VibroSense: Recognizing Home Activities by Deep Learning Subtle Vibrations on an Interior Surface of a House from a Single Point Using Laser Doppler Vibrometry

WEI SUN*, Institute of Software Chinese Academy of Sciences, School of Computer Science and Technology, University of Chinese Academy of Sciences, and Cornell University
TUOCHAO CHEN*, Cornell University and Peking University
JIAYI ZHENG*, Cornell University and SUNY-University at Buffalo
ZHENYU LEI, Cornell University and Huazhong University of Science and Technology
LUCY WANG, Cornell University
BENJAMIN STEEPER, Cornell University
PENG HE, Cornell University and Hangzhou dianzi University
MATTHEW DRESSA, Cornell University
FENG TIAN, School of Artificial Intelligence, University of Chinese Academy of Sciences and State Key Laboratory of Computer Science, Institute of Software Chinese Academy of Sciences
CHENG ZHANG, Cornell University

Smart homes of the future are envisioned to have the ability to recognize many types of home activities such as running a washing machine, flushing the toilet, and using a microwave. In this paper, we present a new sensing technology, VibroSense, which is able to recognize 18 different types of activities throughout a house by observing structural vibration patterns on a wall or ceiling using a laser Doppler vibrometer. The received vibration data is processed and sent to a deep neural network which is trained to distinguish between 18 activities. We conducted a system evaluation, where we collected data of 18 home activities in 5 different houses for 2 days in each house. The results demonstrated that our system can recognize 18 home activities with an average accuracy of up to 96.6%. After re-setup of the device on the second day, the average recognition accuracy decreased to 89.4%. We also conducted follow-up experiments, where we evaluated VibroSense under various scenarios to simulate real-world conditions. These included simulating online recognition, differentiating between specific stages of a device's activity, and testing the effects of shifting the laser's position during re-setup. Based on these results, we discuss the opportunities and challenges of applying VibroSense in real-world applications.

*First, second, and third authors contributed equally to the paper.

Authors' addresses: Wei Sun, sunwei2017@iscas.ac.cn, Institute of Software Chinese Academy of Sciences, School of Computer Science and Technology, University of Chinese Academy of Sciences, Cornell University; Tuochao Chen, chentuochao@pku.edu.cn, Cornell University, Peking University; Jiayi Zheng, jzheng35@buffalo.edu, Cornell University, SUNY-University at Buffalo; Zhenyu Lei, leizhenyu1225@gmail.com, Cornell University, Huazhong University of Science and Technology; Lucy Wang, ljw222@cornell.edu, Cornell University; Benjamin Steeper, bds238@cornell.edu, Cornell University; Peng He, ph369@cornell.edu, Cornell University, Hangzhou dianzi University; Matthew Dressa, mtd67@cornell.edu, Cornell University; Feng Tian, tianfeng@iscas.ac.cn, School of Artificial Intelligence, University of Chinese Academy of Sciences, State Key Laboratory of Computer Science, Institute of Software Chinese Academy of Sciences; Cheng Zhang, chengzhang@cornell.edu, Cornell University, 244 Gates Hall, Ithaca, New York, 14853.

2474-9567/2020/9-ART96 $15.00
https://doi.org/10.1145/3411828

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing**; • **Ubiquitous and mobile computing systems and tools**;

Additional Key Words and Phrases: Home activity recognition, Laser Doppler Vibrometry, Deep learning, Structural vibration

ACM Reference Format:
Wei Sun, Tuochao Chen, Jiayi Zheng, Zhenyu Lei, Lucy Wang, Benjamin Steeper, Peng He, Matthew Dressa, Feng Tian, and Cheng Zhang. 2020. VibroSense: Recognizing Home Activities by Deep Learning Subtle Vibrations on an Interior Surface of a House from a Single Point Using Laser Doppler Vibrometry. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 96 (September 2020), 28 pages. https://doi.org/10.1145/3411828

1 INTRODUCTION

Humans conduct a wide variety of activities at home, including interacting with different electrical appliances (e.g., dryers, dishwashers, microwaves, fridges, electric kettles, range hoods, heating systems) and non-electrical objects (e.g., water faucets). Recognizing these home activities has been of great interest to the Ubicomp community for decades, as it can help computers to better understand human behaviors and needs, with the hope of improving human-machine interfaces.

The future vision for smart homes assumes that homes will be able to understand various human activities and respond intelligently in a timely manner. For example, washing clothes in the washer can take up to one hour. In practice, people commonly turn on the washer and then leave their home, or become preoccupied with other tasks. Quite often, people completely forget about the contents in the washer until days later, when the garments must be washed again before drying as they have stayed wet for too long. If a smart home could monitor the washer status and remind residents when necessary, many situations such as this one could be addressed. Existing technologies that monitor home activities can either only sense activities in a small area of the home (e.g., kitchen [29]), or require pre-setup on each individual device or object [69] to monitor. In order to monitor these activities throughout an entire house with several rooms over many floors, multiple devices need to be setup in different locations throughout the house, which can be time consuming and inconvenient.

To address this issue, researchers have developed sensing technologies that instrument the infrastructure of a house to monitor water, electricity, and gas related activities [10, 17, 19]. However, these techniques require a separate sensor for each infrastructure type. For instance, to recognize water, gas, and electricity related activities, three separate sensors must be installed on each pipeline.

Evidently, there is an imminent need for a new sensing technology which is able to recognize multiple activities throughout a home with a simple setup from just one central location. In this paper we present VibroSense, a device which explores the detection of home activities across different rooms and floors of a house, through the modeling of subtle structural vibrations on ceilings and walls. This technology uses a laser Doppler vibrometer (LDV) pointed at a single spot on a wall or ceiling surface to capture the subtle structural vibrations caused by these activities. LDVs are known for their high sensitivity and ability to detect vibrations as small as 0.5 μm/s.

The captured vibration data was used to train a deep neural network to distinguish between up to 18 different types of home activities, which are outlined in Table 1 and 2. The term "activities" refers to the status of various home appliances, units and infrastructure utilities (either electrical or water related). Some activities such as showers, faucets, and microwaves are initiated and controlled by humans, while others such as air conditioners and water heaters may be controlled by a central computing unit.

To evaluate the performance of VibroSense, we collected activity data from 5 houses (averaging 1228 sq ft in size), where we spent 2 days at each house. The results showed that VibroSense can recognize the 18 activities with an average accuracy of over 96.6% when we did same-day training and testing. After re-setup of the device on the second day, the average accuracy across 5 houses dropped to 89.4%. To simulate real-world performance, we collected data continuously for at least 5 hours in each house, where different types of activities and noise (e.g., background noise, street noise, walking) were collected. This collection was designed to simulate the online classification were the system to be deployed in the real world. In the simulated online classification, VibroSense could recognize activities with an average accuracy of 90.99%. Furthermore, we conducted several preliminary experiments to evaluate VibroSense under potential real-world scenarios, including recognizing different stages of an appliance's activity and testing various laser re-setup positions (different locations, angles, and distances between the wall and the laser head). Based on these results, we further discuss the opportunities and challenges of deploying VibroSense in real-world applications.

Although structural vibrations have been used for activity recognition in previous projects [69], no prior work has demonstrated the ability to recognize activities throughout an entire house by measuring structural vibrations from a single point. To the best of our knowledge, to recognize the activities provided in our paper, all prior work necessitates the installation of multiple sensors across a home (e.g., cameras [13], microphones [27], Geophones [41]). VibroSense is the first technology that can detect activities throughout a house (across rooms and floors) by using a single sensing device.

The key research question this paper explores is *whether we can recognize different activities in a home across different rooms by monitoring subtle structural vibrations, which are generated by different activities, and transmitted through the building infrastructure (e.g., walls, ceilings) at a single point.*

The contributions of this paper are:

- We developed a sensing system using a laser Doppler vibrometer and deep learning, which demonstrated that subtle structural vibrations captured from a single point can recognize 18 activities throughout a house with an accuracy of up to 96.6%.
- We evaluated VibroSense in a 2-day experiment for each of the 5 houses, which collected 58 hours of data on over 4936 total activities. The results revealed that VibroSense can recognize 18 activities with an average cross validation evaluation accuracy of 96.6%, and an accuracy of 89.4% after the re-setup of the equipment on the second day.
- We conducted several follow-up experiments to evaluate VibroSense under different scenarios to simulate real-world conditions, including simulated online recognition, differentiation of unique statuses of a device, and the testing of different variations of re-setup.
- Based on the results of this experiment, we discussed the opportunities and challenges of applying VibroSense in real-world applications.

2 RELATED WORK

In-home activities can generate a variety of signals (e.g., sounds, structural vibrations, electromagnetic fields), with each activity presenting a unique signal pattern. VibroSense recognizes activities across the house by learning the signatures of subtle structural vibrations generated by in-home activities. These subtle structural vibrations are captured using a laser Doppler vibrometer, which points at a single point on an internal surface of the building (i.e. wall or ceiling). In this section, we define related work at large as the prior projects that recognize in-home activities by instrumenting sensing devices in the environment to capture and analyze signals generated by home activities, with a focus on those that detect structural vibrations.

2.1 Indoor Activity Recognition Using Non-Contact Sensing

In order to recognize indoor activities, many previous projects have explored using non-contact sensing methods, where the sensing device does not need to directly touch the object of interest. Computer vision is one of the most popular non-contact sensing methods for indoor activity recognition, which has been widely explored under different scenarios [28, 34, 46, 56, 68, 72]. This camera-based method needs to directly capture the object of interest without any occlusion (direct line of sight view), which is not always feasible in real-world settings.

Furthermore, the requirement of directly seeing the objects also limits the range that this method can be effective. For instance, in order to monitor activities in a house, multiple cameras are required to be installed in different locations.

Wireless sensing is another popular non-contact sensing method for activity recognition, which does not require the sensing device to touch nor see the object of interest. This type of method usually requires the sensing device to broadcast wireless signals in the house. Then the object of interest would reflect wireless signals back to the sensing source, and the received signals would be analyzed to detect the corresponding activities of the object. A variety of wireless signals (e.g., WIFI [31, 32, 39, 45, 62–64], electromagnetic interference signals [9, 19, 71], RFID [30, 52, 53], radio waves [6, 51]) have been used to detect different applications, including indoor localization, sleep monitoring [32], status of appliances [9, 19, 39], and motion detection [11, 30, 45, 52, 71]. These wireless sensing methods are usually good at detecting larger scale human activities, which are more effective in reflecting wireless signals. However, they are less sensitive in recognizing activities that cause subtle vibrations in the house (e.g., water dropping in the basin).

2.2 Indoor Activity Recognition Using Contact Sensing

In contrast to non-contact sensing methods, which indirectly observe indoor activities from reflected signals, a significant amount of previous work recognizes indoor activities by directly instrumenting various sensing devices such as accelerometers [58, 59], magnetic sensors [25], or a set of sensors [65] directly onto or near the source objects themselves.

These types of methods require installing sensors on each object of interest and communicating the sensor data to a processing computer for recognition, where deployment of the system necessitates significant efforts. For instance, Pan [41] used a vibration sensor and an electrical sensor to recognize fine-grained activities in the kitchen. To deploy this method in different rooms of a house, multiple sets of hardware would need to be instrumented. Therefore, to reduce the workload of excessive instrumentation and provide the ability to monitor whole-house activities, researchers developed another set of technologies, also known as Infrastructure Mediated Sensing (IMS). The idea of IMS is to instrument one sensor instead of multiple sensors onto a piece of infrastructure in a house (e.g., water pipeline), which can then be used to monitor related activities throughout the entire house. This method has been proven successful in monitoring various house activities in the areas of electricity [9, 19, 44], water [16, 17], HVAC [43], and gas [10]. However, IMS requires different sensors for different types of activities. For instance, to recognize water, gas, and electricity based activities, three separate sensors must be installed on each pipeline. There is a clear need for a simple sensing technology with an easy setup that is able to monitor various types of activities across an entire house.

2.3 Indoor Activity Recognition Using Vibrations

VibroSense uses structural vibrations to recognize home activities. This concept is not entirely new, as several previous works have explored using vibrations of the source object to recognize home activities. Davis, et al. [12, 13] used high-speed cameras to reconstruct the status of an object (e.g., sound) by decoding its vibration from captured videos. VibroSight recognized the status of an object by directly shooting a laser beam at the object to capture vibration data [69].

As previously stated, vibrations from objects associated with various indoor activities can travel through the infrastructure of a house causing structural vibrations. Structural vibrations can in turn be highly informative for distinguishing between different activities. Many prior research projects developed technologies to recognize activities from these structural vibrations. GeoPhone is a sensing device that is deployed in various locations around a house to recognize human activities by capturing vibrations on surfaces. Applications include identifying the quantity [15], identity [42], positions [36, 37, 49, 50], gait [35], activity levels [70], and working habits [7] of people in the house. For instance, GeoPhone has been instrumented around a washbasin to monitor water-related activities (e.g., hand washing) [14]. Besides Geophone, researchers have also explored using microphones spread out in different locations in a house to capture vibration data for activity recognition (e.g., kitchen [27, 66], washbasin [14, 16, 26]).

Although structural vibrations have been used in the past to recognize home activities, they are limited in that they either 1) only recognize human activities from floor vibrations, or 2) need to deploy the sensors (e.g., Geophone, microphone) directly on or next to the source object itself for activity recognition. To the best of our knowledge, we have not found any prior work that demonstrates the use of a single sensor capable of recognizing activities across rooms in an entire house by sensing structural vibrations from a single point.

2.4 Applications Using Laser Doppler Vibrometer

Laser Doppler vibrometry (LDV) offers a non-contact method to capture subtle vibrations on a surface (i.e. nano-meter level vibrations), which are challenging to capture using other traditional vibration sensing devices, such as an accelerator. LDVs use the frequency shift generated by the Doppler effect related to vibration velocity, which can be used to calculate the velocity of a surface demodulation with higher accuracy and sensitivity [8, 48]. LDVs have been widely applied in many fields, including structural condition monitoring [18, 40, 55], MEMS fabrication [61], biomedical imaging [57], otology[4, 22, 23, 38], and acoustics [2, 5, 47, 73]. Due to their high sensitivity to subtle vibrations, LDVs have also been used to reconstruct audio information from a distance by pointing a laser beam at an object to capture vibrations on its surface [60, 67]. A similar idea has also been implemented in Vibrosight [69], which shoots a laser beam directly at various appliances to detect their statuses. However, this method necessitates a clear view between the sensing device and the object, which limits its range of operation. Unlike Vibrosight, VibroSense does not require a separate laser beam for each device of interest. Instead, VibroSense indirectly detects multiple activities throughout a home by measuring structural vibrations with just a single LDV.

2.5 Summary

Compared to prior work, VibroSense is the first non-contact sensing technology that is able to recognize multiple activities (e.g., water, electrical appliances) across rooms in a house by capturing structural vibrations from a single point.

3 THEORY OF OPERATION

This paper is centered around the principle that most activities performed in a home will generate vibrations which will then propagate throughout the home's infrastructure (e.g., floors, walls, ceilings). Previous work [3] has shown that the mechanical wave propagation in solids (like walls) is influenced by the structure and material of its travel path. Hence, the vibration propagation in a building can be approximated into a physics model:

$$Y(f) = H(f)X(f) + N(f) \tag{1}$$

$X(f)$ is the source vibration signal caused by indoor activities, and $Y(f)$ is the vibration signal received at the point of detection on the wall, ceiling, or floor. $H(f)$ represents the frequency response, which is determined by the structure and material of the building and the vibrations' travel path. $N(f)$ represents background noise. Based on the above vibration propagation model, we will explain how a building's structural vibrations can be utilized to recognize indoor activities.

Firstly, structural vibrations will carry with them unique signature features belonging to the sources creating them. This means that different activity sources generate different signal patterns $X(f)$.

Secondly, the travel path from the vibration source to the point of detection differs depending on the location of the source object. Each path serves as its own unique band-pass filter which alters the frequency response of the signals generated by the activity source. Different travel pathway materials and structures result in different frequency responses to vibrations $H(f)$. Therefore, the travel path (band-pass filter with unknown parameters) together with the activity source (signal source) construct unique signatures in the received signals, which can then be extracted and used to differentiate between activities.

Furthermore, due to the high accuracy and sensitivity of LDVs, we are able to sense these micro vibrations from a single point on the surface of a house. These vibrations will propagate along the home infrastructure, such as walls and ceilings. Admittedly, the vibrations will attenuate as the distance increases. However, the LDV we used has a very high velocity resolution of 0.5 μm/s, because it uses interferometry to extract the frequency shift caused by the Doppler effect. Therefore, it is possible to utilize this property to recognize many activities performed throughout the home from a single probe point.

To verify our above hypothesis, we conducted a series of preliminary experiments. As shown in Fig. 1, we used a laser Doppler vibrometer (LDV) to capture the structural vibrations from a house's interior surface for 18 different types of indoor activities (such as microwaves, heaters, kitchen faucets, toilet faucets, etc) and visualized their spectrums. Fig. 1 shows that different kinds of activities have visually distinct patterns in frequency domains. Fig. 1 subfigures 10 and 11 show that the same activity (faucet) in different locations (bathroom vs kitchen) present visually distinct patterns.

These preliminary experiments were very encouraging, which motivated us to build the system, which we will detail in the next section.

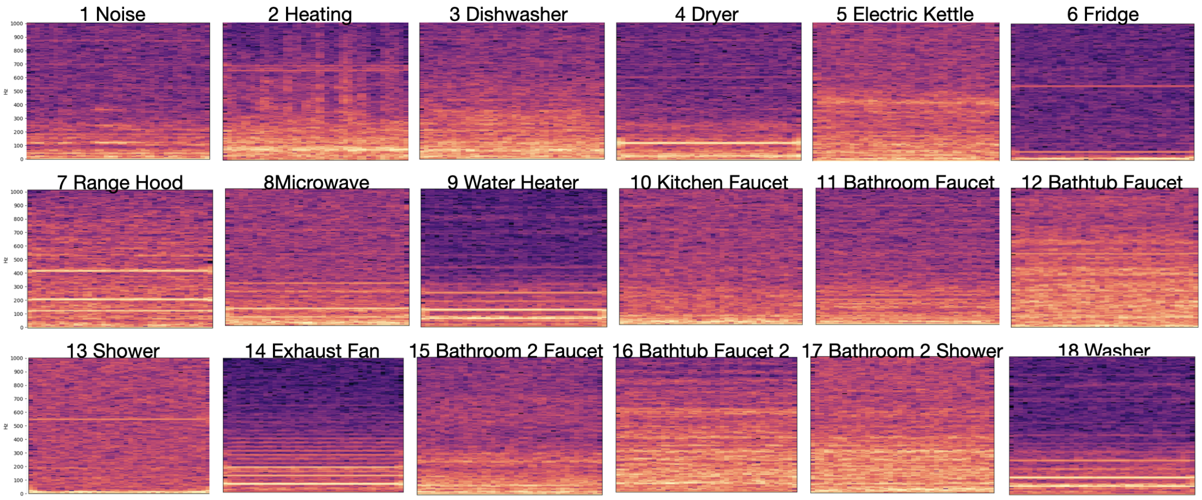


Fig. 1. Waveform and Spectrum (0 - 1.5K Hz) of different activities

4 VIBROSENSE SYSTEM

In this section, we describe the hardware components, data collection system, and data processing pipeline (including pre-processing methods and deep learning algorithms) of the VibroSense system.

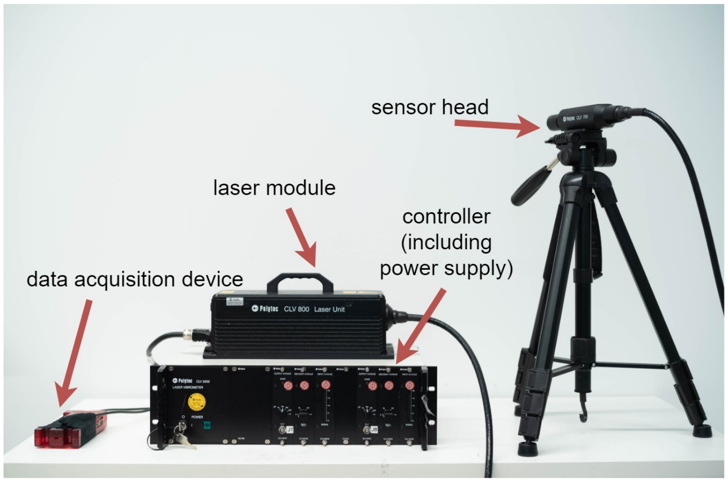


Fig. 2. The overview of the data collection system

4.1 Hardware and Data Collection System

To capture subtle vibrations of electrical or water appliances from a single point in the house, we need a sensing device that is extremely sensitive to subtle vibrations. Therefore, we decided to use a laser Doppler vibrometer as the sensing device, which is known for its high sensitivity to subtle vibrations.

Our hardware data collection system consists of a commercial LDV from Polytec, a LabJack T7 Pro (16-bit ADC), and a tripod, as shown in Figure 2. The LDV includes a CLV-700 sensor head, a laser module, and a controller. The focus (VF25) of the sensor head is variable and its stand-off ranges from 0.11 to 10m. The sensor head is connected to the laser module (He-Ne), which emits the laser beam with a central wavelength of 632.8 nm. The controller provides the power supply for the optic components, and implements signal processing including decoding, signal conditioning, and low-pass filtering (f_{cutoff} = 5 kHz). An analog output ranging from -10 v to +10 v is generated from the controller, which is proportional to the instantaneous velocity of the detecting point. The controller can detect vibration velocities of up to 1.25 m/s with a resolution of 0.5 μm/s. The LabJack is used to transmit the output data from the LDV to a laptop MacBook Pro (13-inch, 2017) with a sample rate of 10K. Finally, a laptop is running a python program to store the data set for later processing.

4.2 Data Processing Pipeline

As shown in Fig. 3, the VibroSense data processing pipeline can be divided into two parts: pre-processing and deep learning. The process of our system works as follows. First, the laser module captures the vibration velocity on the surface (wall or ceiling) and digitizes the vibration signal. Then, the digital signal is windowed, pre-processed, and sent to generate the feature vectors. Finally, these feature vectors are input into a customized, complex neural network to recognize the pre-defined home activities.

4.2.1 Data Preprocessing. To process the continuous data stream from the LDV, we applied a sliding window of 2 seconds with a 50% overlap. We chose a 2 second window after considering the trade-offs of using different window lengths. On one hand, the length of the sliding window determines the interval of two contiguous events that our system can detect (50% overlap is used in our system). So the smaller the window length we choose, the higher the time resolution (frequency) for detecting activities. On the other hand, according to the Fourier Transformation, the longer the sliding window length we select, the higher the resolution in the frequency

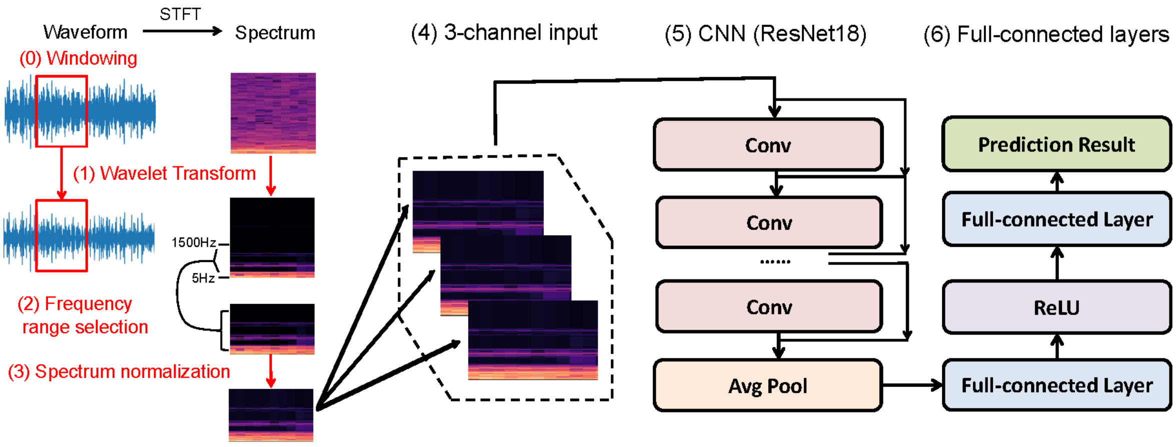


Fig. 3. Data Processing Pipeline

domain, which may be more informative. After considering these two factors, practical situations, previous research [69], , and conducting pilot experiments, we settled on a 2 second window length and a 50% overlap. Since the sampling rate of our LDV is $10KHz$, each small signal frame contains 20000 data points. For each 2 second window, we conducted the following data processing steps, including wavelet transformation, Shot-time Fourier Transformation (STFT), and Normalization.

Wavelet Denoising According to our physics model $Y(f)$ for vibration, only the part, $H(f)X(f)$ can provide valid information for event recognition. Therefore, for better and robust recognition, we tried to reduce the interference of $N(f)$ by introducing the Wavelet Transformation to denoise the original signal. However, in real-world implementation, denoising is not a simple task. Firstly, the background noise and the event signal have an aliasing in the low-frequency range. Hence, the traditional filtering method (such as using a Gaussian filter) cannot directly subtract the noise from the event signal. Secondly, since the agents causing background noise (like human walking, traffic outside, and so on) are complex and vary over time, the method of directly subtracting the 'noise floor' cannot work either. Therefore, we decided to apply wavelet denoising using the PyYAWT library to filter out background noise. Specifically, we first apply the discrete wavelet transform (DWT) to decompose the original signal (decomposition level=5). Then we shrunk the coefficients in the wavelet domain using the "heursure" rule and soft thresholding [33]. Finally, we used the selected coefficients to reconstruct the signal. Fig. 3 displays the changes in temporal domain and frequency domain after wavelet transformation.

Feature Extraction with Short-Time Fourier Transform To acquire the information in the frequency domain, we apply the Short-Time Fourier Transform (STFT) (with win_length = 4096 and hop_length = 512) on the filtered signal and acquire a two-dimensional spectrum (1025x40). The y-axis (1025) of the spectrum is frequency and the x-axis (40) is the time line. As the spectrum displays in Fig. 3 step 2, we only selected the frequency range from 10Hz to 1500Hz. This is because the interference of background noise is mostly below 10 Hz, and the signal above 1500 HZ is not informative as shown in the figure above.

Spectrum Normalization In real-world implementation, it is impossible to maintain the exact position of the LDV. Different incident angles and distances of the LDV from the wall would influence the amplitude of the received signal. Hence, the actual signal received by the LDV can be approximately regarded as the scaling of the $Y(f)$. To reduce the influence of this scaling problem, we normalized the amplitude of the spectrum. More specifically, we first calculated the average amplitude for each window, and then divided the 2D spectrum of each window by its average value as shown in Fig. 3 step 3.

4.2.2 Deep Learning. In order to decide which machine learning algorithms we should use for our system, we compared performances with traditional machine learning methods (supporter vector machine [SVM]) and deep learning models (convolutional neural network [CNN]). We extracted similar frequency domain features (STFT, FFT) in both models. The results demonstrated that the performance for both models are encouraging (over 85%) but CNN outperforms SVM. More importantly, CNN seems to generalize well on different houses. In other words, the variance of model performance is smaller with CNN than with SVM. Furthermore, the input is a 2-dimensional spectrum image, and a convolutional neural network (CNN) is popularly used to handle image-related tasks. As figure 1 shows, the temporal information in each event is limited, as the frequency response does not vary much within each sample. Therefore, we decided to use CNN over a recurrent neural network (RNN) after comparing deep learning models.

Network Architecture Among the many types of CNNs, the 18-layer residual network [20] (ResNet-18) has been proven to be highly effective for visual recognition tasks and less prone to over-fitting. (We tested both VGG and ResNet in our system, and the results showed that the classification accuracy of ResNet is around 2% higher than VGG). The structure of our deep learning model is shown in Fig. 3. A block in ResNet includes several convolutional operations, each followed by batch normalization [24] and rectified linear unit (ReLU). A global average pooling is connected to the end of the ResNet18 to extract a vector representation of the spectrum. Then, the extracted vector is input into two fully connected layers with ReLU in between them and a dropout layer [54] ($p = 0.5$). The final full connected layer outputs the prediction result.

Data Formatting and Training Process Since the ResNet18 requires 3-channel (RGB) 2D images for input [20], we copied the spectrum after preprocessing twice and then combined the original spectrum and its 2 copies together to construct the 3-channel data matrix as the input for the deep learning model. The output of our deep learning model is a one-hot vector for classification result. The training parameter is based on common practices in previous CNN research [21]: batch size 30, learning rate $1e-4$, weight decay $1e-4$ and training epochs 40. We selected cross entropy (CE) as the loss function and the adaptive moment estimation (Adam) algorithm as the training optimizer.

5 DATA COLLECTION

To evaluate the performance of VibroSense in real houses, we collected data using the hardware set (LDV + Labjack) mentioned in the previous section to collect structural vibration data for 18 activities in five different houses. The 18 activities are associated with electrical or water appliances as detailed in figure 1. In this section, we outline the details for data collection, including information about each house as well as the study procedure.

5.1 House Information

To understand how VibroSense would perform in different homes, we deployed the system in five different houses. Each house was located in a different location, and varied in size, structure, and home appliances. Four houses (H1, H2, H3, H4) were rented through Airbnb. We acquired approval from the home owners to conduct our experiments. H5 was one of the researchers' home. H2, H3, and H5 were houses with two floors, and H1 and H4 were houses with a single floor. The sizes of the homes ranged from 850 square feet to 1400 square feet. The floor plans of H1 to H5 were 2B1B, 2B1.5B, 3B1B, 4B1.5B, and 3B3.5B respectively. Among these 5 houses, only H5 was a townhouse. The other 4 houses were all single family homes. We intentionally chose houses with different floor plans (number of floors, size, layouts) and settings (e.g., appliances) to maximize variance in the evaluation data sets. Figure 4 shows the detailed floor plan of all five houses. For each individual house, we all collected activity 1 and 2, which were distributed throughout the house.

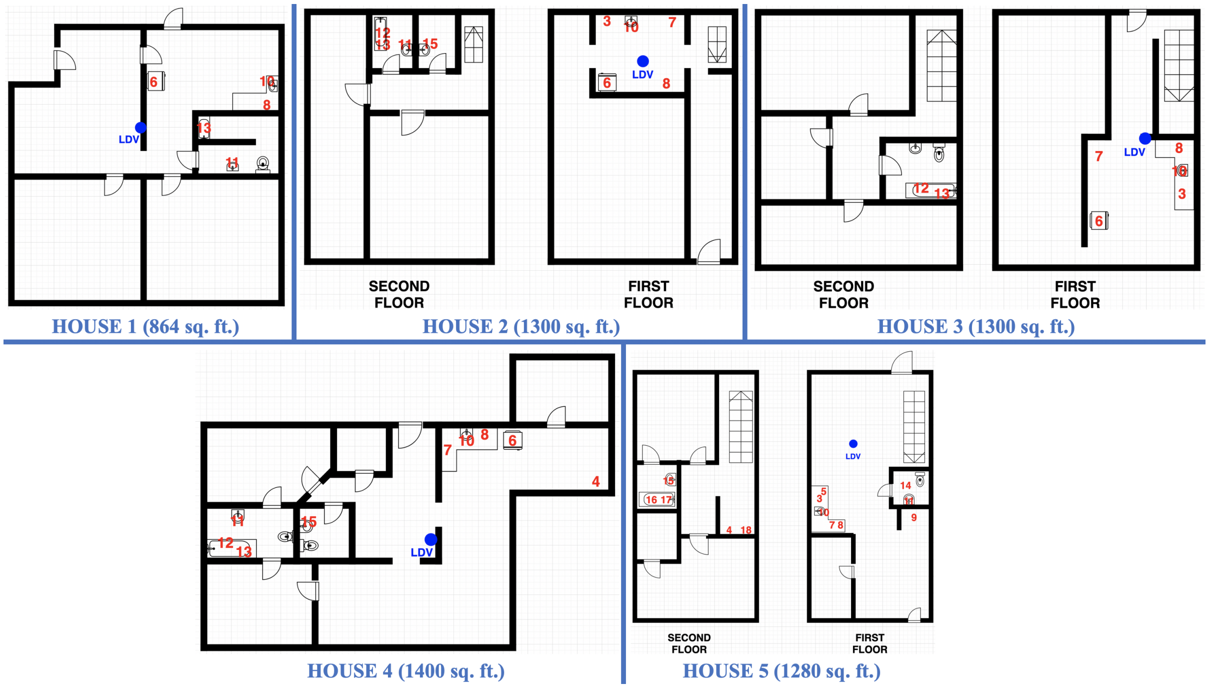


Fig. 4. The floor plan, locations of activities, and location of LDV in houses 1-5

In the preliminary experiment, we deployed our system in apartments and it worked adequately. However, an apartment usually shares floors, ceilings, and walls with other neighboring apartments. Consequently, the high sensitivity of the vibration capturing process causes our system to include activities from other apartments which are not our targets. On one hand, this raises significant privacy concerns, as our system is extremely sensitive to subtle vibrations. For instance, previous studies have shown that an LDV can be used to reconstruct voices from vibrations on glass. On the other hand, data collection in an apartment can introduce complex noise into our data set. For example, the structural vibrations generated by washers in other apartments can be captured by our system. We cannot possibly verify the ground truth for all of these activities. Therefore, we decided to exclude apartments from our current experiment, with the intent of exploring them in future studies.

5.2 Data Collection Procedure

In this section, we present the procedure for our data collection experiment. In total, we collected data for 18 different types of activities from common electrical appliances (e.g., dishwashers and dryers) and water appliances (e.g., faucets and toilets) in five houses. However, not every home contained all the appliances needed for all 18 activities. Therefore, we collected a subset of the 18 activities in each house corresponding to the appliances located in that particular dwelling. In the end, we collected data for 7 activities in H1, 11 activities in H2, 9 activities in H3, 11 activities in H4 and 15 activities in H5. Figure 4 details which activities we collected in each house.

5.2.1 Device Setup. At the beginning of each experiment within each house, the first step was to set up the system. Based on our pilot study results, we developed some guidelines to help us choose an optimal location to set up the device. First, we placed the device at a central location in the house, as shown in figure 4. Since we intended to detect activities which are distributed over various locations in the house, placing the device near the center of the house would minimize the average distance between the device and the appliance of interest. The greater the distance, the more attenuation we may observe in the received signal. We pointed the laser head at the ceiling or a wall depending on the house layout. If the house had two floors, we pointed the laser at the ceiling. If the house only had one floor, we pointed the laser at the interior wall of the building. The laser head was attached on a tripod, which was placed 30-50 cm from the wall or the ceiling. The angle between the laser beam and the surface was approximately 90 degrees. As figure 4 shows, following this rule, the laser was pointed at an interior wall in H1 and H4, and at the ceiling in H2, H3, and H5.

After deciding the setup location, we connected different components of the data acquisition system and fixed the sensor head on the tripod. To enhance the reflected light signal, we attached a piece of kinesiology tape along with a layer of retro-reflective tape on top of the surface at which the laser would be pointing. We borrowed this common practice from prior work [69]. After attaching the retro-reflective tape, we pointed the laser beam at the tape and performed minor adjustments to the laser's position to optimize reflected signal strength. Finally, we initiated a computer program in Python, which ran on a separate laptop to collect the data.

We admit that different device locations and setup settings (angles and distances between the laser head and surface) may influence the received signals and results. We conducted a follow-up study to explore the influence of different levels of re-setup on the results, which is detailed in the discussion section.

5.2.2 *Experiment Procedure.*

Two researchers collaborated in collecting data at each house. One researcher (R1) was responsible for starting and ending data collection on the laptop in addition to labeling the ground-truth. The other researcher (R2) performed daily activities by turning on or off appliances around the house. For each sample of each activity, R2 kept the device on for at least 15 seconds. Throughout this process, both researchers were in constant communication to ensure that the data being collected was in sync with the status of the device (on/off), for the 15 second window. During data collection, we intentionally abstained from turning on any other activities of interest while one activity was being collected. We did, however, allow random background noise to be mixed in with the recorded data. Additionally, we randomized the sequence of different activities during data collection.

We opted to choose a window of 15 seconds since our chosen daily activities usually took roughly this amount of time to complete (e.g., hand washing). We also referenced prior work that recognized similar activities [17, 43, 44]. We recorded every activity for 15 seconds for consistency. Data collection at each house lasted for 2 days.

Discrete Samples Collection on Day 1 On the first day of data collection, the researcher repeated activities of noise (including passive background noise, human voices, and footsteps), electrical appliances, and water appliances (Table 1 and 2) in a random order around 30 times (30 samples), where each sample was saved as a 15 second data segment. We decided to record 30 samples for each activity due to two major considerations. On the one hand, the more samples we collect, the more likely our performance would be higher. However, this would also increase workload for users in the future. For this reason, we needed to find a balance between data collection effort and performance. Based on our preliminary data set, we did find that as the size of training samples increased, the accuracy gradually increased as well. However, the curve flattened at around 20-30 samples (detailed follow-up experiment in section 7.3). Therefore, we decided to collect 30 samples for each activity, to be divided for training and testing in cross-validation. We will present the evaluation protocol in later sections.

Cross-day re-setup Experiment on Day 2 One challenge VibroSense may face if being deployed in a real-world setting is that background noise may vary depending on the time of day or the day of the week. Furthermore, external environmental factors may nudge the device and alter the received signal (e.g., human interference, temperature of the device and room). In order to test how VibroSense would perform in these scenarios, we re-setup VibroSense on the second day before collecting data. This involved turning off the entire system and then slightly readjusting the angle and distance between the laser head and surface (ceiling or wall). After re-setup, the laser beam was still pointing at the retro-reflective tape on the surface, but in a slightly different location. This re-setup was designed to simulate a real-world scenario where the device could be accidentally bumped and reset. We call this a "non-significant setting change" when the device's position or angle is slightly altered while still pointing at the same retro-reflective tape. After re-setup, we collected around 10 samples for each activity, in accordance with the procedure in Day 1. This collected data was used to test the model which was trained using the data collected from day 1 before re-setup.

In addition, we understand that re-setup can significantly influence the received signal. Therefore, we did follow-up experiments to explore the influence of different levels of setup on the received signal and the corresponding model performance. These details will be discussed in the discussion section 7.1.

Data Collection for Simulated Online Testing In the previous data collection sections, we collected discrete activity samples, with each sample lasting 15 seconds. Using these discrete data segments, we were able to verify the feasibility of our theory of operation. However, this method would present issues if implemented in real time in a real home. In a real home, the system will be primarily dealing with background noise most of the time. In other words, the majority of the data stream would be background noise, and activities of interest would randomly appear in the data stream. Furthermore, a real-world system would need to handle a continuous data stream instead of data segments of 15 seconds. Therefore, understanding how the system would perform under these scenarios is critical to estimating the realistic potential performance of VibroSense.

To simulate this scenario, we collected data continuously for 5 hours in each home on both day 1 and day 2. For each data collection session, we first started the data collection system and then randomly chose an appliance to turn on for a random duration of at least 10 seconds. We then turned off that appliance before turning on the next appliance. We also collected a random number of samples for each activity ranging from 1 sample to 29 samples. Table 1 and Table 2 present details on the sample sizes collected at each house for this experiment. The process was recorded by a GoPro for ground-truth labeling. During this process, background noise (including researchers walking and talking around the house) were recorded with the continuous data. These continuous data recordings were then used for a simulated online testing experiment, which will be detailed in section 6.3.

Table 1. Single Activity Data Collection 1/2

Activity Number		House 1			House 2			House 3	
	Day 1 (Samples)	Day 2 (Samples)	Online Testing (Samples)	Day 1 (Samples)	Day 2 (Samples)	Online Testing (Samples)	Day 1 (Samples)	Day 2 (Samples)	Online Testing (Samples)
1 Noise	294	146		294	146		280	140	
2 Heating	30		1	30	12	1	26	10	2
3 Dishwasher				31	11	22	32	10	2
4 Dryer									
5 Electric Kettle									
6 Fridge	33	11	18	32	10	1	33	10	3
7 Range Hood				31	10	22	35	10	5
8 Microwave	30	10	18	32	10	18	31	10	13
9 Water Heater									
10 Kitchen Faucet	31	10	19	30	10	19	31	10	12
11 Bathroom 1 Faucet	30	7	9	33	10	9			
12 Bathroom 1 Bathtub Faucet				33	10	12	30	10	15
13 Bathroom 1 Shower	30	9	9	33	10	7	30	10	19
14 Bathroom 1 Exhaust Fan with Light									
15 Bathroom 2 Faucet				30	11	3			
16 Bathroom 2 Bathtub Faucet									
17 Bathroom 2 Shower									
18 Washer									
Activities (in total)	478	193	74	609	250	114	528	220	71

Table 2. Single Activity Data Collection 2/2

Activity Number	House 4			House 5			Sample Size of Five Houses (in total)
	Day 1 (Samples)	Day 2 (Samples)	Online Testing (Samples)	Day 1 (Samples)	Day 2 (Samples)	Online Testing (Samples)	
1 Noise	420	126		470	168		2484
2 Heating	30	10	2	31	10	2	197
3 Dishwasher				30	10		148
4 Dryer	29	10	4	30	10	5	88
5 Electric Kettle				35	10	5	50
6 Fridge	32	11	7				201
7 Range Hood	30	10	8	30	10	25	226
8 Microwave	30	10	8	30	10	29	289
9 Water Heater				38	13		51
10 Kitchen Faucet	30	10	7	30	10	23	282
11 Bathroom 1 Faucet	30	10	5	30	10	13	196
12 Bathroom 1 Bathtub Faucet	31	10	9				160
13 Bathroom 1 Shower	34	10	7				208
14 Bathroom 1 Exhaust Fan with Light				31	10	19	60
15 Bathroom 2 Faucet	30	10	10	30	10	20	154
16 Bathroom 2 Bathtub Faucet				30	10	13	53
17 Bathroom 2 Shower				30	10	7	47
18 Washer				30	12		42
Activities (in total)	726	227	67	905	313	161	4936

6 EVALUATION RESULTS

In this section, we present the results of the VibroSense System. The evaluation of the system was divided into three parts: discrete same-day testing, discrete cross-day testing and simulated online testing. The same-day testing evaluated the system's ability to identify home activities after setting up the laser once, using 10-fold cross validation. The cross-day testing aimed to recognize the same home activities on the next day after resetting up the laser (altering its angle and distance from the wall). Lastly, we tested the system's ability to continuously track home activities over a 5-hour long period at each house, simulating real-time online applications. All x-axis and y-axis values of the confusion matrices in this section correspond to various activities, as outlined in Figure 5, 6, 7, and 9.

6.1 Discrete Same-Day Testing

For discrete same-day testing, the data sets collected on the first day with sample lengths of 15 seconds were used for training and testing. As stated in section 5, on the first day of each house, we collected around 30 samples for each activity using a sample length of 15 seconds. All data was collected after setting up the laser just once (same incident angle and distance for all samples). We then implemented a 10-fold cross-validation by dividing the samples into 10 groups based on chronological order. We then selected one group for testing and used the remaining 9 groups for training. After dividing samples into training and testing sets, we applied our data processing pipeline (detailed in section 4.2) to process these sets for training and testing the CNN models. One thing worth mentioning is that there was a discrepancy between the 5 second and the 2 second window needed for the deep learning model. To address this issue (as described in section 4.2) we applied a sliding window of 2 seconds with a 50% overlap on each activity sample of 15 seconds to generate the actual training and testing instances of 2 seconds. The 2-second instances generated from one sample will only be used for training or testing. In other words, we did not use one part of the 2-second instances from the same sample for training and the other part for testing. If one 2-second instance generated from a sample of 15 seconds was used for training, then all 2-second instances generated by this sample were used for training.

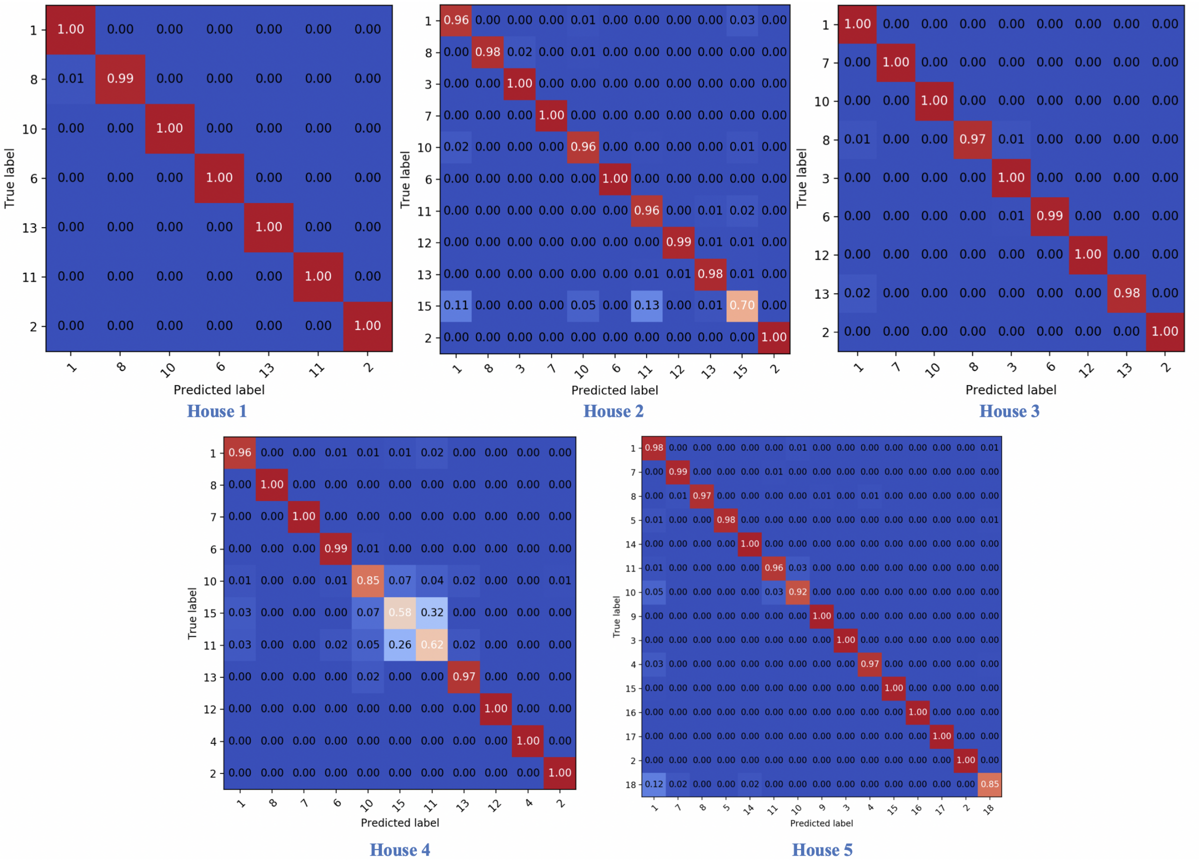


Fig. 5. Confusion matrices for 5 houses on discrete same-day testing

Furthermore, we calculated the recognition accuracy and repeated the above process 9 more times until each group had functioned as the testing group. The final accuracy for each house was calculated by taking the mean of all 10 accuracies. The final recognition accuracies of the 5 houses for discrete same-day testing were 99.67%, 95.9%, 99.42%, 90.52%, and 97.47%, respectively. The average recognition accuracy of the five houses was 96.6%, verifying that our system demonstrated feasibility in identifying house appliances, when setting standards were left unchanged. House 4 presented the lowest accuracy among these five houses. House 4 is a single floor house with the largest size, meaning that appliances are spread out and quite distant from the sensor. This could have led to a reduction in performance as longer propagation distances may lead to lower signal noise ratios (SNR), which influences recognition performance.

The confusion matrix for each house is presented in Figure 5. By observing each confusion matrix, we found that the system performed very well when recognizing electrical appliances. However, the system sometimes confused water appliances with background noise. For example, the system struggled to detect the faucet located on the second floor in bathroom 2 of house 2 (activity 15). In house 4, there was also confusion between the water appliances (e.g., activities 10, 11, 15). Activities 10, 11, and 15 refer to the Kitchen Faucet, Bathroom 1 Faucet, and Bathroom 2 Faucet, respectively. As Figure 4 shows, these faucets were scattered around the house. We found that some water appliances were similar to noise and similar to each other regarding their waveforms, frequencies, and amplitudes. We discovered this when analyzing FFT frequency vs. amplitude results in section 6.2.

Based on our observations, water-related activities could have been more difficult to be distinguished for two reasons. Firstly, the vibrations generated by water appliances were relatively weaker than electrical appliances. These signals were further attenuated after propagating through the building infrastructure. By the time the waveform reached the sensor, the signal-to-noise ratio (SNR) was relatively low. When SNR is low, background noise may play a more important role in making the classification decision. Secondly, water pressure levels may have varied across different samples, which may have led to a variance in frequency response as well as strength of the vibration. However, these factors are only our suspicions after observing the data patterns. A more thorough experiment is needed to draw a data-informed conclusion in the future.

Another observation we made was that we only collected activity 18 (washer) in house 5, as no other houses were equipped with this appliance. The accuracy of activity 18 was the lowest in house 5. One possible explanation for this, is that a washing machine has multiple stages of operation. We did not control for washing cycle stages while recording data. Different stages of the washer may have generated different vibration patterns which could have confused our system and reduced accuracy. We will further explore the recognition of multiple stages of an appliance in our discussion section.

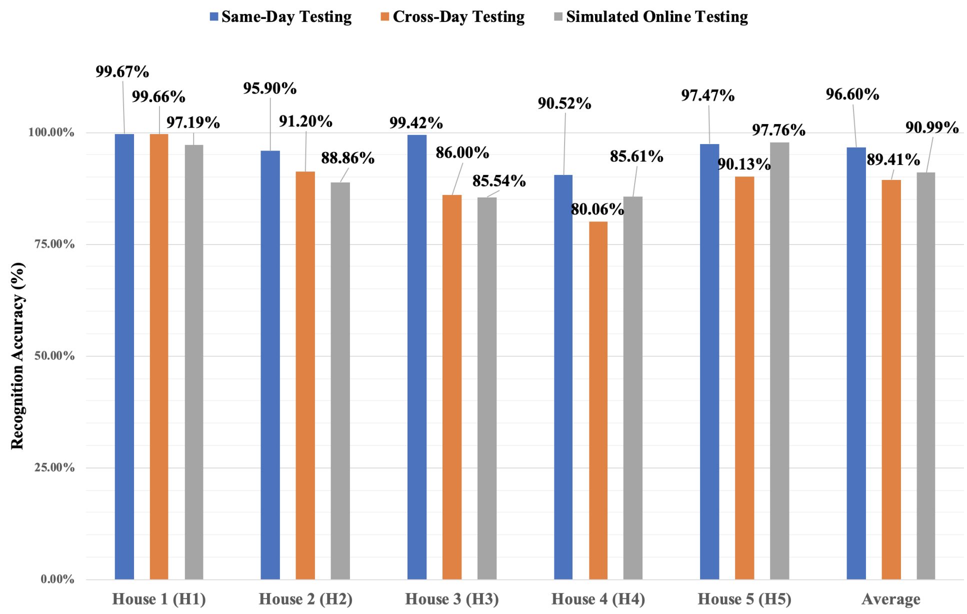
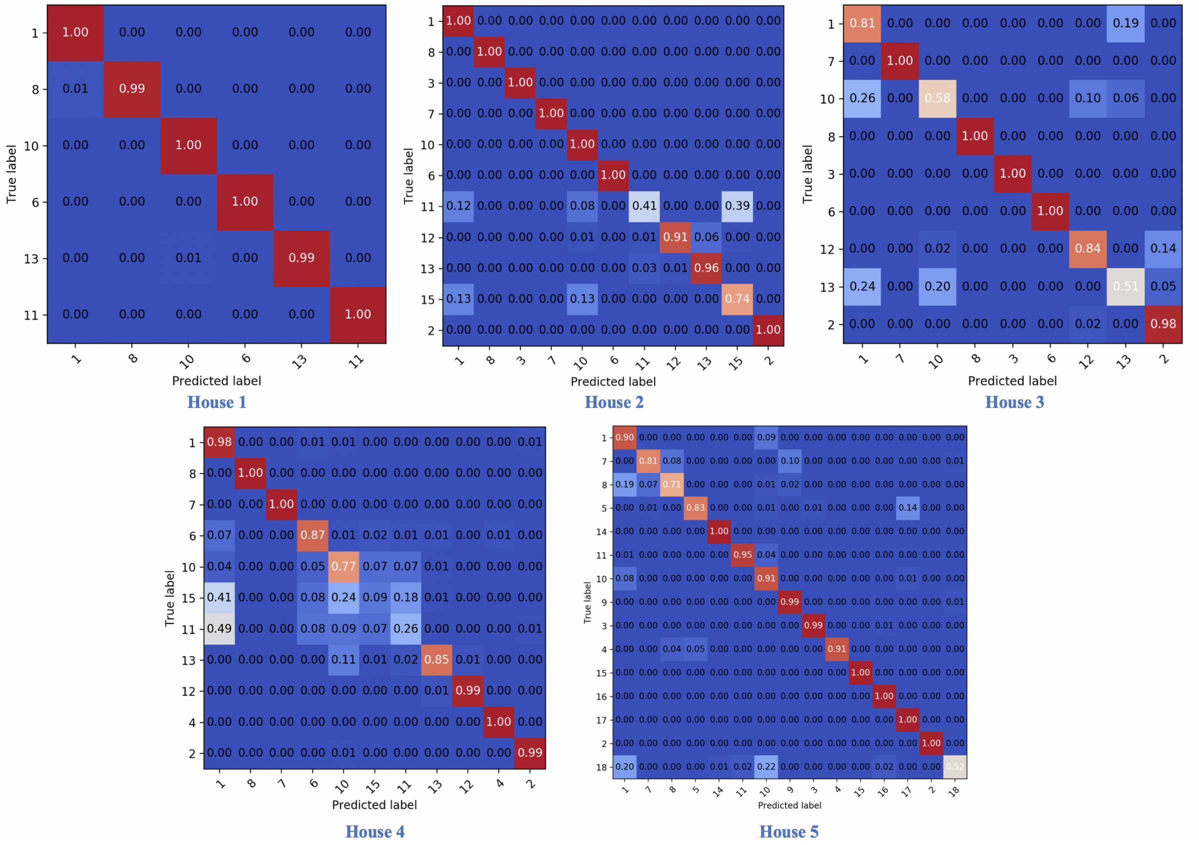


Fig. 6. Summary of experiment results on each house including discrete day 1 testing, discrete day 2 testing, and simulated online testing

6.2 Cross-Day Re-setup Experiment

In real world scenarios, it's likely that the laser head's position and angle could be altered - whether it's bumped by accident or re-setup for other reasons (e.g., power outage). Thus, to further evaluate our system, we also conducted a cross-day study to test VibroSense's performance after being reset. We reset the system after data collection on the first day, and put the tripod and sensor head away. We then turned off the system. The next day



House 1

	1	8	10	6	13	11
1	1.00	0.00	0.00	0.00	0.00	0.00
8	0.01	0.99	0.00	0.00	0.00	0.00
10	0.00	0.00	1.00	0.00	0.00	0.00
6	0.00	0.00	0.00	1.00	0.00	0.00
13	0.00	0.00	0.01	0.00	0.99	0.00
11	0.00	0.00	0.00	0.00	0.00	1.00

House 2

	1	8	3	7	10	6	11	12	13	15	2
1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
11	0.12	0.00	0.00	0.00	0.08	0.00	0.41	0.00	0.00	0.39	0.00
12	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.91	0.06	0.00	0.00
13	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.96	0.00	0.00
15	0.13	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.74	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

House 3

	1	7	10	8	3	6	12	13	2
1	0.81	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.00
7	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.26	0.00	0.58	0.00	0.00	0.00	0.10	0.06	0.00
8	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
12	0.00	0.00	0.02	0.00	0.00	0.00	0.84	0.00	0.14
13	0.24	0.00	0.20	0.00	0.00	0.00	0.00	0.51	0.05
2	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.98

House 4

	1	8	7	6	10	15	11	13	12	4	2
1	0.98	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01
8	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.07	0.00	0.00	0.87	0.01	0.02	0.01	0.01	0.00	0.01	0.00
10	0.04	0.00	0.00	0.05	0.77	0.07	0.07	0.01	0.00	0.00	0.00
15	0.41	0.00	0.00	0.08	0.24	0.09	0.18	0.01	0.00	0.00	0.00
11	0.49	0.00	0.00	0.08	0.09	0.07	0.26	0.00	0.00	0.00	0.01
13	0.00	0.00	0.00	0.00	0.11	0.01	0.02	0.85	0.01	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.99	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
2	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.99

House 5

	1	7	8	5	14	11	10	9	3	4	15	16	17	2	18
1	0.90	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.81	0.08	0.00	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.01
8	0.19	0.07	0.71	0.00	0.00	0.00	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.01	0.00	0.83	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.14	0.00	0.00
14	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	0.01	0.00	0.00	0.00	0.00	0.95	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.08	0.00	0.00	0.00	0.00	0.00	0.91	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.01
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.01	0.00	0.00	0.00
4	0.00	0.00	0.04	0.05	0.00	0.00	0.00	0.00	0.00	0.91	0.00	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
18	0.20	0.00	0.00	0.00	0.01	0.02	0.22	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.52

Fig. 7. Confusion matrices of discrete cross-day resetup experiment on 5 houses

(Day 2), we re-setup the tripod and sensor head in the same general location. The details of the data collection procedure have been presented in section 5.

For discrete cross-day testing, we trained the VibroSense system using the data (around 30 samples per activity) collected on Day 1 (before re-setup), and tested the system using the data (around 10 samples per activity) collected on Day 2 (after re-setup). The purpose of this evaluation was to test our system's robustness after being adjusted (e.g., nudged or reset). The experiment results for each house are summarized in Figure 6. Confusion matrices for each house are shown in Fig. 7. The recognition accuracies for H1 to H5 were 99.66%, 91.2%, 86.0%, 80.06%, and 90.13%, respectively. Notably, the average recognition accuracy of the five houses after re-setup was 89.4%, which is expectantly lower than the same-day testing result of 96.6%. We found similar confusions between water appliances and noise in our cross-day testing session as we did in our same-day testing sessions (e.g., activity 15 in H2 or activities 10, 11, and 15 in H4).

We think there are two possible factors that may have contributed to the decrease in accuracy before and after re-setup. Firstly, altering the angle and distance between the laser head and the wall or ceiling, may have influenced the received signal (primarily affecting the amplitude). Secondly, data before and after re-setup was collected on different days so background noise could have varied. This may have led to different frequency responses for the same activity before and after re-setup of the device, as the Figure 8 shown.

We think the performance of our system on this cross-day re-setup evaluation, may be closer to the realistic performance of our system in the real-world, compared to our discrete same-day testing results. This is because background noise may vary in real world settings. Even after re-setup, however, the performance of VibroSense is still very encouraging, recognizing 18 home activities across 5 houses with an average accuracy of 89.4%.

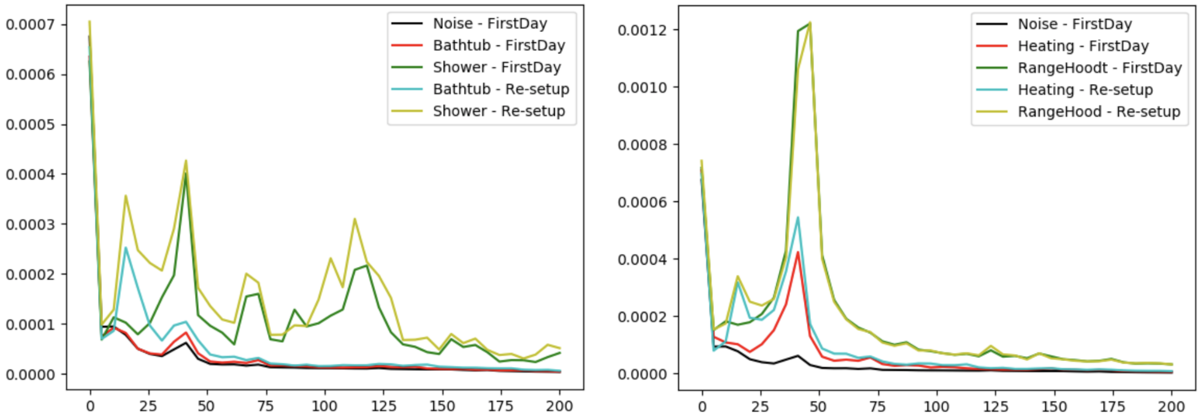


Fig. 8. The visualization for the FFT result of some activity signal in House4 across day1 and day2

6.3 Simulated Online Testing

To further evaluate the feasibility of our system in real-world scenarios, we collected a continuous stream of data for approximately 5 hours in each house to simulate online testing. The deep learning model was trained using data collected in the discrete same-day experiment. We then fed the data into our trained system to predict which activities were active every seconds. Then, we applied a sliding window of 2 seconds with a 50% overlap. The system outputted a result every second using a voting mechanism. The process of the voting mechanism is described as follows. First, the threshold was set to detect the start of an activity. Starting from the first second, when five consecutive predicted labels (each with 2 second window) were found to be the same, the system would regard the first second as the beginning of this activity. Likewise, if the system was currently predicting an activity, once it predicted five consecutive noise labels, it would note the time of the first noise label in the sequence as the end of the activity. If the duration of the activity was under 8 seconds, we discarded the activity, as we assumed all activities should last longer than 10 seconds. Overall, 12 samples were discarded for this reason in all 5 houses. Otherwise, we chose the activity with the most outputs in voting during that window as the recognition result. We compared the label activity of this identified window with the ground truth of this window. Within this window, if at any second the recognized activity was the same as the activity labeled by the ground-truth, we marked it as a correct classification. Otherwise, it was marked as an error.

The predicted results of all activities during the 5-hour time frame were then compared with the actual activity labels to calculate an overall accuracy along with the number of false-positive (FP) and false-negative (FN) errors for each house. If our system detected an activity and the ground-truth did not include any activity besides noise, we marked the detection as a false-positive (FP) error. If our system did not capture an activity (recognized as all noise), and the ground-truth in the window did in fact contain an activity of interest, we marked it as a false-negative error (FN). The average recognition accuracy across all five houses was 90.99%. The detailed results of each home are presented below:

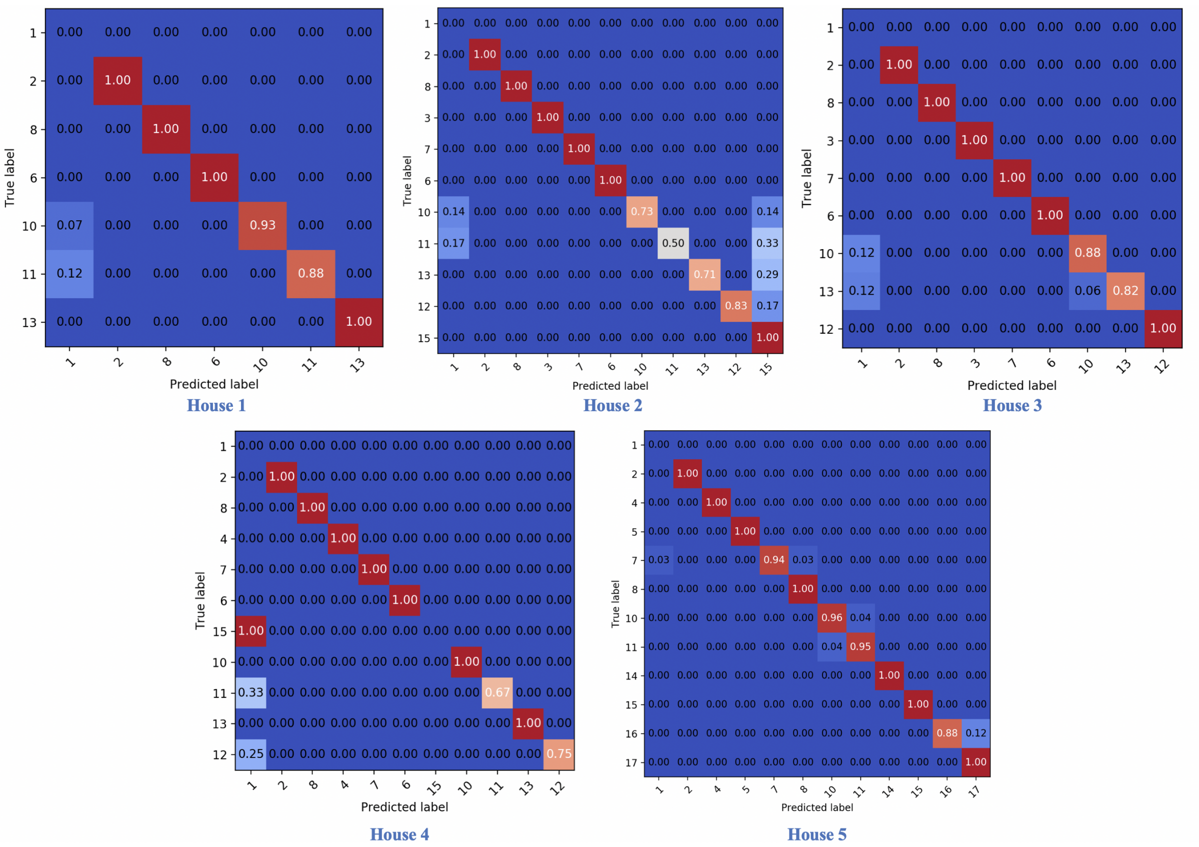


Fig. 9. Confusion matrices of Simulated Online Testing

- House 1's average accuracy was 97.19%. No false-positive error occurred among a total of 74 activity samples. One false-negative error occurred among a total of 74 activity samples. In activities 10 and 11 there was some confusion with noise. Activities 10 and 11 represent the Kitchen Faucet and Bathroom 1 Faucet, respectively.
- House 2's, average accuracy was 88.86%. Three false-positive errors occurred among a total of 114 activity samples. No false-negative errors occurred among a total of 114 activity samples. Namely, all water-related activities were either confused with one another or confused with noise. Activities 10, 11, 12, and 13 represent the Kitchen Faucet, Bathroom 1 Faucet, Bathroom 1 Bathtub Faucet, and Bathroom 1 Shower, respectively.
- House 3's average accuracy was 85.54%. Four false-positive errors occurred among a total of 71 activity samples. One false-negative error occurred among a total of 71 activity samples. In activities 10 and 13 there were confusions with noise. Activities 10 and 13 represent the Kitchen Faucet and Bathroom 1 Shower, respectively.
- House 4's average accuracy was 85.61%. 19 false-positive errors occurred among a total of 67 activity samples. Three false-negative errors occurred among a total of 67 activity samples. In activities 11, 12, and

15, there were confusions with noise. Activities 11, 12, and 15 represent the Bathroom 1 Faucet, Bathroom 1 Bathtub Faucet, and Bathroom 2 Faucet, respectively.

- House 5's average accuracy was 97.76%. Six false-positive errors occurred among a total of 161 activity samples. Two false-negative errors occurred among a total of 161 activity samples. Activity 16 was mistaken for activity 17. Activities 16 and 17 represent the Bathroom 2 Bathtub Faucet and Bathroom 2 shower, respectively.

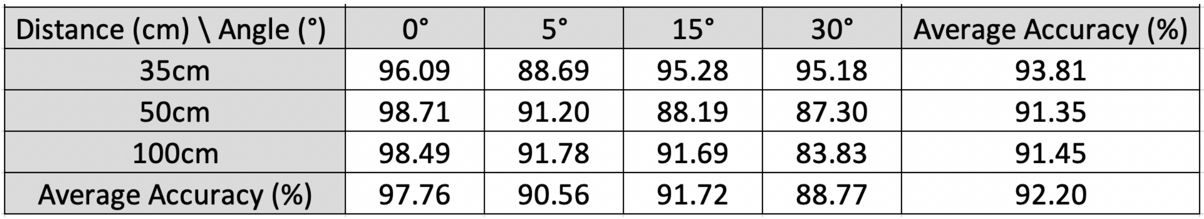
7 DISCUSSION

Overall, our findings are promising in showing the capability of using a single laser device for activity detection throughout a home. In addition to the evaluations in the previous section, we have also conducted studies testing different positions and locations of the laser, as well as identifying different stages of activities. In this section, we will discuss the results of these added experiments, the limitations of our system, and our future plans to improve VibroSense.

7.1 Influence of System Setups with Different Angles and Distances between the Laser Head and Wall

As previously discussed, it's inevitable that the LDV's position and orientation will shift over time in real-world scenarios (whether bumped accidentally or re-setup purposefully). We have already shown that such alterations in the laser's position do in fact lower the performance of our deep learning model. However, it is unclear how exactly a change in angle or distance will affect the results. To address this question, we conducted an extra experiment where we tested the LDV's performance at various angles and distances from the wall in House 3.

For the purposes of this experiment, we tested three distances from the wall (35 cm, 50 cm, and 100 cm) with four orientation angles (0, 5, 15, and 30 degrees) across nine different activities (1, 2, 3, 6, 7, 8, 10, 12, 13). For each distance and angle, we collected 30 samples for each activity, totaling $3 * 4 * 9 * 30 = 3240$ samples. We then utilized 90% of the samples collected at $50cm$ and 0 degrees (this position is consistent with the position in the evaluation) as the training data to train the deep learning model. We utilized the remaining 10% of those samples along with the remaining samples from the other 11 positions to test the model. The results are shown in Fig. 10.

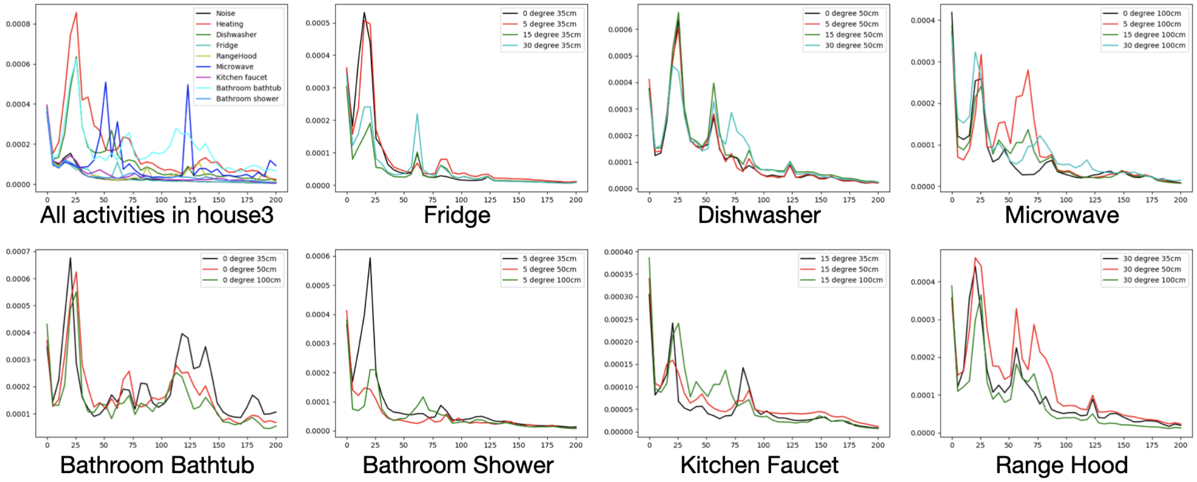


Distance (cm) \ Angle (°)	0°	5°	15°	30°	Average Accuracy (%)
35cm	96.09	88.69	95.28	95.18	93.81
50cm	98.71	91.20	88.19	87.30	91.35
100cm	98.49	91.78	91.69	83.83	91.45
Average Accuracy (%)	97.76	90.56	91.72	88.77	92.20

Fig. 10. The accuracy of the model when testing with different setup settings

The overall average accuracy of the 12 testing samples was 92.2%, demonstrating the robustness of our system despite slight changes in the position and angle of the LDV. To examine the effects of distance and angle further, we ran a one way ANOVA test. The results of the angle effect test revealed F(3,8) = 4.3 and p = 0.04. The distance effect test produced results of F(2,9) = 0.23 and p = 0.79. Overall, we found the statistically significant effect that as the orientation angle increased, accuracy decreased. However, when the distance from the wall changed, accuracy of the system was not significantly affected. To explain this phenomenon, we analyzed the physics principles behind it based on our theory of operation. Specifically, we note that when the LDV position and angle change, there are two main factors that are influenced: strength of the received signal and the location of the projected point.

Strength of the Received Signal: According to optical principles and the visualization in Fig. 11, when the LDV distance and incident angle increases, the strength of the received signal decreases. This decrease of the valid activity signal would lead to a lower Signal-Noise-Ratio (SNR), which may cause a decline in recognition accuracy.



All activities in house3 | Fridge | Dishwasher | Microwave

Bathroom Bathtub | Bathroom Shower | Kitchen Faucet | Range Hood

Fig. 11. The visualization for FFT result of the activities signal with different LDV locations and angles

The Projected Point: When the orientation of the LDV rotates, its projected point on the wall also shifts. Therefore, according to the vibration model in our Theory of Operation section, this movement may cause $H(f)$ to change. Since $X(f) = H(f)X(f) + N(f)$, a change in $H(f)$ will also lead to a change in the spectrum distribution of the received signal, which may confuse the deep learning model and reduce performance.

7.2 Results of Different LDV Locations in House 3

To explore the influence of LDV locations on our physics model and recognition performance, we selected three laser projected points in different locations as shown in Fig.12(a) (location 1 is the location we used for house 3 in section 6). Next, we collected data for activities 1, 2, 6, 7, and 10 in these three locations. The experiment involved two parts. First, we used data from location 1 to train a model. Then, we used data from locations 2 and 3 to test the model. The results show that the accuracies of the data from location 2 and location 3 are only 53.45% and 61.31%, respectively. There is a relatively large drop in accuracy compared to the results found in section 6, which can be explained by our physics model. When the LDV changes locations, the propagation path from the activities to the projected point also changes, which causes $H(f)$ to change. Next, we applied 10-fold cross validation on each data set from the three different locations (each location used their own data for training and testing). The recognition accuracy of the three locations was 99.83% for location 1, 98.6% for location 2, and 99.75% for location 3. Overall, these results demonstrate that our system can work well when set up in a variety of locations. We then analyzed the results of each activity in different locations. We found most activity accuracies to be around 99%. Activity 13 in location 2, however, had an accuracy of only 92%. As shown in Fig. 12(a) and (b), when the LDV is placed in location 2 on the first floor, the distance between the projected point and the bathroom 1 shower (activity 13) on the second floor is quite large, which may cause greater signal attenuation during propagation (lower SNR) and subsequently lower recognition accuracy.

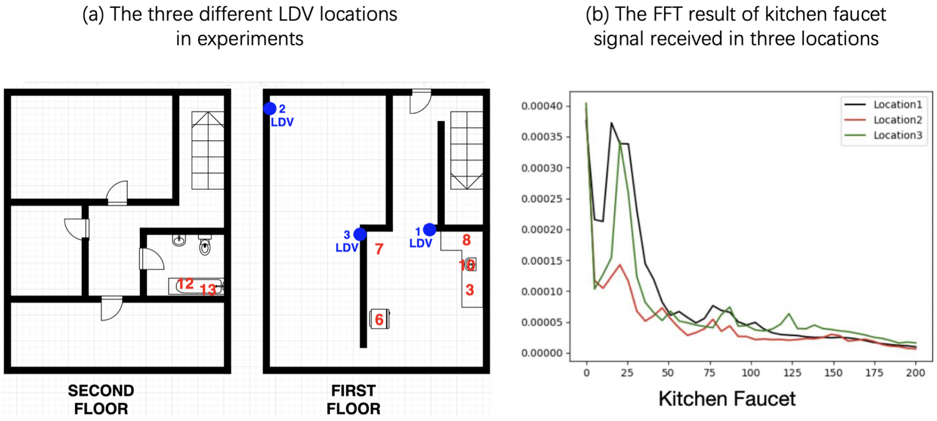


Fig. 12. The influence of different LDV locations on VibroSense system recognition

7.3 The Influence of Training Sample Size on Recognition Performance

To test the influence of training sample size on recognition performance, we began by collecting 60 samples for 9 different activities (1, 2, 3, 6, 7, 8, 10, 12, 13). Each sample lasted for a duration of 15 seconds. We then divided each sample into 14 windows (2-second length and 50% overlap). Next, we randomly selected 6 samples for testing data. Among the remaining 54 samples, we selected groups of 3, 5, 10, 15, 20, 30, and 40 samples to train our deep learning model and test them with the same testing data. The results are shown in Fig. 13. We found that recognition accuracy increases as the sample size increases. When the sample size increases to 30, the final accuracy exceeds 99%. This demonstrates that the 30 samples we collected for each activity in each house is an appropriate amount.

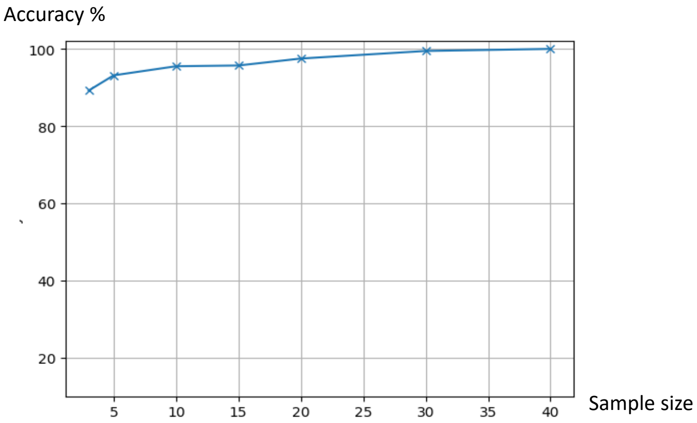


Fig. 13. The influence of training sample size on recognition performance

7.4 Concurrent Activities

In reality, home activities do not always occur in separate intervals. Instead, they often overlap or occur concurrently. For example, the microwave and water faucet are commonly used together while residents are cooking.

An ideal system for activity recognition should be able to distinguish concurrent activities. However, recognizing concurrent activities is a challenging topic for all similar systems. For instance, it is hard to collect data for all possible combinations of 18 single home activities. In our experiment, we collected a small amount of concurrent activities data of two single activities at 5 houses. The initial results were encouraging, but we feel more data is needed to draw any conclusion. Therefore, we decided to leave this section for future work.

7.5 Recognizing Stages of an Activity

During our study, we only collected data from one particular stage of each single event. For example, we collected data from the boiling stage of the electric kettle. However, in a real-world scenario, many appliances have different stages of operation. This experiment would expand the applicability of our system for it to incorporate detection of all activities in each house regardless of the stages of operation. To test whether our approach can further diversify activities into different stages, we selected the toilet and the washer to represent each of the appliance types (water appliances and electrical appliances). We divided toilet activities into two stages: flushing and replenishing. We then collected 38 complete cycles of flushing and replenishing. Each cycle lasted around 1.2 minutes - 12 seconds flushing and 60 seconds replenishing. Similarly, we divided washer activities into washing, rinsing, and spinning. We collected 2 complete cycles of washer activity data. Each cycle lasted a total of 20 minutes. During data collection, we made video and audio recordings using our smartphones to track the stages in order to manually label the ground-truth. In total for the toilet, we labeled 240-seconds of flushing data and 38-minutes of replenishing data. As for the washer, we labeled 12-minutes of washing data, 10-minutes of rinsing data, and 2-minutes of spinning data. Finally, we conducted a 10-fold cross validation on the collected data using the same machine learning pipeline, which resulted in an average accuracy of 97.5%. The confusion matrix in Fig. 14 visualizes our multiple stage experiments. The results show that our system is able to recognize multiple stages of a single appliance.

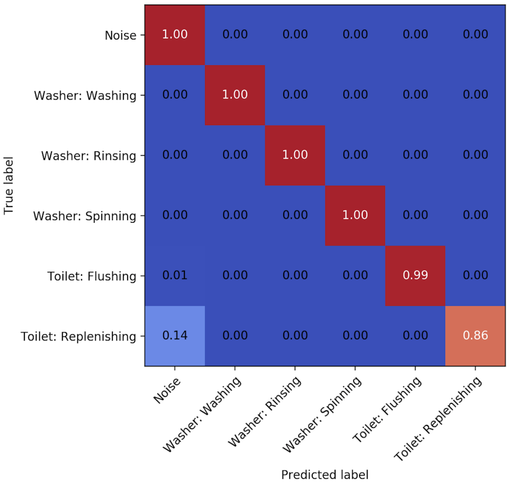


Fig. 14. Confusion Matrix of recognizing multiple stages of activities

7.6 Applying VibroSense to Other Sensing Methods

The research question we answered in this paper is whether or not subtle structural vibrations captured from one point in a house can be used to recognize different activities throughout an entire house. We chose to use a laser Doppler vibrometer as the sensing device, which is relatively expensive at this moment. However, our findings and research does not have to be limited to an LDV. We recognize that other sensing devices (e.g., high-end accelerometer, high-end GEOPhone) that can provide high sensitivity in capturing structural vibrations may also perform well with this task. Hence, given the data format is similar, the data processing pipeline we presented in VibroSense can be quickly applied to data collected from other sensing devices as well. In other words, the research findings in this paper do not need to be limited to one type of sensing device (laser Doppler vibrometer).

7.7 The Sensing Range of VibroSense

VibroSense uses a laser Doppler vibrometer to capture the structural vibration caused by different activities. The performance of the system can be influenced by the quality of the received signal. These vibration signals usually attenuate as the travel distance increases. If the travel distance is too long, it is possible that structural vibration can be too weak to be picked up and recognized using VibroSense. In other words, VibroSense may not work in a setting where the house is beyond the range of sensing. In our study, we deployed and tested the system in five houses. The size of H1 to H5 were 864, 1300, 1300, 1400, and 1280 square feet, respectively. Based on these floor plans, our study demonstrated that our system could cover a range of 864 square feet to 1400 square feet. At this moment, we do not know how VibroSense would perform if the size of a house is significantly larger than 1400 square feet. Besides size, a different floor plan may also influence the performance, as we only tested in houses with 1 and 2 floors. In order to fully understand this problem, a larger and more thorough study with more types of houses is needed. We plan to explore this further in the future.

7.8 Deep Learning Model Improvement

In the future, our current deep learning model could be improved in three aspects. Firstly, a data augmentation method could be implemented to increase data scale and diversity, which may help to improve the adaptability and performance of our deep learning model. Specifically, we could collect different kinds of background noises and add them to the activity signal before feeding them into the deep learning model. Secondly, we collected different amounts of data for different kinds of activities (due to the differences in appliances across houses) resulting in imbalanced data. A weighted loss function could be implemented to address this issue, possibly improving performance. Thirdly, we could work on developing a new architecture for our deep learning model. For example, we could append a GRU and LSTM module behind our Resnet module, which may improve recognition accuracy.

7.9 Recognizing Activities With Short Time Spans

In our current system, all the activities recognized occur continuously over a relatively long period of time. However, in the real world, there are still other home activities that occur in shorter time spans (e.g., doors opening/closing, windows opening/closing, humans walking around etc). Confronted with such activities, our 2-second windowing method may fail. Thus, to segment these types of activities, we would need to put forward a new algorithm. For instance, in the future, a Recurrent Neural Network (RNN) model could be used to recognize activities with shorter time spans.

7.10 Applications

Our system has many potential applications in the future of smart homes. For example, since our system can detect both the occurrence of an indoor activity as well as the time of the activity, it could be used to estimate electricity and water usage rates and provide energy-saving advice for homeowners. Furthermore, our system could improve energy security by monitoring energy usage and preventing water and electrical leakage, as well as electrical failures such as short-circuits in home appliances.

7.11 Limitations and Future Work

Although the above results are promising, there are still some limitations to our system. We outlined some of these limitations and included a discussion below.

7.11.1 Privacy. Privacy may be a serious concern in the potential implementation of our system into the real world. For this reason, we chose not to conduct our experiments in any apartments. In an apartment, there are many people living in the same building within close proximity to one another. It's very possible that our system may collect information from people living in the apartment next door without their permission. Although there are many methods that can protect privacy in these data (e.g., extracting features on the fly without saving raw data), the specific influence of our system on personal privacy necessitates further study.

7.11.2 Weak Vibrations. Even though our evaluation results for activity recognition were quite high, our system performed poorly when recognizing certain activities at certain houses. Namely, Activity 10 (Kitchen Faucet), Activity 11 (Bathroom 1 Faucet), and Activity 15 (Bathroom 2 Faucet) in House 4 after re-setup caused confusions. There are two potential causes for this less-than-optimal performance. Firstly, structural vibrations caused by water flow may be relatively weak. Unlike other activities, water-related activities introduce a new medium. Secondly, House 4 had the largest square footage among the five houses. Thus, the distance was greater between the source of the activities and the LDV, which further increases signal attenuation. These two reasons potentially led to lower signal noise ratios (SNR). This issue can be further amplified when the device was re-setup in the second day, where the system was confused among water activities and background noise.

7.11.3 House Independent Models. Currently, our deep learning model is house-dependent, meaning that implementing this smart-home technology in a new house would require the recollection of data and the training of a new model. However, the same indoor activities may have similar vibration patterns across different homes. A large amount of training data may enable the deep learning model to learn such subtle similarities across houses. Hence, if the trained data set were large enough, it could be possible to apply the migration process of the deep learning model from one house to another without the need to retrain the model, using only a small collection of new data.

7.11.4 Device Portability. To begin with, our device is cumbersome and expensive. However, the device we used was fabricated in 2008. Modern-day LDVs are much smaller and more portable [1]. Furthermore, they can also be fiber-based and use a micro-lens, which will further reduce their size in the future.

7.11.5 Device Cost. As for the cost, the high price point of an LDV can be attributed to the device's high-quality components. However, many of these components are not always necessary since we do not require such high levels of precision in regular scenarios. Hypothetically, we could create a makeshift DIY LDV composed of cheaper components. Additionally, almost every LDV has a Bragg Cell to help with identifying the direction of the vibration. However in this paper, this functionality is not required. With continual improvements in technology and these points in mind, costs can be significantly reduced.

8 CONCLUSION

In this paper, we present VibroSense, a novel sensing technology that can recognize up to 18 home activities related to electrical and water appliances by analyzing subtle structural vibrations at a single point on the wall or ceiling of a house. The vibrations are captured using a laser Doppler vibrometer which is used to train a deep neural network (CNN) for activity recognition. Based on the data collected over 5 houses, VibroSense was able to distinguish between these 18 home activities with an average accuracy of 96.6% when the data was collected on the same day, 89.4% after re-setup on the second day, and 91% for simulated online testing. It presents an encouraging step towards using a single, non-contact, sensing device to recognize home activities across rooms in a single house.

ACKNOWLEDGMENTS

This work is supported by the Information Science Department at Cornell University. We would like to thank the owners of the houses used in our study for granting us permission to conduct our study in their houses. We would also like to thank the reviewers for their constructive comments, and lab mates in the Cornell SciFi lab for providing early feedback.

REFERENCES

[1] Accessed: Feb. 2020. VibroGo. https://www.polytec.com/eu/vibrometry/products/single-point-vibrometers/vibrogo/

[2] R Abbaszadeh, A Rajabipour, H Ahmadi, MJ Mahjoob, and M Delshad. 2013. Prediction of watermelon quality based on vibration spectrum. *Postharvest biology and technology* 86 (2013), 291–293.

[3] Jan Achenbach. 2012. *Wave propagation in elastic solids*. Elsevier.

[4] JRM Aerts and JJJ Dirckx. 2010. Nonlinearity in eardrum vibration as a function of frequency and sound pressure. *Hearing research* 263, 1-2 (2010), 26–32.

[5] Yekutiel Avargel and Israel Cohen. 2011. Speech measurements using a laser Doppler vibrometer sensor: Application to speech enhancement. In *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays*. IEEE, 109–114.

[6] Daniel Avrahami, Mitesh Patel, Yusuke Yamaura, and Sven Kratz. 2018. Below the surface: Unobtrusive activity recognition for work surfaces using RF-radar sensing. In *23rd International Conference on Intelligent User Interfaces*. 439–451.

[7] Amelie Bonde, Shijia Pan, Hae Young Noh, and Pei Zhang. 2019. Deskbuddy: an office activity detection system: demo abstract. In *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*. 352–353.

[8] P Castellini, M Martarelli, and EP Tomasini. 2006. Laser Doppler Vibrometry: Development of advanced solutions answering to technology's needs. *Mechanical systems and signal processing* 20, 6 (2006), 1265–1285.

[9] Ke-Yu Chen, Sidhant Gupta, Eric C Larson, and Shwetak Patel. 2015. DOSE: Detecting user-driven operating states of electronic devices from a single sensing point. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. IEEE, 46–54.

[10] Gabe Cohn, Sidhant Gupta, Jon Froehlich, Eric Larson, and Shwetak N Patel. 2010. GasSense: Appliance-level, single-point sensing of gas activity in the home. In *International Conference on Pervasive Computing*. Springer, 265–282.

[11] Gabe Cohn, Sidhant Gupta, Tien-Jui Lee, Dan Morris, Joshua R Smith, Matthew S Reynolds, Desney S Tan, and Shwetak N Patel. 2012. An ultra-low-power human body motion sensor using static electric field sensing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. 99–102.

[12] Abe Davis, Katherine L Bouman, Justin G Chen, Michael Rubinstein, Fredo Durand, and William T Freeman. 2015. Visual vibrometry: Estimating material properties from small motion in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5335–5343.

[13] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham J Mysore, Frédo Durand, and William T Freeman. 2014. The visual microphone: passive recovery of sound from video. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 79.

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

[15] JONATHON FAGERT, MOSTAFA MIRSHEKARI, SHIJIA PAN, PEI ZHANG, and HAE YOUNG NOH. 2019. Vibration Source Separation for Multiple People Gait Monitoring Using Footstep-Induced Floor Vibrations. *Structural Health Monitoring 2019* (2019).

[16] James Fogarty, Carolyn Au, and Scott E Hudson. 2006. Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*. ACM, 91–100.

[17] Jon E Froehlich, Eric Larson, Tim Campbell, Conor Haggerty, James Fogarty, and Shwetak N Patel. 2009. HydroSense: infrastructure-mediated single-point sensing of whole-home water activity. In *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, 235–244.

[18] D Goyal and BS Pabla. 2016. The vibration monitoring methods and signal processing techniques for structural health monitoring: a review. *Archives of Computational Methods in Engineering* 23, 4 (2016), 585–594.

[19] 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, 139–148.

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[21] Tong He, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. 2019. Bag of tricks for image classification with convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 558–567.

[22] Alexander M Huber, Geoffrey R Ball, Dorothe Veraguth, Norbert Dillier, Daniel Bodmer, and Damien Sequeira. 2006. A new implantable middle ear hearing device for mixed hearing loss: a feasibility study in human temporal bones. *Otology & neurotology* 27, 8 (2006), 1104–1109.

[23] Alexander M Huber, Christoph Schwab, Thomas Linder, Sandro J Stoeckli, Mattia Ferrazzini, Norbert Dillier, and Ugo Fisch. 2001. Evaluation of eardrum laser Doppler interferometry as a diagnostic tool. *The Laryngoscope* 111, 3 (2001), 501–507.

[24] Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167* (2015).

[25] Younghun Kim, Thomas Schmid, Zainul M Charbiwala, and Mani B Srivastava. 2009. ViridiScope: design and implementation of a fine grained power monitoring system for homes. In *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, 245–254.

[26] Stacey Kuznetsov and Eric Paulos. 2010. UpStream: motivating water conservation with low-cost water flow sensing and persuasive displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1851–1860.

[27] Gierad Laput, Karan Ahuja, Mayank Goel, and Chris Harrison. 2018. Ubicoustics: Plug-and-play acoustic activity recognition. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 213–224.

[28] Gierad Laput, Walter S Lasecki, Jason Wiese, Robert Xiao, Jeffrey P Bigham, and Chris Harrison. 2015. Zensors: Adaptive, rapidly deployable, human-intelligent sensor feeds. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1935–1944.

[29] Gierad Laput, Yang Zhang, and Chris Harrison. 2017. Synthetic sensors: Towards general-purpose sensing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 3986–3999.

[30] Hanchuan Li, Can Ye, and Alanson P Sample. 2015. IDSense: A human object interaction detection system based on passive UHF RFID. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2555–2564.

[31] Shengjie Li, Xiang Li, Qin Lv, Guiyu Tian, and Daqing Zhang. 2018. WiFit: Ubiquitous bodyweight exercise monitoring with commodity Wi-Fi devices. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 530–537.

[32] Xuefeng Liu, Jiannong Cao, Shaojie Tang, and Jiaqi Wen. 2014. Wi-Sleep: Contactless sleep monitoring via WiFi signals. In *2014 IEEE Real-Time Systems Symposium*. IEEE, 346–355.

[33] Guomin Luo and Daming Zhang. 2012. *Wavelet Denoising*. https://doi.org/10.5772/37424

[34] Shota Mashiyama, Jihoon Hong, and Tomoaki Ohtsuki. 2015. Activity recognition using low resolution infrared array sensor. In *2015 IEEE International Conference on Communications (ICC)*. IEEE, 495–500.

[35] Mostafa Mirshekari, Jonathon Fagert, Amelie Bonde, Pei Zhang, and Hae Young Noh. 2018. Human gait monitoring using footstep-induced floor vibrations across different structures. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. 1382–1391.

[36] 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.

[37] Mostafa Mirshekari, Pei Zhang, and Hae Young Noh. 2016. Non-intrusive occupant localization using floor vibrations in dispersive structure. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*. 378–379.

[38] Pieter GG Muyshondt, Joris AM Soons, Daniël De Greef, Felipe Pires, Peter Aerts, and Joris JJ Dirckx. 2016. A single-ossicle ear: acoustic response and mechanical properties measured in duck. *Hearing research* 340 (2016), 35–42.

[39] Kazuya Ohara, Takuya Maekawa, and Yasuyuki Matsushita. 2017. Detecting state changes of indoor everyday objects using Wi-Fi channel state information. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–28.

[40] Wieslaw Ostachowicz, Maciej Radzieński, and Pawel Kudela. 2014. 50th anniversary article: comparison studies of full wavefield signal processing for crack detection. *Strain* 50, 4 (2014), 275–291.

[41] 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 *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. 149–158.

[42] 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), 1–31.

[43] Shwetak N Patel, Matthew S Reynolds, and Gregory D Abowd. 2008. Detecting human movement by differential air pressure sensing in HVAC system ductwork: An exploration in infrastructure mediated sensing. In *International Conference on Pervasive Computing*. Springer, 1–18.

[44] 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.

[45] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking*. 27–38.

[46] Marcus Rohrbach, Sikandar Amin, Mykhaylo Andriluka, and Bernt Schiele. 2012. A database for fine grained activity detection of cooking activities. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1194–1201.

[47] NB Roozen, Ludovic Labelle, Monika Rychtáriková, and Christ Glorieux. 2015. Determining radiated sound power of building structures by means of laser Doppler vibrometry. *Journal of Sound and Vibration* 346 (2015), 81–99.

[48] SJ Rothberg, MS Allen, P Castellini, D Di Maio, JJJ Dirckx, DJ Ewins, Ben J Halkon, P Muyshondt, N Paone, T Ryan, et al. 2017. An international review of laser Doppler vibrometry: Making light work of vibration measurement. *Optics and Lasers in Engineering* 99 (2017), 11–22.

[49] Laixi Shi, Mostafa Mirshekari, Jonathon Fagert, Yuejie Chi, Hae Young Noh, Pei Zhang, and Shijia Pan. 2019. Device-free Multiple People Localization through Floor Vibration. In *Proceedings of the 1st ACM International Workshop on Device-Free Human Sensing*. 57–61.

[50] Laixi Shi, Yue Zhang, Shijia Pan, and Yuejie Chi. 2020. Data Quality-Informed Multiple Occupant Localization using Floor Vibration Sensing. In *Proceedings of the 21st International Workshop on Mobile Computing Systems and Applications*. 98–98.

[51] Shuyu Shi, Stephan Sigg, and Yusheng Ji. 2012. Passive detection of situations from ambient fm-radio signals. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. 1049–1053.

[52] Joshua R Smith, Kenneth P Fishkin, Bing Jiang, Alexander Mamishev, Matthai Philipose, Adam D Rea, Sumit Roy, and Kishore Sundara-Rajan. 2005. RFID-based techniques for human-activity detection. *Commun. ACM* 48, 9 (2005), 39–44.

[53] Andrew Spielberg, Alanson Sample, Scott E Hudson, Jennifer Mankoff, and James McCann. 2016. RapID: A framework for fabricating low-latency interactive objects with RFID tags. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 5897–5908.

[54] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15, 1 (2014), 1929–1958.

[55] WJ Staszewski, BC Lee, L Mallet, and F Scarpa. 2004. Structural health monitoring using scanning laser vibrometry: I. Lamb wave sensing. *Smart Materials and Structures* 13, 2 (2004), 251.

[56] Jaeyong Sung, Colin Ponce, Bart Selman, and Ashutosh Saxena. 2012. Unstructured human activity detection from rgbd images. In *2012 IEEE international conference on robotics and automation*. IEEE, 842–849.

[57] Habib Tabatabai, David E Oliver, John W Rohrbaugh, and Christopher Papadopoulos. 2013. Novel applications of laser Doppler vibration measurements to medical imaging. *Sensing and Imaging: An International Journal* 14, 1-2 (2013), 13–28.

[58] Emmanuel Munguia Tapia, Stephen S Intille, and Kent Larson. 2004. Activity recognition in the home using simple and ubiquitous sensors. In *International conference on pervasive computing*. Springer, 158–175.

[59] Emmanuel Munguia Tapia, Stephen S Intille, and Kent Larson. 2007. Portable wireless sensors for object usage sensing in the home: Challenges and practicalities. In *European Conference on Ambient Intelligence*. Springer, 19–37.

[60] AA Veber, A Lyashedko, E Sholokhov, A Trikshev, A Kurkov, Y Pyrkov, AE Veber, V Seregin, and V Tsvetkov. 2011. Laser vibrometry based on analysis of the speckle pattern from a remote object. *Applied Physics B: Lasers and Optics* 105, 3 (2011), 613–617.

[61] JF Vignola, X Liu, SF Morse, BH Houston, JA Bucaro, MH Marcus, DM Photiadis, and L Sekaric. 2002. Characterization of silicon micro-oscillators by scanning laser vibrometry. *Review of scientific instruments* 73, 10 (2002), 3584–3588.

[62] Hao Wang, Daqing Zhang, Junyi Ma, Yasha Wang, Yuxiang Wang, Dan Wu, Tao Gu, and Bing Xie. 2016. Human respiration detection with commodity wifi devices: do user location and body orientation matter?. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 25–36.

[63] Hao Wang, Daqing Zhang, Yasha Wang, Junyi Ma, Yuxiang Wang, and Shengjie Li. 2016. RT-Fall: A real-time and contactless fall detection system with commodity WiFi devices. *IEEE Transactions on Mobile Computing* 16, 2 (2016), 511–526.

[64] Zhu Wang, Bin Guo, Zhiwen Yu, and Xingshe Zhou. 2018. Wi-Fi CSI-based behavior recognition: From signals and actions to activities. *IEEE Communications Magazine* 56, 5 (2018), 109–115.

[65] Daniel H Wilson and Chris Atkeson. 2005. Simultaneous tracking and activity recognition (STAR) using many anonymous, binary sensors. In *International Conference on Pervasive Computing*. Springer, 62–79.

[66] Jason Wu, Chris Harrison, Jeffrey P Bigham, and Gierad Laput. 2020. Automated Class Discovery and One-Shot Interactions for Acoustic Activity Recognition. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[67] Zeev Zalevsky, Yevgeny Beiderman, Israel Margalit, Shimshon Gingold, Mina Teicher, Vicente Mico, and Javier Garcia. 2009. Simultaneous remote extraction of multiple speech sources and heart beats from secondary speckles pattern. *Optics express* 17, 24 (2009), 21566–21580.

[68] Chenyang Zhang and Yingli Tian. 2012. RGB-D camera-based daily living activity recognition. *Journal of computer vision and image processing* 2, 4 (2012), 12.

[69] Yang Zhang, Gierad Laput, and Chris Harrison. 2018. Vibrosight: Long-Range Vibrometry for Smart Environment Sensing. In *The 31st Annual ACM Symposium on User Interface Software and Technology*. ACM, 225–236.

[70] Yue Zhang, Shijia Pan, Jonathon Fagert, Mostafa Mirshekari, Hae Young Noh, Pei Zhang, and Lin Zhang. 2018. Occupant Activity Level Estimation Using Floor Vibration. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on*

Pervasive and Ubiquitous Computing and Wearable Computers. 1355–1363.

[71] Yang Zhang, Chouchang Yang, Scott E Hudson, Chris Harrison, and Alanson Sample. 2018. Wall++ Room-Scale Interactive and Context-Aware Sensing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–15.

[72] Zhongna Zhou, Xi Chen, Yu-Chia Chung, Zhihai He, Tony X Han, and James M Keller. 2008. Activity analysis, summarization, and visualization for indoor human activity monitoring. *IEEE Transactions on Circuits and Systems for Video Technology* 18, 11 (2008), 1489–1498.

[73] L Zipser and H Franke. 2004. Laser-scanning vibrometry for ultrasonic transducer development. *Sensors and Actuators A: Physical* 110, 1-3 (2004), 264–268.