment—neither of which is desirable.

Fortunately, it is also possible to sense state and events indirectly, without having to physically couple to objects. For example, work by Kim and colleagues [29] explored sensing of appliance usage with a sensor installed nearby. When an appliance is in different modes of operation (e.g., refrigerator compressor running, interior lights on/off), it emits characteristic electromagnetic noise that can be captured and recognized. Similarly, Ward and colleagues [59] were able to recognize tool use in a workshop through acoustic sensing. Indeed, many sensors are specifically designed for indirect sensing, including non-contact thermometers, rangefinders, and motion sensors.

Overall, indirect sensing allows for greater flexibility in placement, often allowing sensors to be better integrated into the environment or even hidden, and thus less aesthetically and socially obtrusive. Ideally, it is possible to relocate to a nearby wall power outlet, eliminating the need for batteries. However, this typically comes at the cost of some sensing fidelity – the further you move away from an object or area of interest, the harder it becomes to sense and segment events. Moreover, some sensors require line-of-sight, which can make some sensor placements untenable.

**General-Purpose Sensing**

Increasingly, sensor “boards” are being populated with a wide variety of underlying sensors that affords flexible use (Table 1). Such boards might be considered general purpose, in that they can be attached to a variety of objects, and without modification, sense many facets. However, this is still ultimately a one-sensor to one-object mapping (e.g., Sen.se’s Mother [52]), and thus is more inline with the tenants of a distributed sensing system (many-to-many).

The ideal sensing approach occupies the top-left of our taxonomy, wherein one sensor can enable many sensed facets, and more specifically, beyond any one single instrumented object. This one-to-many property is challenging, as it must be inherently indirect to achieve this breadth. The ultimate embodiment of this approach would be a single, omniscient sensor capable of digitizing an entire building.

Computer vision has come closest to achieving this goal. Cameras offer rich, indirect data, which can be processed through e.g., machine learning to yield sensor-like feeds. There is a large body of work in video-based sensing (see e.g., [15,40,53]). Achieving human-level abstractions and accuracy is a persistent challenge, leading to the creation of mixed CV- and crowd-powered systems (e.g., [6,18,35]).

Most closely related to our current work is Zensors [34], which explicitly use a sensor metaphor (as opposed to a Q/A metaphor [6]). Using a commodity camera, a wide variety of environmental facets could be digitized, such as

---

**Table 1.** An inventory of research and commercial sensors offering varying degrees of general-purpose sensing. Our prototype sensor is the union of these capabilities, with the notable absence of a camera.
“how many dishes are in the sink?” To achieve this level of general-purposeness, Zensors initially uses crowd answers, which are simultaneously used as labels to bootstrap machine learning classifiers, allowing for a future handoff.

While these CV-based sensing approaches are powerful, cameras have been widely studied and recognized for their high level of privacy invasion and social intrusiveness [3, 7, 8, 28], and thus carry a heavy deployment stigma. To date, this has hindered their use in many environments ripe for sensing, such as homes, schools, care facilities, industrial settings and work environments. In this work, we show that we can achieve much of the same sensing versatility and accuracy without the use of a camera.

EXPLORATORY STUDIES
As a first step in our exploration of general purpose sensing, we conducted two focused probes. This grounded basic assumptions and informed the design of our system.

Survey of Sensor Boards
There is an emerging class of small, screen-less devices equipped with a array of sensors and wireless connectivity, often described as “sensor boards” or “tags”. For example, the Texas Instruments SimpleLink SensorTag packs five sensors into a matchbook-sized, coin-battery-powered package [56]. These devices are intended to facilitate “quick and easy prototyping of IoT experiences.” We performed an extensive survey of these boards, drawn from commercial and academic systems [4, 9, 14, 16, 17, 31, 39, 47, 52, 56], allowing us to build an inventory of sensing capabilities. The high level results of our search are offered in Table 1.

Facet & Privacy Elicitation Study
In our second probe, we sought to better understand the perceived utility of a “perfect” and omnipotent, general-purpose sensor. For this, we conducted an elicitation study (10 interaction design Masters students, 4 female, mean age 24.4, two hours, paid $20) that allowed us to gather facets of interest about six local environments (a common area, kitchen, workshop, classroom, office and bathroom). In total, following a group affinity diagramming exercise, 107 unique facets were identified. We also used this opportunity to informally inquire about the perceived privacy implications of such sensed facets. During discussion, participants unanimously desired that “sensor data be stored in a way that cannot identify individuals.” We asked participants to rank facet privacy on a scale of 0 (no privacy danger) to 5 (high privacy danger). Unsurprisingly, facets along “who” dimensions (mean 3.76, SD 1.34) ranked significantly higher (p<0.01) in their privacy invasiveness than those along “what” dimensions (mean 0.92, SD 1.02). Reinforcing our initial notion, and what many others have also found [3, 7], participants uniformly rejected the use of a camera.

CUSTOM SENSOR TAG
We set out to design a novel sensor tag (Figures 3 and 4), which integrates the union of the sensing capabilities across all of the devices in Table 1, minus a camera. Not only does this serve as an interesting vehicle for investigation (e.g., what sensors are most accurate and useful?), but also an extreme embodiment of board design using many low-level sensors – one that we hoped could approach the versatility of camera-based approaches, but without the stigma and privacy implications. We incorporated nine physical sensors capturing twelve distinct sensor dimensions (see Figure 3).

The heart of our sensor tag design is a Particle Photon STM32F205 microcontroller with a 120MHz CPU. We strategically placed sensors on the PCB to ensure optimal performance (e.g., ambient light sensor faces outwards), and we spatially separated analog and digital components to isolate unintended electrical noise from affecting the performance of neighboring components. For connectivity, we considered industry standards such as Ethernet, ZigBee, and Bluetooth, but ultimately chose WiFi for its ubiquity, ease-of-setup, range and high bandwidth.

Finally, our board uses a Type A USB 2.0 connector, which can be used for power (e.g., with a DC wall wart) or to deploy software. We intentionally designed our board so that it can be easily plugged-in to a power outlet (in line with our goals of being maintenance free). From this placement, we hope to be “omniscient” through clever signal processing and machine learning. For this reason, power consumption was not a design objective (for reference: approximately 120mA at 5V when fully streaming).
PILOT DEPLOYMENT & FINDINGS
At this stage of development, our sensor tags provided a raw stream of high fidelity data (e.g., an audio-quality microphone stream), which was logged to a secured server. Two questions were paramount: 1) was the captured sensor data sufficiently descriptive to enable general-purpose sensing? And 2) was the sensed data adequately preserving occupant privacy, especially identity?

To explore possible tradeoffs, we deployed five sensor tags across thirteen diverse environments we controlled for a collective duration of 6.5 months. During this period, we iteratively refined our sensor’s software, affording us the ability to test parameters and system features live and in real world settings. This led to several critical insights:

Immediate Featurization. We found there was little advantage in transmitting raw data from our sensor boards, and instead, all data could be featurized on-sensor. Not only does this reduce network overhead, but it also denatures the data, better protecting privacy while still preserving the essence of the signal (with appropriate tuning). In particular, we selected features (discussed later) that do not permit reconstruction of the original signal.

Sensing Fidelity. Our pilot deployments showed that our sensor tags were capable of capturing rich, nuanced data. For example in Figure 5, we can see not only the coarse event of “microwave in use”, but also its door being opened and closed, as well as the completion chime, revealing state. In general, we found that sensed signals could be broadly categorized into three temporal classes: sub-second, seconds-to-minutes, and minutes-to-hours scale. For example, a knock on a door lasts a fraction of a second, and so for a sensor to capture this (e.g., acoustically), it must operate at a sampling interval capable of digitizing sub-second events. On the other hand, other facets change slowly, such as room temperature and humidity.

In response, we tuned our raw sensor sampling rates over the course of deployment, collecting data at the speed needed to capture environmental events, but with no unnecessary fidelity. Specifically, we sample temperature, humidity, pressure, light color, light intensity, magnetometer, Wifi RSSI, GridEye and PIR motion sensors at 10Hz. All three axes of the accelerometer are sampled at 4 kHz, our microphone at 17 kHz, and our and EMI sensor at 500 kHz. Note that when accelerometers are sampled at high speed, they can detect minute oscillatory vibrations propagating [33] through structural elements in an environment (e.g., dry-walls, studs, joists), very much like a geophone.

Sensor Activation Groups. Our pilot deployments revealed that events tend to activate particular subsets of sensor channels. For instance, a “lights on” event will activate the light sensor, but not the temperature or vibration sensors. Similarly, a door knock might activate the microphone and x-axis of our accelerometer, but not our EMI sensor. We use “activate” to mean any statistical deviation from the environmental norm. As we will discuss, we can leverage these sensor activation groups to improve the system’s robustness to noise by only dispatching events to our classification engine if the appropriate set of sensors are activated.

SYNTHETIC SENSORS
Our exploratory studies revealed that while low-level sensor data can be high-fidelity, it often does not answer users’ true intent. For example, the average user does not care about a spectrogram of EMI emissions from their coffee maker – they want to know when their coffee is brewed. Therefore, a key to unlocking general-purpose sensing is to support the “virtualization” of low-level data into semantically relevant representations. We introduce a sensing abstraction pattern that enables versatile, user-centered, general-purpose sensing, called Synthetic Sensors.

In this framework, sensor data exposed to end-users is “virtualized” into higher-level constructs, ones that more faithfully translate to users’ mental models of their contexts and environments. This “top-down” approach shifts the burden away from users (e.g., “what can I do with accelerometer
data?”) and unto the sensing system itself (e.g., user demonstrates a running faucet while the system learns its vibrational signature). Such output better matches human semantics (e.g., “Is the laser cutter exhaust running?”) and end-user applications can use this knowledge to power rich, context-sensitive applications (e.g., “If exhaust is turned off, send warning about fumes”).

Overall Architecture
First, as already discussed, we detect events that manifest in an environment through low-level sensor data. For example, when a faucet is running, a nearby sensor tag can pick up vibrations induced by service pipes behind the wall, as well as characteristic acoustic features of running water. Next, a featurization layer converts raw sensor data into an abstract and compact representation. This happens in our embedded software, which means raw data never leaves the sensor tag. Finally, the “triggered” sensors form an activation group (e.g., X- and Z-axis of accelerometer, plus microphone), which becomes the input to our machine learning layer.

We support two machine learning modalities: manual training (e.g., via user demonstration and annotation using a custom interface we build; see Video Figure) or automatic learning (e.g., through unsupervised clustering methods). The output of the machine learning layer is a “synthetic sensor” that abstracts low-level data (e.g., vibration, light color, EMI sensors) into user-centered representations (e.g., coffee ready sensor). Finally, data from one or more synthetic sensors can be used to power end-user applications (e.g., estimating kitchen water usage, sending a text when the laundry dryer is done).

On-Board Featurization
Data from our high-sample-rate sensors are transformed into a spectral representation via a 256-sample sliding window FFT (10% overlapping), ten times per second. Note that phase information is discarded. Our raw 8x8 GridEye matrix is flattened into row and column means (16 features). For our other low-sample-rate sensors, we compute seven statistical features (min, max, range, mean, sum, standard deviation and centroid) on a rolling one-second buffer (at 10Hz). The featurized data for every sensor is concatenated and sent to a server as a single data frame, encrypted with 128-bit AES.

Event Trigger Detection
Once data is sent, we perform automatic event segmentation on the server side. To reduce the effects of environmental noise, the server uses an adaptive background model for each sensor channel (rolling mean and standard deviation). All incoming streams are compared against the background profile using a normalized Euclidean distance metric (similar to [38]). Sensor channels that exceed their individual thresholds are tagged as “activated”. We also apply hysteresis to avoid detection jitter. Thresholds were empirically obtained by running sensors for several days while tracking their longitudinal variances. All triggered sensors form an activation group.

Of note, classification of simultaneous events is possible, especially if the activation groups are mutually exclusive. For overlapping activation groups, cross talk between events is inevitable. Nonetheless, our evaluations suggest that many events contain discriminative signals even when using shared channels.

Server-Side Feature Computation
The set of activated sensors serves as useful metadata to describe an event. Likewise, we use activation groups to assemble an amalgamated feature vector (e.g., a boiling kettle event combines features extracted from the GridEye, accelerometer and microphone). Then, if any high-sample-rate sensors were included, we compute additional features on the server. Specifically, for vibrations, acoustics and EMI, we compute band ratios of 16-bin downsampled spectral representations (120 additional features), along with additional statistical features derived from the FFT (min, max, range, mean, sum, standard deviation, and centroid). For acoustic data, we also compute MFCCs [63]. Data from all other sensors are simply normalized. Finally, these features are fed to a machine learning model for classification.

Learning Modalities
In manual mode, users train the system by demonstrating an event of interest, à la “programming by demonstration” [11,24,25], supplying supervised labeled data (see Video Figure for an interactive demonstration). The feature sets, along with their associated labels are fed into a plurality-based ensemble classification model (implemented using the Weka Toolkit [21]), similar to the approach used by Ravi et al. [45]. We use base-level SVMs trained for each synthetic sensor, along with a global (multi-class) SVM trained on all sensors. This ensemble model promotes robustness against false positives, while supporting the ability to detect simultaneous events.

In automatic learning mode, the system attempts to extract environmental facits via unsupervised learning techniques. We use a two-stage clustering process. First, we reduce the dimensionality of our data set using a multi-layer perceptron configured as an AutoEncoder [27], with five non-overlapping sigmoid functions in the hidden layer. Because the output of the AutoEncoder is the same as the input values, the hidden layer will learn the best reduced representation of the feature set. Finally, this reduced feature set is used as input to an expectation maximization (EM) clustering algorithm. These were implemented using python scikit-learn [43], Weka [21] and Theano [37].

EVALUATION
We explored several key questions to validate the feasibility of synthetic sensors. Foremost, how versatile and generic is our approach across diverse environments and events? Are the signals captured from the environment stable and consistent over time? How robust is the system to environmental noise? And finally, once deployed, how accurate can synthetic sensors be?
Deployment

To answer these questions, we conducted a two-week, in situ deployment across a range of environmental contexts. Specifically, we returned to five of the six locations we explored in our Face & Privacy Elicitation Study: a kitchen (~140 sq. ft.), an office (~71 sq. ft.), a workshop (~446 sq. ft.), a common area (~156 sq. ft.) and a classroom (~1000 sq. ft.), spanning an entire building at our institution. In each room, a single sensor tag was plugged into a centrally located, available, electrical wall socket. Building occupants went about their daily routines uninterrupted. Each tag ran continuously for roughly 336 hours (for a cumulative period of roughly 1,700 hours). Featured data was streamed and stored to a secure local server for processing and analysis.

Versatility of General Purpose Sensing

We examined the list of environmental facet questions (i.e., synthetic sensors) that our earlier study participants elicited, which we pruned to facets that could be practically sensed. Specifically, we removed facets that required a camera (e.g., “what is written on the whiteboard?”), and likewise eliminated facets that did not manifest physical output that could be sensed (“where did I leave my keys?”). This pruned the original set from 107 facets to 59. From the remaining facets, we selected 38 to be our test synthetic sensors (Table 2).

Signal Fidelity and Sensing Accuracy

To understand the fidelity and discriminative power of the signals produced from our sensor tags, we conducted an accuracy evaluation, spanning multiple days and locations. In each test location, we demonstrated instances of each facet of interest (mean repeats = 6.0, max 8). For instance, in the workshop, we collected data for the “Laser Cutter Exhaust” synthetic sensor by turning on and off the exhaust several times. We collected data every day for a week, which was labeled offline using a custom tool. This yielded a total of ~150K labeled data instances spanning our 38 synthetic sensors located in five locations. The labeling process took approximately one hour per sensor for a day’s worth of data. This task was tedious, but critical, as it established a ground truth from which to assess accuracy. Note that we also captured “null instances” (i.e., no event) that were derived from captured background instances.

To evaluate accuracy, we started by training the classifier using data from day 1, and then testing the classifier using data collected on day 2. This effectively simulates, post hoc, what the accuracy would have been on day 2. We then repeat this process, using data from days 1 and 2 for training, and testing on day 3’s data. We continue this process up to day 7, which is trained on data from days 1 through 6. In this way, we can construct a learning curve, which reveals how accuracy would improve over time (Figure 6).

Across our 38 synthetic sensors in five locations, spanning all seven days, our system achieved an average sensing accuracy of 96.0% (SD=5.2%). Note that the accuracy on day 2 is already relatively high (91.5%, SD=11.3%). We also reiterate that a “day’s worth” of data, but simply the demonstrated instances for a day (i.e., a few minutes of demonstration data per sensor).

Sensing Stability Over Time

It is possible for environmental facets to change their physical manifestation over time (due to e.g., ambient air temperature, or shifts in physical position). Therefore, it is important to explore whether signals are sufficiently reliable over time to enable robust synthetic sensors. Thus, in addition to collecting data on days 1 through 7, we also collected and labeled one additional round of data on day 14. This weeklong separation (without intervening data collection days) is a useful, if basic test of signal stability.

More specifically, we trained our system on data from days 1 through 7, and tested on day 14’s data. Overall, across our 38 synthetic sensors in five locations, the system was 98.0% accurate (SD=2.1%), similar to day 7’s results, showing no

<table>
<thead>
<tr>
<th>ID Sensor</th>
<th>Accel</th>
<th>Mic</th>
<th>EM</th>
<th>Temp</th>
<th>Baro</th>
<th>Humid</th>
<th>Light</th>
<th>Colar</th>
<th>Mag</th>
<th>RSSI</th>
<th>IR</th>
<th>G Spray</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kettle On</td>
<td>9%</td>
<td>5%</td>
<td>4%</td>
<td>3%</td>
<td>2%</td>
<td>2%</td>
<td>4%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
<td>5%</td>
</tr>
<tr>
<td>Paper Towel Dispenser</td>
<td>64%</td>
<td>38%</td>
<td>30%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO Detector Alarming</td>
<td>73%</td>
<td>5%</td>
<td>2%</td>
<td>1%</td>
<td>5%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dishwasher Running</td>
<td>39%</td>
<td>3%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faucet Running</td>
<td>37%</td>
<td>5%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fridge Door Closed</td>
<td>77%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soap Dispenser</td>
<td>54%</td>
<td>4%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dishwasher Done Chime</td>
<td>74%</td>
<td>6%</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Door Closed</td>
<td>49%</td>
<td>4%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Projector Running</td>
<td>36%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person Speaking</td>
<td>20%</td>
<td>23%</td>
<td>2%</td>
<td>2%</td>
<td>1%</td>
<td>5%</td>
<td>1%</td>
<td>3%</td>
<td>3%</td>
<td>1%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room Lights Off</td>
<td>39%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>5%</td>
<td>1%</td>
<td>3%</td>
<td>3%</td>
<td>3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roughing On</td>
<td>15%</td>
<td>11%</td>
<td>2%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td>5%</td>
<td>1%</td>
<td>3%</td>
<td>3%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microwave Running</td>
<td>41%</td>
<td>20%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coffee Maker Brewing</td>
<td>32%</td>
<td>29%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>3%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microwave Keypad Press</td>
<td>37%</td>
<td>10%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microwave Door Opened</td>
<td>46%</td>
<td>3%</td>
<td>2%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microwave Door Closed</td>
<td>41%</td>
<td>3%</td>
<td>2%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toaster Done Bep</td>
<td>27%</td>
<td>56%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W. Fountain Chiller Active</td>
<td>23%</td>
<td>49%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoke Detector Alarming</td>
<td>27%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water Fountain Dispensing</td>
<td>42%</td>
<td>6%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microwave Oven Chime</td>
<td>58%</td>
<td>16%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone Ringing</td>
<td>60%</td>
<td>6%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knock on Door</td>
<td>56%</td>
<td>3%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Door Closed</td>
<td>50%</td>
<td>3%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person Talking</td>
<td>52%</td>
<td>3%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone Loud Speaker On</td>
<td>49%</td>
<td>4%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Old Printer Running</td>
<td>44%</td>
<td>6%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper Towel Dispenser</td>
<td>52%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hammering</td>
<td>53%</td>
<td>6%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wood Saw Running</td>
<td>45%</td>
<td>3%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dremel Running</td>
<td>79%</td>
<td>4%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shop VAC Running</td>
<td>47%</td>
<td>5%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dust GoGo Running</td>
<td>63%</td>
<td>3%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Power Drill Running</td>
<td>43%</td>
<td>5%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sanding Running</td>
<td>45%</td>
<td>3%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laser Cutter Running</td>
<td>58%</td>
<td>4%</td>
<td>1%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. List of synthetic sensors studied across our two-week deployment. Percentages are based on a sensor’s feature merit (i.e., normalized SVM weights).

![Figure 6. Learning curves for our synthetic sensor deployment, combined per test location.](Image)

![Kitchen](Image)  ![Classroom](Image)  ![Common Area](Image)  ![Office](Image)  ![Workshop](Image)
degradation in accuracy. Day 14’s results are also plotted in Figure 6, and we further provide the full confusion matrices for each room’s sensors in Figure 7. Note that a vast majority of the synthetic sensors perform well – near or at 100% accuracy – with just three sensors performing poorly (in the 60% accuracy range). Overall, we believe these results are encouraging and suggest that our sensors boards and synthetic sensors do indeed achieve their general-purpose aim.

**Noise Robustness**

Human environments are noisy, not just in the acoustic channel, but all sensor channels. A robust system must differentiate between true events and a much larger class of false triggers. In response, we conducted a brief noise robustness study that examined the behavior of synthetic sensors when exposed to deliberately noisy conditions.

We selected a high-traffic location (common area) and manually monitored the performance of our classifier (trained on data from days 1-7). An experimenter logged location activity *in vivo*, while simultaneously monitoring classification output. The range of activities observed was diverse, from "sneezing" and "clipping nails", to "people chatting" and a "FedEx delivery." The experimenters also injected their own events, including jumping jacks, whistling, clapping, and feet stomping.

The observation lasted for two hours, and within this period, the experimenter recorded 13 false positive triggers. Admittedly, a longer duration would have been preferable, but the labor involved to annotate a longer period was problematic at the time of the study. Regardless, we believe this result is useful and promising, though the false positive rate is higher than we hoped. It suggests future work is needed on mitigating false positives, perhaps by supplying more negative examples or employing more sophisticated ML techniques, like dropout training.

**Automatic Event Learning**

We performed a preliminary evaluation of our system’s ability to automatically extract and identify events of interest *without* user input (i.e., segmentation or labeling). As briefly discussed in the implementation section, we used a two-step process: multi-layer perceptron followed by an EM clustering algorithm. To evaluate the effectiveness of auto-generated clusters, we used labeled data from days 1 though 7 and performed a cluster membership evaluation.

We found mixed results, ranging from a high of 88.1% mean accuracy in the classroom location (five synthetic sensors, thus chance=20%) to a low of 30.0% in the workshop setting (chance=11%). In most locations, clusters were missing for some user-labeled facets, and often featured scores of unknown clusters for things the system had learned by itself. Much future work could be done in this area. More sophisticated clustering and information retrieval techniques could help [48,60], as could correlating sensor data with known events and activities (e.g., room calendar) to power a knowledge-driven inference approach [44,62].

**SENSOR TYPE & SAMPLE RATE IMPLICATIONS**

Most of the synthetic sensors we have described so far have been *sub-second scale* events, and thus most heavily rely on our board’s high-sample-rate sensors. This bias can be readily seen in Table 2, which provides a weighted breakdown of merit as calculated by SVM weights when all features are supplied to the classifier. Three sensors in particular stand out as most useful: microphone (17 kHz sample rate), accelerometer (4 kHz) and EMI (500 kHz) – the three highest sample rate sensors on our board.

Foremost, we stress that this result should not be overgeneralized to suggest that using these three sensors alone are sufficient for general purpose sensing. In many cases, the other sensors provide useful data for *e.g.*, edge cases, and can make the difference between 85% and 95% accuracy. Second, as already noted, the sensor questions elicited from our participants were heavily skewed towards instantaneous events, chiefly because we spent roughly ten minutes in each location during the Facet & Privacy Elicitation study, which likely inhibited participants from fully considering environmental facets that might change over longer durations (like a draft in a poorly sealed window).

Finally, every environment is different and there is no doubt a long tail of questions that could be asked by end users. Although microphone, accelerometer and EMI might enable 90% of possible sensor questions, to be truly general purpose, a sensor board will need to approach 100% coverage. You can see in Table 2 that other, infrequently-used sensors are occasionally critical in classifying some questions, for example, the GridEye sensor for detecting kettle use (Table 2A), and the light color sensor for detecting when the classroom lights are on/off (Table 2, M/L). If these two sensors

![Confusion Matrices](image-url)
were dropped from our board, these two questions would likely have been unanswerable (at acceptable accuracies).

As an initial exploration of how other sensors can come into play – especially for sensing longer-duration environmental facets – we ran a series of small, targeted deployments in mostly new locations and at different times of the year. We plot raw data and highlight events of interest to underscore the potential utility of other sensor channels.

**Room Temperature Fluctuation.** We used our sensor tag to capture temperature variation in a room with a window-mounted air conditioner on a warm summer’s night (Figure 8). Note the accelerated slope of the temperature when the AC is turned on and off. Another example of HVAC cycling can be seen in Figure 15 (note also the change in behavior when the thermostat target temperature is moved from 70 to 72°).

**Non-Contact Temperature Sensing.** The GridEye sensor acts like a very low-resolution thermal camera (8×8 pixels), which is well suited for detecting localized thermal events. For example, in our kitchen location, the kettle occupies part of the sensor’s field of view, and as such, the radiant heat of the kettle can be tracked, from which its operational state can be inferred (Figure 9).

**Light Color Sensing.** We also found ambient light color to be a versatile sensing channel. For example, colors cast by artificial lighting (Figures 12-15), or sunrise/sunset (Figure 10), can provide clues about the state of the environment (e.g., bedroom window open, office door closed, TV on).

Additionally, many devices communicate their operational states through color LEDs (Figure 11), which can be captured and aid classification of different states.

**Multiple Sensors.** Figures 12 through 15 offer more complex examples of how multiple sensors can work together to characterize environments. For example, in Figure 13, we can see that opening a garage door causes a temperature, color and illumination change in the garage. Having multiple confirmatory signals generally yields superior classification accuracies. Figures 14 and 15 show how high- and low-sample-rate sensors can work together to capture the state of a car and an apartment.

**SECOND-ORDER SYNTHETIC SENSORS**

Up to this point, the synthetic sensors that we have discussed all operate in a binary fashion (e.g., is the “faucet running?” Possible outputs: yes or no). These are what we call first-order synthetic sensors. We can build more complex, non-binary, second-order synthetic sensors by leveraging first order outputs as new ML features. We explored three second-order classes: state, count and duration. The first-order sensors used in this section were trained using data from days 1 though 7 in the deployment study.

**State**

Two or more first-order synthetic sensors can be used as features to produce a second-order synthetic sensor for tracking the multi-class state of an object or environment. For example, in our two-week deployment study, we had five first-order synthetic sensors about a single microwave

---

Figure 8: Room temperature variation caused by an AC unit.

Figure 9: Heat radiating from the kettle can be captured by the GridEye sensor.

Figure 10: Light color captured over a one-hour period from an example sunrise (top) and sunset (bottom).

Figure 11: Various states of a MakerBot Replicator 2X over a ~30 minute period as captured by a light color sensor.

Figure 12: Data captured over a ~24 hour period in an outdoor parking lot on a warm summer day.

Figure 13: Data captured in a two-car garage over a ~36 hour period during winter.
As a real-world demonstration, we built a second-order count sensor that tracked the number of towels dispensed by the dispenser in our kitchen location. To test this counter’s accuracy, we manually dispensed 100 towels. At the end, our sensor reported 92 towels dispensed. As shown here, errors can accumulate, but nonetheless offers a reasonable approximation. Similar to our previous example, when the dispenser runs low, an order for more supplies could be automatically placed.

**Duration**

Similar to count, it is also possible to create second-order synthetic sensors that track the cumulative duration of an event, for example energy consumption or water usage (Figure 17). We performed two simple evaluations.

First, using our microwave running first-order sensor (Table 2N), we built a second-order “microwave usage” duration sensor. To test it, we ran the microwave 15 times with random durations (between 2 and 60 seconds). At the end of each run, we compared our sensor’s estimated duration to the real value. Across 15 trials, our microwave usage sensor achieved a mean error of 0.5 seconds (SD=0.4 sec).

As a second example, we used our faucet running first order sensor (Table 2E) to estimate water usage. To convert time into a volume of water, we used a calibration process similar to UpStream [32]. To test this sensor, we filled a large measuring cup with a random target volume of water (between 100-1000mL), and compared the true value to the classifier’s output. We repeated this procedure ten times, revealing a mean error of 75mL (SD=80mL).
Nth-Order Synthetic Sensors

Importantly, there is no reason to stop at second-order synthetic sensors. Indeed, first-order and second-order synthetic sensors could feed into third-order synthetic sensors (and beyond), with each subsequent level encapsulating richer and richer semantics. For example, appliance-level second-order sensors could feed into a kitchen-level third-order sensor, which could feed into a house-level sensor, and so on. A house-level synthetic sensor, ultimately drawing on scores of low-level sensors across many rooms, may even be able to classify complex facets like human activity. We hope to extend our system in the near future to study and explore such possibilities.

CONCLUSION

In this work, we introduced Synthetic Sensors, a sensing abstraction that unlocks the potential for versatile and user-centered, general-purpose sensing. This allows everyday locations to become "smart environments" without invasive instrumentation. Guided by formative studies, we designed and built novel sensor tags, which served as a vehicle to explore this broad problem domain. Our real-world deployments show that general-purpose sensing can be flexible and robust, able to power a wide range of applications.

ACKNOWLEDGMENTS

This research was generously supported by Google as part of the GIoTTTo Expedition Project, as well as with funding from the David and Lucile Packard Foundation. We especially thank Robert Xiao for his help with early development, and also Yuvraj Agarwal, Anind Dey, and Anthony Rowe for input throughout. Finally, we are also grateful to Dorothy and Liz Carter for their thoughtful ideas during the early stages of this project.

REFERENCES


