Fine-Grained Recognition of Activities of Daily Living through Structural Vibration and Electrical Sensing

Shijia Pan
span24@ucmerced.edu
University of California Merced
Merced, CA

Mario Berges
marioberges@cmu.edu
Carnegie Mellon University
Pittsburgh, PA

Juleen Rodakowski
jur17@pitt.edu
University of Pittsburgh
Pittsburgh, PA

Pei Zhang
peizhang@cmu.edu
Carnegie Mellon University
Moffett Field, CA

Hae Young Noh
noh@cmu.edu
Carnegie Mellon University
Pittsburgh, PA

ABSTRACT

Fine-grained non-intrusive monitoring of activities of daily living (ADL) enables various smart building applications, including ADL pattern assessments for older adults at risk for loss of safety or independence. Prior work in this area has focused on coarse-grained ADL recognition at the activity level (e.g., cooking, cleaning, sleeping), and/or course-grained (hourly or minutely) activity duration segmentation. It also typically relies on a high-density deployment of a variety of sensors. Finer-grained (sub-second-level and event-based) ADL recognition, could provide more detailed ADL information, which is crucial for enabling the assessment of patients' activity patterns and potential changes in behavior. To achieve this fine-grained ADL monitoring, we present a system that combines two emerging non-intrusive sparse sensing mechanisms: 1) vibration sensors to capture the action-induced structural vibration and 2) electrical sensor to capture appliance usage. To evaluate our system, we conducted real-world experiments with multiple human subjects to demonstrate the complementary information from these two sensing modalities. Results show that our system achieved an average 90% accuracy in recognizing activities, which is up to 2.6× higher than baseline systems considering each state-of-the-art sensing modality separately.

CCS CONCEPTS

• Information systems → Sensor networks; **• Human-centered computing → Ubiquitous and mobile computing design and evaluation methods**.

KEYWORDS

Fine-grained ADL; Structural vibration sensing; Electrical load sensing; Non-intrusive; Ensemble; Complementary

BuildSys '19, November 13–14, 2019, New York, NY, USA

ACM ISBN 978-1-4503-7005-9/19/11…$15.00
https://doi.org/10.1145/3360322.3360851

ACM Reference Format:
Shijia Pan, Mario Berges, Juleen Rodakowski, Pei Zhang, and Hae Young Noh. 2019. Fine-Grained Recognition of Activities of Daily Living through Structural Vibration and Electrical Sensing. In *The 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '19), November 13–14, 2019, New York, NY, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3360322.3360851

1 INTRODUCTION

The Internet of Things (IoT) and its rapid development enables various smart home applications that have the potential to support independent living for older adults [2, 20]. Engagement in activities of daily living (ADL) is an important metric for these smart home applications to monitor, as engagement in ADL is associated with the risk of disability and all cause mortality for older adults [33]. One way that variation in an ADL can be detected is by the length of time or missed steps within an ADL. For example, an older adult with cognitive impairments may insidiously decline in engagement of ADL (e.g., take longer to perform an ADL or miss steps within an ADL) as their cognitive impairments progress [31]. Non-intrusive, fine-grained and in-home ADL monitoring provides a critical platform to detect variation in ADL and ensure safety and independence in the home.

Prior work in ADL monitoring mainly focus on duration segmentation and type recognition to describe ADL patterns. ADL duration segmentation relies on dense deployment of sensors that capture a sequence of human interaction with ambient objects (e.g., drawers, doors) to determine the duration of an activity [18, 19]. Activity type recognition methods leverage learning algorithms to improve the accuracy and robustness for classifying given sensing signals [4, 6]. Combined, these efforts focus on activity-level information with the time-resolution of minute or hour, which is coarse-grained. It is indeed challenging to achieve *fine-grained*, which we define as *sub-second-level and event/action-level*, ADL recognition *non-intrusively and sparsely* because each ADL consists of several events or actions. Nonetheless, fine-grained ADL monitoring provides detailed ADL action information, which enables a nuanced understanding of ADL patterns and, most importantly, provides knowledge of when changes in ADL patterns occur. A potential change in ADL patterns may be an indication of changes in disease status or safety for living independently. Prior attempts for fine-grained ADL monitoring combine electrical sensors and passive RFID sensors, where the on-wrist RFID provides locations and electrical sensor provides appliance usage information. These methods (e.g., [8]) require high-density sensor deployment and people carrying devices or tags during their activities. Older adults, especially those with cognitive impairments, may find it difficult to remember to wear or uncomfortable to wear such devices.

Two lines of research suggest that structural vibration and electrical load monitoring provide distinct and unique information about ADL patterns. On the one hand, researchers have noted that when people interact with their ambient environments their actions induce the structures to vibrate, and have used this vibration to infer various types of information [23, 27] including the action (or motion) of the person [7]. Additionally, non-intrusive load monitoring methods have been shown to detect appliance usage duration[3] from aggregate measurements at the main electrical meter. Thus, we combine these two complementary non-intrusive and passive sensing modalities to cover two important aspects of ADL patterns – occupant action and appliance usage – with fine granularity.

Our system conducts event detection and event-based ADL recognition on these two sensing modalities. The system integrates these estimates over high-resolution time windows using an ensemble algorithm. The contributions of this work are as follows.

- We introduce a fine-grained (sub-second-level, event/action-level) ADL detection and recognition system using structural vibration and electrical sensing.
- We present an event-based ADL detection and recognition framework, and an ensemble algorithm to fuse ADL predictions from structural vibration and electrical sensing.
- We conduct real-world experiments to evaluate our systems and demonstrate its effectiveness and complementary nature of the selected sensing modalities.

The rest of the paper is organized as follows. Section 2 discusses the related work and contrast it to our approach. Then, Section 3 presents the design of our system in detail. Next, Section 4 demonstrates the results and analysis of real-world experiments and evaluations. Finally, we discuss limitations as well as future directions of this work in Section 5 and conclude in Section 6.

2 RELATED WORK

With the growth of the Internet of Things (IoT), devices or systems with sensing abilities have become ubiquitous in daily life. As a results, many systems and learning algorithms are explored to investigate the problem of activity of daily living (ADL) recognition. We summarize them into the following categories.

2.1 Smart Home ADL Monitoring Systems

Various sensing systems have been developed to monitor smart home ADL. Kokku et al. proposed the concept of activity signatures in the smart home environment with a variety of sensors [18–20]. They proposed to utilize temporal cut, sensor cut, and frequency cut to finally determine the activity segments at a relatively coarse grained task-level, e.g., sleep, bath, eat. Fortin-Simard et al. fused electrical load and the passive RFID sensors to achieve home activity recognition, where the on-wrist RFID provides locations and electrical load sensor provides appliance usage information [8]. Moriya et al. explored this direction with Echonet Lite [24]. Compared to these, our approach can detect more types of movements without requiring human subjects to wear any device nor dense deployment on each monitored appliance.

2.2 Vibration Sensing for Human Monitoring

In recent years, vibration based human sensing has been developed, providing us with a non-intrusive passive sensing modality for indoor occupant monitoring. It has been explored to obtain occupant information [30] including identity [27], location [23], heart rate [17], hand washing activities [7], etc. When the occupant interacts with ambient objects, such as floor, table, walls, bed, sink, etc., the interactions induce the structure to vibrate in a unique way such that their frequency components reflect the mode excited by different excitation sources [7]. Compared to other sensing modalities, it does not require that the user wear a device and, as a result, it allows ubiquitous indoor occupant activity monitoring. Since it detects action induced surface vibration, it allows fine-grained ADL monitoring, which is the goal of this work.

2.3 Electrical Sensing for Appliance Usage Monitoring

Non-Intrusive Load Monitoring (NILM) has been explored as an efficient way to monitor in-home appliance usage and related activities by disaggregating the total electrical usage of a building into its constituent components (i.e, appliances)[14, 22]. Though the field has largely focused on inferring appliance usage using a variety of different approaches (e.g., voltage noise [9, 12, 29], harmonic power [3, 11], etc.) new research has started to look into derivative objectives such as fault detection and diagnosis of appliance patterns and, relevant to this work, monitoring ADLs (e.g., [1]). Though our envisioned solution would make use of NILM algorithms, the work presented here is intended to serve as a proof-of-concept and, as such, we only analyze the appliance-level power consumption. In light of this, existing NILM work on appliance-level identification (e.g., [10]) is most relevant.

Accurate detection and recognition of different appliance usage is an important aspect of ADL monitoring, and these prior works have shown the feasibility and robustness of the recognition. As a result, we believe the combination of structural vibration and electrical load covers two important aspects of human activity in home scenario – with or without using appliance. We focus on the combination of these two non-intrusive sensing modalities.

3 SYSTEM DESIGN

To achieve non-intrusive fine-grained ADL detection and recognition for long-term monitoring, we measure two essential aspects of in-home activity – human actions (via structural vibration sensor) and appliances usage (via electrical sensor). As shown in Figure 1, our system first obtains signals from both structural vibration and electrical sensors (Section 3.2). Then, it conducts event detection on signals from each sensor (Section 3.3). Next, the system classifies the activities at the event-level (Section 3.4). Finally, our system conducts an event-based prediction ensemble to provide accurate recognition for sub-second time windows (Section 3.5).

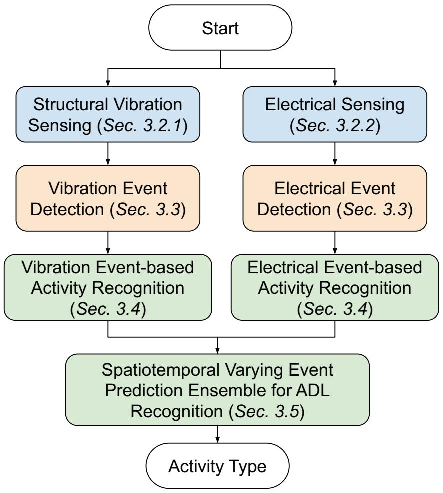


Figure 1: System overview.

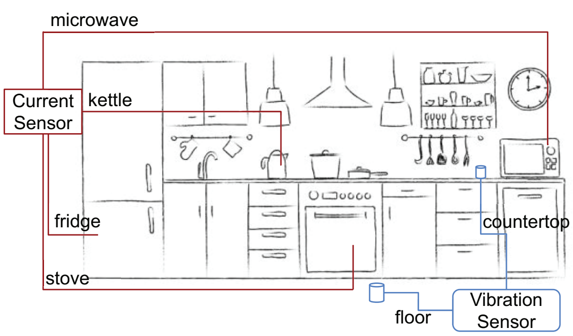


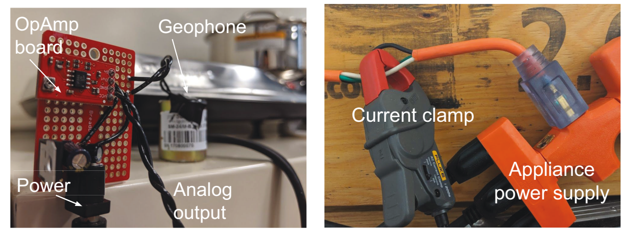
Figure 2: Sensing system conceptual diagram.

3.1 Complementary Sensing Modalities

To achieve non-intrusive monitoring for smart home applications, we selected structural vibration and electrical sensor as the primary sensing modalities. These two sensing modalities are selected because they are both non-intrusive – indirectly inferring activities instead of directly measuring – and complementary with each other. The structural vibration sensing captures the human interaction with the ambient environment, which is mostly the impulsive excitation [7, 13, 23, 27, 30]. The structural vibration sensing also captures the appliance machinery vibration, such as a motor or a compressor, as well as appliance usage induced vibration, such as water or food boiling. On the other hand, the electrical sensor precisely detects the appliance usage time and duration, which the structural vibration sensing may not [11, 14, 28, 32]. Since the appliance usage, as well as human motion or interaction with the appliance, are the two significant aspects of human activities [8, 24], these two sensing modalities are complementary for our purpose.

3.2 Sensing System

Our sensing system consists of the structural vibration sensing and the electrical sensing. Figure 2 demonstrate a conceptual scenario of our system deployed in a kitchen as an example. The load sensor measures the appliances . The primary structural surfaces (e.g., the countertop and the floor) are equipped with vibration sensors to capture human action caused vibration.



(a) Vibration sensor (b) Electrical sensor

Figure 3: Sensing hardware.

3.2.1 Structural Vibration Sensing. The structural vibration can be used to infer the occupant actions causing the vibration. When people interact with objects or structure around them, the interaction causes the surface of the object or structure to deform [23, 27, 30]. The surface deformation causes mechanical waves dominated by the Rayleigh-Lamb wave. These waves propagate through the structural and are captured by the sensor. Since different activities or interactions excite different modes of the structure [7], they induce vibration with different frequency domain characteristics, which can be used as features to recognize them. We place vibration sensors on the surfaces, including floor and countertops, where the human and appliance interact directly. A vibration sensor consists mainly of three modules – a geophone that obtains the surface vibration, an amplifier module that amplifies the surface waves, and an ADC module that converts the analog signal to digital. Figure 3(a) shows an example of the structural vibration sensor on a countertop. The Geophone in the figure is SM-24 [16]. The opamp board has a modified gain for the monitored surface.

3.2.2 Electrical Sensing. All electrical measurements were collected using a 16-bit National Instruments (NI-9215) data acquisition interface. Current was measured with a Fluke i200 AC current clamp with a cut-off frequency of 10 kHz, and voltage was measured with a Pico-TA041 Oscilloscope probe. Both measurements were done on a power strip to which all appliances in the testbed were connected to, as shown in Figure 3(b). The setup we used is similar to the one used in [10].

3.3 Event Detection

For each type of sensors, an *event* is defined as a segment of signal that has distinguishing characteristics compared to the signal segment when no human activities occur. Our system conducts event detection on raw sensor signals. The intuition is that when there are no activities – neither human interacting with the ambient environment nor appliance usage – the signals obtained by the two sensing modalities are considered as the ambient noise. We analyze the signal with a sliding window. The sliding window that covers the ambient noise signals has different signal energy distribution compared to that of an event [25]. We model the sliding windows

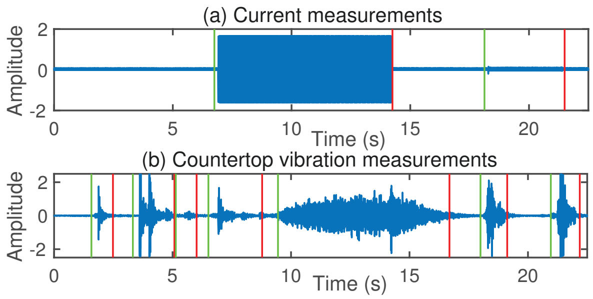


Figure 4: Examples of event detection. Solid blue lines are the detected signals. Green lines mark the beginning of events, and red lines mark the end of events.

that cover the segments of signals known as ambient noise with Gaussian distribution, and we consider the tested sliding window that has significantly higher signal energy as part of an event. One event contains consecutive sliding windows that are detected as part of an event. Figure 4 shows an example of the ambient noise for both current sensing and vibration sensing between x-axis 0 and 1 second. The solid green and red lines indicate the start and the end of an event for these sensing signals, which have a significant signal energy difference compared to the ambient noise signal. As a result, we use anomaly detection algorithms [27] to detect and extracts the signal segments that are not ambient noise as *events*.

3.4 Event-based Activity Recognition

Our system enables non-intrusive passive sensing for action- or event-level ADL recognition. The system first conducts feature extraction on the detected event signal segments (Section 3.4.1). Then the features of the labeled data are used to train a classifier using support vector machine (SVM) (Section 3.4.2).

3.4.1 Feature Extraction and Normalization. For a detected event signal segment, our system extracts its frequency domain characteristics – the power spectral density – as features.

Electrical sensing. The load sensor signals have an energy concentration approximately at 60 Hz in the frequency domain. It is because the local alternating current is of 60 Hz. On the other hand, the non-linear loads in the circuit often induce the current harmonics, which is unique for the particular circuit of the appliance. As a result, the current harmonics induce frequency characteristics that are distinguishable for different appliances. Therefore, for events extracted from electrical sensing signals $Event\_Load_i$, where $i = 1...N_{load}$, N_{load} is the number of events detected by the electrical load sensor, we first normalize the signal by its energy to reduce the variation caused by different appliance usage duration.

For appliances that are mainly linear load, i.e., their frequency components do not show current harmonics, which may make the accurate classification difficult with only features from the electrical load sensor. Therefore, our system extracts the signal of the same time duration of $Event\_Load_i$ from vibration sensors where the monitored surface has multiple appliances on it. We refer to this signal segment as $Event\_Load\_Vib_{i,j}$, where $j = 1...S_{multi}$, S_{multi} is the number of surfaces that has multiple appliances on them. Note that for signals collected from vibration sensors, the signal segment of the same time duration may not be detected as an event or a part of an event. The system takes the frequency domain characteristics of $Event\_Load_i$ and $Event\_Load\_Vib_{i,j}$ and concatenate them as the feature for the i^{th} event detected by the electrical load sensor.

Structural vibration sensing. Various human actions cause different parts of the structural surface to vibrate. Most of the human interaction with an ambient surface are impulsive, meaning the interaction excites the surface and induces vibrations dominated by the Rayleigh-Lamb waves [26]. For varying excitations occurring on the same surface, they will generate different responses in the natural frequencies of the surface structure [7]. For the events detected by vibration sensor on surface k during t_i, we refer to this signal segment as $Event\_Vib_{k,i}$, where $i = 1...N_{vib}$, $k \in [1...N_{surface}]$. N_{vib} is the number of events detected by the vibration sensor on the k^{th} surface. $N_{surface}$ is the number of monitored surfaces.

To take into account the spatial characteristics of the signal, we further extract the signal segment of the same time duration t_i of $Event\_Vib_{k,i}$ from vibration sensors on another surface l, which we refer to as $Event\_Vib\_Vib_{k,l,i}$, where $l = 1...N_{surface}, l \neq k$. Our system extracts the frequency components of $Event\_Vib_{k,i}$ and $Event\_Vib\_Vib_{k,l,i}$ after normalizing the signal by energy, and concatenates the frequency components as the features for the i^{th} event on surface k.

3.4.2 Classification with Support Vector Machine. Our system utilizes support vector machine (SVM) [5] to conduct the classification for the events detected by each sensor. SVM is a widely used classifier, and it aims to find the maximum-margin hyperplane $\boldsymbol{w}$ by minimizing the following loss function (for binary classification):

$$\min_{\boldsymbol{w},b} \quad \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{l}\max(1 - y_i(\boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b), 0), \tag{1}$$

where $(y_1, \boldsymbol{x}_1), \ldots (y_l, \boldsymbol{x}_l)$ is training data, $\boldsymbol{x}_i \in R^n, \forall i$ are training samples, $y_i = \pm 1, \forall i$ are training labels, and C is the penalty parameter that controls the generalization of the model. Based on our feature analysis, which will be introduced in details later in Section 4.4, features of events detected by vibration sensors are not linearly separable. As a result, our system trains the nonlinear SVM model using the kernel function $\phi(\cdot)$ (RBF kernel) to ensure high class separability. Since we have more than two types of activities to classify, we decompose the multi-class classification problem to two-class classification problems to solve [5, 15].

3.5 Ensemble Events for ADL Recognition

Once the system obtains the event-based classification predictions from each sensor, it further conducts prediction ensemble on the sliding window at a sub-second level, and then outputs the activity recognition at the event-level as shown in Figure 5. Since the type and duration of events detected by different sensors vary, the ensemble occur at a sub-seconds sliding window level instead of the event-level. We selected the sliding window that is smaller than a single impulsive structural vibration signal segment empirically.

Our system first applies a sliding window through the target sensing time duration. For each sliding window on the signal from each sensor, our system assigns it as an event if the majority of

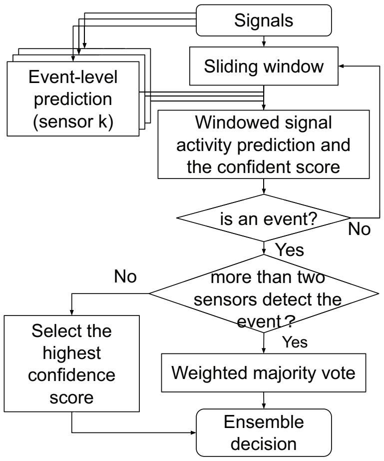


Figure 5: Ensemble algorithm.

the samples are part of a detected event. Since we use the SVM to predict the event categories, a confidence score between 0 and 1 can be calculated based on the distance between the data point and the margin [5]. Therefore, for a sliding window of a particular event, we also assign the prediction score to it.

For the target sliding window, the system collects the prediction and the prediction score from all sensors. The system first conducts a weighted majority vote if there are more than two sensors detects the sliding window as part of an event. The weight are assigned equally over vibration sensors monitoring each surface and the electrical load sensor. For example, if there are multiple sensors on the same surface, e.g., floor, the information from the multiple sensors will be combined. This means that each sensor type is weighted equally. In addition, since different sensors may detect different event durations and types, there may be no more than two sensors detecting an event within a sliding window. When that happens, the system outputs the single sensor decision with the highest prediction score as the final decision of the system instead of conducting a majority vote.

4 EVALUATION

To evaluate our system, we facilitated engagement in ADL through real-world experiments with multiple appliances in a kitchen scenario. In this section, we first define the ADL, at the event-level, that were conducted in the kitchen in the scenario in Section 4.1. Then, we introduce the experiments we conducted and the metrics and ground truth used for evaluation in Section 4.2. We also explain the dataset characteristics and justify our evaluation metrics and its presentation. The results for event detection module, classification module, and the overall system performance are analyzed in Section 4.3, 4.4, and 4.5 respectively.

4.1 Kitchen Scenario Definition

ADL are critical for older adults. Performance of ADL is critical to ensure safety and independence in the home. Changes in ADL are associated with disability and institutionalization. The Lawton Instrumental Activities of Daily Living Scale is a commonly used tool to assess ability to live independently [21]. Critical components of the Lawton tool include food preparation and housekeeping. We focused on tasks related to meal preparation and housekeeping and they are in Table 1 below. We assessed these ADL in adults with no identified physical or cognitive impairments, as the intent of this study was detection of normal ADL. The table depicts how the structural vibration sensors and electrical load sensors detect different events within the same ADL.

Table 1: Target kitchen activities and their acronyms.

Activity	Electrical	Vibration
use stove	stove (S)	put on stove (PS)
use kettle	kettle (K)	use kettle (K)
use microwave	door (MD)	open/close door (OM)
	heating (M)	put down food (OM)
vacuum floor	vacuum (V)	use vacuum (V)
walking	NA	footsteps (Step)

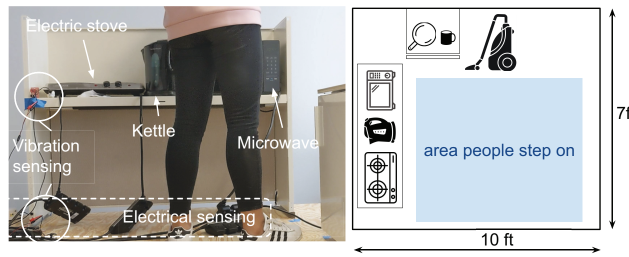


Figure 6: Experimental setup.

4.2 Experiments

We conducted real-world experiments in our laboratory setting to evaluate our system following the guideline of the IRB protocol. The experiments are conducted on a wooden floor structure with an area of 10×7 ft with one electrical load sensor and two structural vibration sensors. We put a countertop on the wooden floor structure, and placed an electrical stove, a kettle, and a microwave on the countertop as shown in Figure 6. We also placed a vacuum cleaner on the side of the floor. Two vibration sensors are used, one placed on the countertop (Figure 3(a)) and one on the floor, as circled out in Figure 6. The target appliances – stove, kettle, microwave, and vacuum – are connected to the outlets of the power strip (Figure 3(b)). We also provide a pot to boil water on the stove, and a mug to heat water using the microwave.

In total, five human subjects are invited to conduct the experiments. We refer to one trial of data collection of a subject conducting a sequence of different types of activities as a *routine*. Two of the five conducted all the listed types of activities in Table 1 in their routine. The rest conducted a subset of the types that reflect their ADL for preparing meal or cooking, e.g., in some trials if they select to use the stove for cooking, they will not use the microwave, and vice verse. Each person conducted five routines. As a result, we first conduct the classification on the two people's data with 80% data for training and 20% data for testing through cross validation

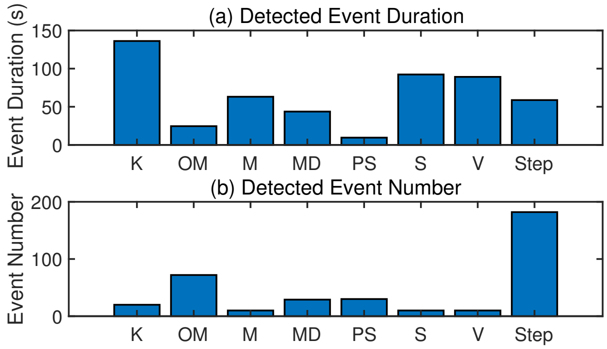


Figure 7: Experiment data stats.

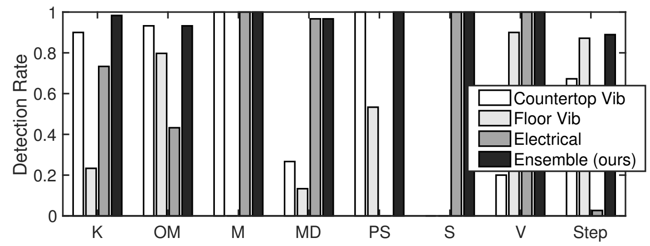


Figure 8: Detection results at event level.

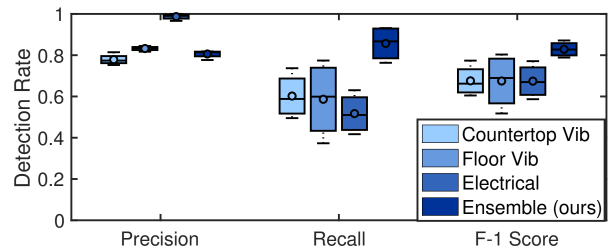


Figure 9: Detection results at sample level (duration).

(Section 4.5.1). Then we train the system with these two people's data and test on the data from the rest three to demonstrate the system robustness with individual variations (Setion 4.5.1).

4.2.1 Metric and Ground Truth. The designed experiments target three different types of accuracy, 1) the event detection accuracy, 2) the event recognition accuracy, and 3) the sub-second sliding window level ensemble activity recognition accuracy, to evaluate our system. In general, we consider four types of baselines including each sensor by themselves and combining two vibration sensors data by fuse them with our ensemble algorithm.

We use a camera to record experiments as the ground truth. The video records from the angle below the waist so that the identity of the human subjects is not recorded. Figure 6 shows an example view of the ground truth recording. The listed events (Table 1) are labeled on a frame by frame basis manually.

4.2.2 Dataset Statistics. We further analyze the collected and labeled data in terms of event duration and number. Figure 7 demonstrates the corresponding stats from the data collected from the two people conducted all activities. Figure 7(a) shows the overall event duration and (b) shows the count of each type of events conducted. We can observe a clear bias in both event duration and number. Furthermore, this bias is also not consistent over the type of events. For example, kettle usage (K) has the highest over all duration, but the number of events are relatively low. On the other hand, the number of footsteps are high, while its duration is low due to the short period of events for each footsteps. Therefore, if we compare all the events' activity recognition accuracy in either event duration or number, it will be a biased.

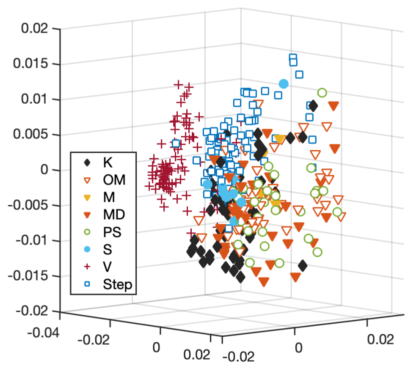
To fairly compare our system's performance to aforementioned baselines, we calculate the activity recognition accuracy for each type of activities (eight types in total) respectively. For the final system performance, we report the average accuracy across all types of activities (average of the eight accuracy values).

4.3 Event Detection Analysis

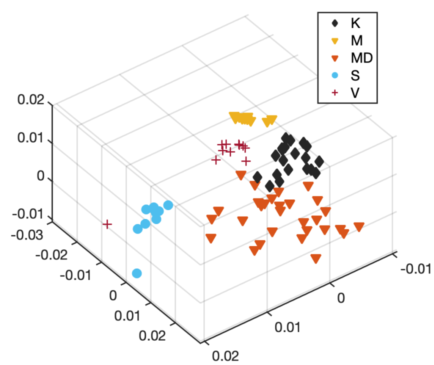
Our system combines multiple sensors that detect different aspects or duration of an event to increase the accuracy. We use the detection with signals from only one sensor as the baseline and compare to our ensemble event detection. To avoid the bias over different event duration and number, we first compare the detection rate of each type of activities at the event level. This detection rate is calculated as follows. For each labeled event, we extract their indexes of $[E_1...E_L]$, where L is the sample number in the event. For the evaluated signal, we extract the detection event indexes $[DA_1...DA_M]$, where M is the sample number of the detected event. If the majority elements of the segment are part of an event, i.e., the number of samples in $[DA_1...DA_M] \cap [E_1...E_L]$ is larger than $L/2$, we consider this event detected.

We present the detection rate in Figure 8. Note that this detection rate is a binary concept, it does not reflect types of (overlapping) events being detected by different sensors. We further evaluate the types of event recognition in Section 4.4. We observe that for events that occur on the floor, i.e., vacuum and walking, the vibration sensor on the floor demonstrates higher detection rates (90% and 87%). For the activity occur on the countertop, i.e., operating kettle, microwave, and stove, the vibration sensor on the countertop achieves a higher accuracy (90%, 93%, and 100%). On the other hand, the electrical load sensor achieves the highest accuracy for the event that only involves activating appliances, e.g., microwave heating, microwave door open, stove heating, and vacuum (100%, 97%, 100%, and 100%). The average detection rate over eight types of events are 62%, 44%, 65%, and 97% in respectively for the three baselines and our approach, which is a 1.5× to 2.2× improvement compared to the baselines.

Since for the event detection rate at the event-level only reflects the true positive rate, we further analyze the sample-level (duration) detection precision and recall rate in Figure 9. We plot the three baselines and our ensemble approach with bars from light blue to dark blue. We observe that the current sensor achieves high precision with a mean value of 99%. However, because not all the activities can be measured by the electrical sensor, the recall for current value is fairly low with a mean of 52%. The vibration sensors at two locations showed similar of precision and recall rate – the



(a) Countertop vibration sensor



(b) Electrical load sensor

Figure 10: Feature analysis with PCA.

mean precision values are 78% and 83% and the mean recall values are 60% and 59%. Because they are complementary in detecting events occur on different surfaces, the ensemble detection achieved a mean recall of 86%, which is a 1.4× improvement compared to that of only using one sensor for detection. Our approach achieves the highest F-1 score of 0.83.

4.4 Event-based Activity Recognition Analysis

For event-level ADL recognition, different sensors perform differently due to their spatio-temporal variation even using the same classification algorithm. In this section, we first conducted analysis on the extracted features, then we discuss the classification accuracy for each sensor.

4.4.1 Feature Analysis. We visualized the features of different types of activities to verify that the selected features are effective representatives for distinguishing these events. We applied the principal component analysis (PCA) on features and plotted the data points (features) by their projections onto the first three components. We plotted the events using the same ground truth labels in the same marker and color. Figure 10(b) shows the features of five types of events detected by the electrical sensor, which are linearly separable. Figure 10(a) shows the features of the eight types of event-level activities detected by the vibration sensor on the countertop. It indicates that nonlinear SVM is sufficient for this feature set.

4.4.2 Event-level Recognition Analysis. We further demonstrate event-level classification confusion matrices in Figure 11 (a), (b), and (c) for each sensor respectively. The electrical sensor achieved an overall (biased number of events) recognition accuracy of 92.5% when only considering types of activities that contain electrical activities. The average classification accuracy for these five detectable activities is 97%. However, we take all eight types of activities into account, this average drops to 61%.

The vibration sensor on the countertop captures all types of activities. Most of the actions (OM,PS,V,Step) induce impulsive structural vibrations, which achieved over 90% prediction accuracy. Appliances that induce signature machinery vibration (M) achieved 100% prediction accuracy. Unlike the microwave, stove (S) and kettle (K) do not cause the machinery vibration via drive motor, however, the boiling may or may not cause the vibration that can be detected by the sensor on the countertop. The misclassification mostly occur between the operation of the microwave (OM) and the open status of the microwave door (MD). This could be caused by the similar spatial characteristics of these events, i.e., both on the microwave, and the similar impulsive signals induced by open/close door and putting down the mug when the microwave door is open. Since the microwave door open induce a signature current change, the ensemble prediction achieves a higher accuracy when taking both sensing modalities into account. The average of the classification accuracy values for eight types of activities with the vibration sensor on the countertop is 81%.

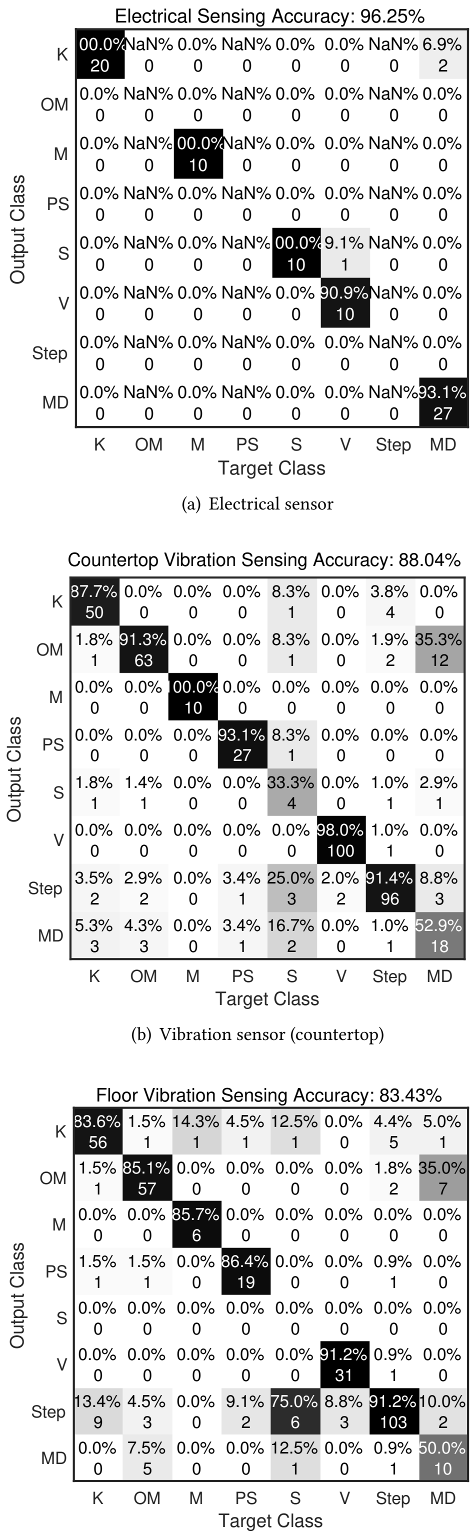
For the vibration sensor on the floor, we observe that most of the stove activities are misclassified as the footsteps (75%). This could be caused by the ambient impulsive floor vibration that is not induced by stove activities being captured by the sensor, i.e., people's micromotion or other building activities. The kettle usage, i.e., turning on/off the kettle, and the footsteps are misclassified with each other at a rate up to 13%, which could be caused by the similarity between these two types of impulses. The average of the classification accuracy value for eight types of activities with the floor vibration sensor is 72%.

The event-based activity recognition module showed 70% activity recognition accuracy for all the sensing modalities and each sensor demonstrated their highest recognition rate for different types of activities. As a result, when these results are fused, the ensemble activity recognition accuracy is higher than that of single sensors.

4.5 System Performance Evaluation

We first compare our ensemble approach to the state-of-the-art baselines (Section 4.5.1). Then we explore the robustness of the system to the variation across human subjects (Section 4.5.2).

4.5.1 Ensemble Activity Recognition. Since the ensemble activity recognition is conducted on each sliding window (Section 3.5), we evaluate the prediction accuracy for sliding windows. Figure 12 shows an example of the prediction values using different approaches. The x-axis is time and the y-axis is approaches. Different



(a) Electrical sensor

(b) Vibration sensor (countertop)

(c) Vibration sensor (floor)

Figure 11: Event-based classification confusion matrix.

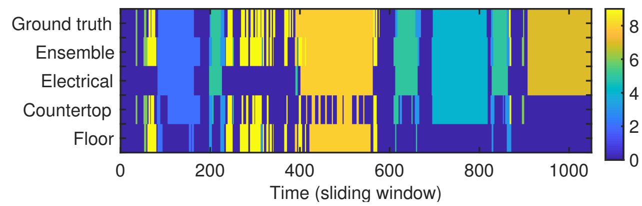


Figure 12: Example of sub-second window-level activity recognition. x-axis is time and y-axis is methods, colors indicate class ID. Our method demonstrates the highest consistency compared to the ground truth.

colors are the prediction output (class IDs corresponding to types of activities). We observe that the ensemble demonstrates a higher similarity to the ground truth compared to the single sensor predictions, indicating a higher recognition accuracy. We further quantify the recognition accuracy over each type of activities.

Figure 13 demonstrates the classification accuracy for each type of activities with bars from light to dark shades representing results using 1) countertop vibration sensor, 2) floor vibration sensor, 3) electrical sensor, and 4) countertop and floor vibration combined. Our ensemble approach is presented by green bars. The observation for the three single sensor baselines are consistent with the confusion matrices discussed in Section 4.4 – each sensor achieves high accuracy for different events. For example, the vacuum is detected mostly by the floor vibration sensor and the electrical sensor and achieved the highest (93%) recognition accuracy among all the methods. The placing item on stove (PS) and the stove heating (S) are detected respectively by the vibration sensor (91%) and the electrical sensor (100%), which results in an average stove usage recognition accuracy of 45% and 50% for these two sensing modalities, and an average accuracy of 92% for our ensemble approach.

The countertop and floor vibration combined approach achieves an average of 64% accuracy over eight types of activities, which is higher than that of the single sensor approach. It is because that the vibration sensors on two surfaces are complementary in special characteristics, which allows high confidence and accuracy for the events occurring on each surface. Our ensemble approach achieved the highest classification accuracy for half of the events, and the average accuracy over eight events is 90%, which is a 1.5× to 2.6× improvement compared to the baselines (56%, 35%, and 61%). The average values for the baselines here are calculated at the sliding window level, which is different from the average values in Section 4.4 calculated at the event level. In summary, the complementary spatial characteristics of the vibration sensors improved the event-level ADL recognition accuracy and the complementary temporal characteristics of the two sensing modalities further increased the accuracy further more.

4.5.2 Personal Variation. The subject performs ADL differently in various aspects, e.g., speed, strength, interaction. As a result, their ADL may cause different signal characteristics (data distribution) [13]. To understand the system robustness to the individual action variation, we further evaluate the model trained on the two persons' data and test it with the three other participants. We plot the average

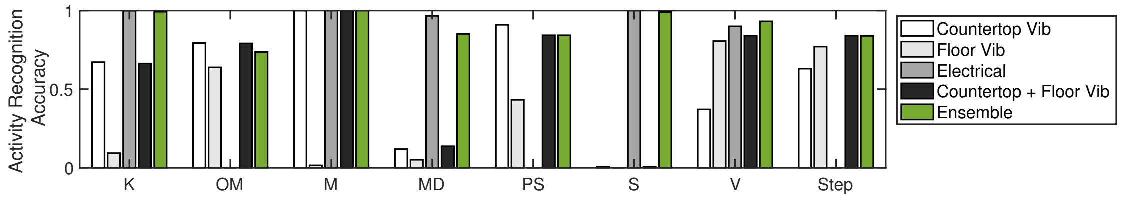


Figure 13: Sliding window level ADL recognition accuracy for eight investigated ADLs.

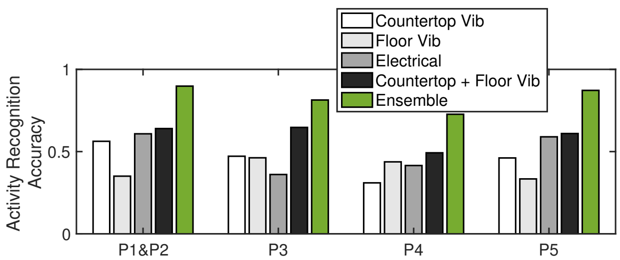


Figure 14: ADL recognition accuracy when training on two persons' data (P1 & P2) and testing on different people.

accuracy over eight types of activities in Figure 14, where bars from light to dark shades represent the accuracy for countertop vibration sensor, floor vibration sensor, electrical sensor, and countertop and floor vibration combined approach. Our ensemble approach is plotted as green bars. We can see that for different people, the accuracy for different methods varies. For example, for Person 5, the electrical sensor achieves higher accuracy compared to Person 3 and 4. While for Person 4, the floor vibration sensor achieves the highest accuracy among baselines. Despite the variation, three baselines for Person 3,4,5 showed comparable accuracy (between 30% to 60%) compared to that of Person 1 and 2. Our ensemble approach achieves the highest accuracy compared to the baselines, which are 81%, 73%, and 87% respectively for Person 3,4,5. Compared to the case where the training and testing data are from the same person different trials, the average accuracy of the three participants drops to 80%, but it is still 20% higher than baselines when the training and testing samples are from the same person.

5 DISCUSSION

This work highlights the potential for non-intrusive fine-grained ADL monitoring to detect patterns of ADL for older adults seeking safety while living in the community. We further discuss the current limitations and future directions of this work in this section. The key directions we plan to further explore beyond this work include: 1) the recognition and segmentation of activities when the same person conducts different activities simultaneously (Section 5.1), and 2) based on the detection and recognition, how can we conduct behavior level monitoring for long term in-home monitoring (Section 5.2). Once these directions have been explored, it would also be important to explore the sensitivity of our solution to errors introduced by NILM algorithms when relying on their estimates for appliance-level measurements.

5.1 Overlapping activities: multiple people, multiple activities

When one or more persons conduct multiple activities within the same sensing area, these activities signal – in both sensing modalities – may overlap. When signal overlapping occurs, it alters the signal characteristics, (i.e., features) used for classification. Prior work on disaggregating information for electrical load monitoring would allow the simultaneous appliance usage monitoring, which can be used to assist the separation or segmentation of the structural vibration sensing events. The direct separation of the structural vibration signal when multiple excitation sources' signal mix, however, is a challenge. Prior work on blind source separation cannot be directly applied here due to the structural dependency of the signal, which makes the assumption of signal independence invalid. However, with the historical data where one activity with one appliance, as well as the disaggregation of the electrical sensing, the prior knowledge on the mixed signals can be inferred for accurate overlapping activity detection and classification.

5.2 Behavior level monitoring: metrics and parameters

Our aging society has a desire to independently live in home. Non-intrusive in-home monitoring of ADL has the potential to support safety and independence for these older adults. Information patterns in engagement in ADL is critical, as engagement is known to be associated with disability, institutionalization, and all-cause mortality. With non-intrusive in-home monitoring, changes in patterns can be immediately identified and appropriate supports or care can be deployed to ensure safety and independence in the home for the older adult. This study is a step in showing that non-intrusive in-home monitoring detect ADL in adults with no known physical or cognitive limitations. These findings may be able to extend to older adults who begin to experience changes in their capacity for performance of ADL. For example, when an older adult is taking longer to engage in a cooking task or forgets the leave the stove on, non-intrusive in-home monitoring could detect a change in behavior that notifies family or health-care providers that can visit the older adult to ensure health and safety in the home. On the other hand, changes in patterns of ADL may cause the interaction between human and the structural changes, which may lead to data distribution change causing learning or classification error. To combat this potential limitation, continuous learning approaches may be needed to adapt to such data distribution changes. Another challenge is the metric to measure the behavior changes, especially for multilevel activity monitoring. The abnormal behavior detected at different granularity may indicate different aspects of the disease progression. A third challenge is the variation between subjects, i.e., the definition and measurement of the anomaly may vary.

6 CONCLUSION

We presented a non-intrusive fine-grained ADL monitoring system through ambient structural vibration and the electrical sensing in this paper. We highlighted the complementary information acquisition for these two sensing modalities and how to acquire high time- and type-resolution monitoring in this work. Both pieces of information may be used for the development of smart home applications seeking to monitor engagement in ADL. Our system first conducts event-level detection and recognition, and then applies an ensemble algorithm on the recognition results from each sensor over the target time period to achieve accurate event-level ADL monitoring. In the real-world experiments we conducted using common kitchen activities, our system achieved an average of 90% ADL recognition accuracy the real-world ADL experiments, which is an up to 2.6× improvement compared to the baselines.

ACKNOWLEDGMENTS

This research was supported in part by Highmark, the National Science Foundation (under grants CMMI-1653550), Intel and Google. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of CMU, UCM, NSF, or the U.S. Government or any of its agencies.

REFERENCES

[1] Jose Alcala, Jesus Urena, Alvaro Hernandez, and David Gualda. 2017. Assessing human activity in elderly people using non-intrusive load monitoring. *Sensors* 17, 2 (2017), 351.

[2] Iman Azimi, Amir M Rahmani, Pasi Liljeberg, and Hannu Tenhunen. 2017. Internet of things for remote elderly monitoring: a study from user-centered perspective. *Journal of Ambient Intelligence and Humanized Computing* 8, 2 (2017), 273–289.

[3] Mario Berges, Ethan Goldman, H Scott Matthews, and Lucio Soibelman. 2008. Training load monitoring algorithms on highly sub-metered home electricity consumption data. *Tsinghua Science and Technology* 13, S1 (2008), 406–411.

[4] Federico Castanedo. 2013. A review of data fusion techniques. *The Scientific World Journal* 2013 (2013), 1–19.

[5] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2, 3 (2011), 27.

[6] Emiro De-La-Hoz-Franco, Paola Ariza-Colpas, Javier Medina Quero, and Macarena Espinilla. 2018. Sensor-based datasets for human activity recognition–A systematic review of literature. *IEEE Access* 6 (2018), 59192–59210.

[7] Jonathon Fagert, Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. 2017. Monitoring hand-washing practices using structural vibrations. *Structural Health Monitoring* shm2017 (2017), 1–10.

[8] Dany Fortin-Simard, Jean-Sebastien Bilodeau, Sebastien Gaboury, Bruno Bouchard, and Abdenour Bouzouane. 2014. Human activity recognition in smart homes: Combining passive rfid and load signatures of electrical devices. In *2014 IEEE symposium on intelligent agents (IA)*. IEEE, Orlando, FL, USA, 22–29.

[9] Jon Froehlich, Eric Larson, Sidhant Gupta, Gabe Cohn, Matthew Reynolds, and Shwetak Patel. 2010. Disaggregated end-use energy sensing for the smart grid. *IEEE pervasive computing* 10, 1 (2010), 28–39.

[10] Jingkun Gao, Suman Giri, Emre Can Kara, and Mario Berges. 2014. Plaid: a public dataset of high-resoultion electrical appliance measurements for load identification research: demo abstract. In *proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, New York, NY, USA, 198–199.

[11] Suman Giri, Mario Berges, and Anthony Rowe. 2013. Towards automated appliance recognition using an EMF sensor in NILM platforms. *Advanced Engineering Informatics* 27, 4 (2013), 477–485.

[12] Sidhant Gupta, Matthew S Reynolds, and Shwetak N Patel. 2010. ElectriSense: single-point sensing using EMI for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, Copenhagen, Denmark, 139–148.

[13] Jun Han, Shijia Pan, Manal Kumar Sinha, Hae Young Noh, Pei Zhang, and Patrick Tague. 2017. Sensetribute: smart home occupant identification via fusion across on-object sensing devices. In *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, Delft, Netherlands, 2.

[14] George William Hart. 1992. Nonintrusive appliance load monitoring. *Proc. IEEE* 80, 12 (1992), 1870–1891.

[15] Chih-Wei Hsu and Chih-Jen Lin. 2002. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks* 13, 2 (2002), 415–425.

[16] Input/Output, Inc. 2006. *SM-24 Geophone Element*. Input/Output, Inc. Rev. 3.

[17] Zhenhua Jia, Amelie Bonde, Sugang Li, Chenren Xu, Jingxian Wang, Yanyong Zhang, Richard E Howard, and Pei Zhang. 2017. Monitoring a Person's Heart Rate and Respiratory Rate on a Shared Bed Using Geophones. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. ACM, Delft, Netherlands, 6.

[18] Palanivel Kodeswaran, Ravindranath Kokku, Madhumita Mallick, and Sayandeep Sen. 2016. Demultiplexing activities of daily living in IoT enabled smarthomes. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE, San Francisco, CA, USA, 1–9.

[19] Palanivel A Kodeswaran, Ravi Kokku, Sayandeep Sen, and Mudhakar Srivatsa. 2016. Idea: A system for efficient failure management in smart iot environments. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, Singapore, Singapore, 43–56.

[20] Ravi Kokku. 2017. Activity Signatures in Smart IoT Environments. In *International Conference on Communication Systems and Networks*. Springer, 72–76.

[21] MP Lawton. 2000. Lawton instrumental activities of daily living scale. *The Merck manual of geriatrics* (2000), 45.

[22] Jing Liao, Lina Stankovic, and Vladimir Stankovic. 2014. Detecting household activity patterns from smart meter data. In *2014 International Conference on Intelligent Environments*. IEEE, Shanghai, China, 71–78.

[23] Mostafa Mirshekari, Shijia Pan, Jonathon Fagert, Eve M Schooler, Pei Zhang, and Hae Young Noh. 2018. Occupant localization using footstep-induced structural vibration. *Mechanical Systems and Signal Processing* 112 (2018), 77–97.

[24] Kazuki Moriya, Eri Nakagawa, Manato Fujimoto, Hirohiko Suwa, Yutaka Arakawa, Aki Kimura, Satoko Miki, and Keiichi Yasumoto. 2017. Daily living activity recognition with echonet lite appliances and motion sensors. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, Kona, HI, USA, 437–442.

[25] Shijia Pan, Amelie Bonde, Jie Jing, Lin Zhang, Pei Zhang, and Hae Young Noh. 2014. Boes: building occupancy estimation system using sparse ambient vibration monitoring. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2014*, Vol. 9061. International Society for Optics and Photonics, 90611O.

[26] Shijia Pan, Ceferino Gabriel Ramirez, Mostafa Mirshekari, Jonathon Fagert, Albert Jin Chung, Chih Chi Hu, John Paul Shen, Hae Young Noh, and Pei Zhang. 2017. Surfacevibe: vibration-based tap & swipe tracking on ubiquitous surfaces. In *2017 16th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, Pittsburgh, PA, USA, 197–208.

[27] Shijia Pan, Tong Yu, Mostafa Mirshekari, Jonathon Fagert, Amelie Bonde, Ole J Mengshoel, Hae Young Noh, and Pei Zhang. 2017. Footprintid: Indoor pedestrian identification through ambient structural vibration sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 89.

[28] Oliver Parson, Siddhartha Ghosh, Mark Weal, and Alex Rogers. 2012. Non-intrusive load monitoring using prior models of general appliance types. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*. ACM, Toronto, Ontario, Canada.

[29] Shwetak N Patel, Thomas Robertson, Julie A Kientz, Matthew S Reynolds, and Gregory D Abowd. 2007. At the flick of a switch: Detecting and classifying unique electrical events on the residential power line (nominated for the best paper award). In *International Conference on Ubiquitous Computing*. Springer, 271–288.

[30] Jeffrey D Poston, R Michael Buehrer, and Pablo A Tarazaga. 2017. A framework for occupancy tracking in a building via structural dynamics sensing of footstep vibrations. *Frontiers in built environment* 3 (2017), 65.

[31] Juleen Rodakowski, Elizabeth R Skidmore, Charles F Reynolds III, Mary Amanda Dew, Meryl A Butters, Margo B Holm, Oscar L Lopez, and Joan C Rogers. 2014. Can performance on daily activities discriminate between older adults with normal cognitive function and those with mild cognitive impairment? *Journal of the American Geriatrics Society* 62, 7 (2014), 1347–1352.

[32] Hao Song, Georgios Kalogridis, and Zhong Fan. 2014. Short paper: Time-dependent power load disaggregation with applications to daily activity monitoring. In *2014 IEEE World Forum on Internet of Things (WF-IoT)*. IEEE, Seoul, Korea, 183–184.

[33] Li-Wei Wu, Wei-Liang Chen, Tao-Chun Peng, Sheng-Ta Chiang, Hui-Fang Yang, Yu-Shan Sun, James Yi-Hsin Chan, and Tung-Wei Kao. 2016. All-cause mortality risk in elderly individuals with disabilities: a retrospective observational study. *BMJ open* 6, 9 (2016), e011164.