Clean Vibes: Hand Washing Monitoring Using Structural Vibration Sensing

JONATHON FAGERT, Baldwin Wallace University, USA
AMELIE BONDE, Carnegie Mellon University, USA
SRUTI SRINIDHI, Carnegie Mellon University, USA
SARAH HAMILTON, University of California, Santa Barbara, USA
PEI ZHANG, University of Michigan, USA
HAE YOUNG NOH, Stanford University, USA

We present a passive and non-intrusive sensing system for monitoring hand washing activity using structural vibration sensing. Proper hand washing is one of the most effective ways to limit the spread and transmission of disease, and has been especially critical during the COVID-19 pandemic. Prior approaches include direct observation and sensing-based approaches, but are limited in non-clinical settings due to operational restrictions and privacy concerns in sensitive areas such as restrooms. Our work introduces a new sensing modality for hand washing monitoring, which measures hand washing activity-induced vibration responses of sink structures, and uses those responses to monitor the presence and duration of hand washing. Primary research challenges are that vibration responses are similar for different activities, occur on different surfaces/structures, and tend to overlap/coincide. We overcome these challenges by extracting information about signal periodicity for similar activities through cepstrum-based features, leveraging hierarchical learning to differentiate activities on different surfaces, and denoting "primary/secondary" activities based on their relative frequency and importance. We evaluate our approach using real-world hand washing data across 4 different sink structures/locations, and achieve an average F1-score for hand washing activities of 0.95, which represents a 8.8X and 10.2X reduction in error over two different baseline approaches.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**; • **Computing methodologies** → *Classification and regression trees*; • **Applied computing → Health care information systems**.

Additional Key Words and Phrases: Hand Washing, Hand Hygiene, Structural Vibration Sensing

1 INTRODUCTION

Proper hand washing is critical for prevention of healthcare-associated infections (HCAIs) (those that occur in or around healthcare settings) and reducing disease transmission rates. During the 2020 COVID-19 outbreak, hand washing was identified as one of the best practices for reducing transmission [24]. Despite this, recent studies have shown that HCAIs affect millions of persons each year, with an estimated incidence rate of 4.5% (1.7 million

Authors' addresses: Jonathon Fagert, jfagert@bw.edu, Baldwin Wallace University, 275 Eastland Road, Berea, Ohio, 44107, USA; Amelie Bonde, Carnegie Mellon University, 5000 Forbes Avenue, Moffett Field, California, 94035, USA, abonde@andrew.cmu.edu; Sruti Srinidhi, ssrinidh@andrew.cmu.edu, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania, 15213, USA; Sarah Hamilton, University of California, Santa Barbara, UC Santa Barbara, Santa Barbara, California, 93106, USA, sihamilton@bren.ucsb.edu; Pei Zhang, University of Michigan, 1301 Beal Ave, Ann Arbor, Michigan, 48109, USA, peizhang@umich.edu; Hae Young Noh, Stanford University, 473 Via Ortega, Stanford, California, 94305, USA, noh@stanford.edu.

2637-8051/2022/1-ART1 $15.00
https://doi.org/10.1145/3511890

patients) annually in the United States of America [74]. To combat these HCAIs and help reduce the spread of COVID-19, the World Health Organization (WHO) and Centers for Disease Control (CDC) have published guidelines for hand washing in healthcare settings as well as for the general public [24, 34, 57]. Despite this, studies show that some healthcare providers only practice proper hand washing less than 50% of the time [11]. As a result, there is a clear need for hand washing detection and monitoring systems.

Existing approaches for detection and monitoring of hand washing in healthcare settings include direct observation as well as sensing-based techniques such as vision, radio frequency (RF), acoustics, and wearables [4, 12, 26, 40, 47, 48, 58, 61]. However, each of these prior approaches is limited in many real-world applications due to deployment restrictions such as line of sight/perceived privacy concerns (vision), sparse/insufficient monitoring (direct observation), sensitivity to ambient noise (acoustics), and requiring users to wear and/or carry a device (RF, wearables). These limitations restrict the ability to accurately and ubiquitously monitor hand washing, which reduces their ability to improve hand hygiene practices.

To overcome the limitations of these prior works, we introduce a new approach which leverages structural vibration sensing to monitor hand washing activity. The primary insight behind this approach is that the various phases of hand washing (i.e., walking to the hand washing station, turning on water, using soap, and rinsing hands in the water), all generate excitations in the sink structure and/or surrounding floor structure. By measuring vibrations of the sink structure, our approach can accurately detect whether each activity has occurred, and monitor its duration to ensure proper compliance with hand washing guidelines. This sensing system enables ubiquitous, non-intrusive monitoring of hand washing in a variety of settings without the need for persons to wear or carry a device. In our prior work, we have shown that structural vibration sensing can accurately detect and classify hand washing activity at small scale (i.e., with one person at one sink structure) [22]. In this work, we expand that work to additional locations and additional users, and modify our approach to one that enables an accurate and robust approach for detecting and monitoring hand washing activities.

The challenges with expanding the system, however, are three-fold as follows:

(1) *Human behavior similarities:* we observe that hand washing activities generate similar vibration responses. For example, footsteps and using a soap dispenser both generate impulsive responses, water running and rinsing hands both generate periodic and less impulsive responses. As such, it is difficult to accurately detect and monitor each phase of hand washing using time series data alone. Figure 1 shows an example of this challenge. In this hand washing-induced vibration signal, we observe that the "water" and "rinsing" responses are very similar and difficult to distinguish using time series data or traditional frequency-based features alone, making it difficult to detect and monitor each stage of hand washing.
(2) *Varying interaction surfaces:* the various human activities associated with hand washing typically occur on different surfaces within the building structure. For example, footsteps occur on the floor, soap dispensers are either on a wall or on the sink, and water runs in the sink basin. As a result, the vibration signals received by the sensors are a mixture of the responses from each structure/surface that the human activity-induced excitation passes through when it propagates. This effect introduces additional noise, changes the signal characteristics, and makes it challenging to characterize the response from each activity.
(3) *Concurrent activities:* various hand washing activities may overlap, which increases the difficulty in uniquely identifying and monitoring their duration. For example, if a person leaves the water running while using a soap dispenser, these two different activities are now overlapping one another, making them more difficult to track individually.

.

To address these research challenges, we introduce a cepstrum-based hierarchical learning approach, which extracts unique signal characteristics for different activities and characterizes responses on different surfaces. This hierarchical approach first determines the presence of hand washing activities, then uses two additional

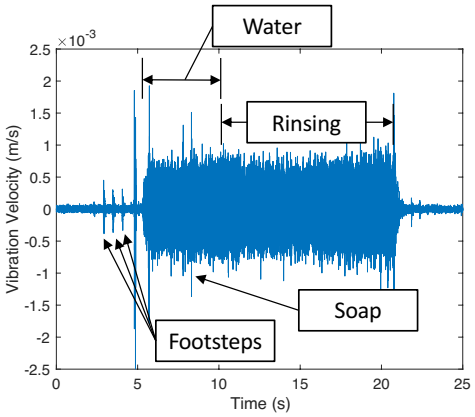


Fig. 1. Vibration response from hand washing activities. Note the similarity between different activity's responses (water, rinsing) and that some are overlapping (e.g., soap and water)

layers to classify those activities that happen away from the sink (i.e., footsteps of someone approaching the sink), and then those activities that occur on/around the sink. For these models we extract unique aspects of each activities' vibration response using a cepstrum analysis. Cepstrum analysis is mostly used in the speech recognition domain to identify and emphasize periodicity in time domain signals [62]. It differs from other approaches for identifying signal periodicity (e.g., Fourier analysis and autocorrelation) by taking the logarithm of Fourier transform amplitudes and then computing the inverse Fourier transform of those values. By combining a Fourier analysis with a logarithm, the periodic components of the signal are emphasized and visible as peaks in the quefrency (time) domain. In the context of hand washing activities, this enables better distinction between seemingly similar activities (i.e., footsteps vs. soap; rinsing vs. water running) because each of these different activities changes the periodicity of the signal. For example, rinsing ones hands in running water interrupts the normal, periodic flow of the water itself, which changes the vibration excitation and response. In this way, our approach is better able to characterize these "similar" activities and achieves a higher accuracy in classifying each. By leveraging a hierarchical learning approach, we are able to characterize the responses from each interaction surface independently. We create one layer for "idle" activities, which is a mixture of each surrounding structure response, one layer for floor interactions like footsteps, and another for sink interactions like water, rinsing, and using a soap dispenser. With this hierarchical approach, we characterize the differences in the signals for hand washing activities by introducing an independent classification layer for each interaction surface. In this way, we overcome the challenge associated with interactions on multiple surfaces. Lastly, to address the challenge of overlapping signals, we identify that the overlapping activities typically consist of a "primary" and "secondary" activity (i.e., water running in the background when soap is being used). Therefore, we label hand washing activities by their primary activity only. In this way, we can robustly monitor hand washing practices and accurately detect/monitor the primary actions.

To validate robustness in different structures and different hand-washing behavior, We have conducted real-world hand washing experiments in four different buildings (with varying sink type and soap dispenser type), with four experimental participants, and with approximately 40 minutes of hand washing activities.

In summary, the primary research contributions of this work are as follows:

(1) We introduce a novel approach for passive and ubiquitous monitoring of hand washing activities using structural vibration sensing.

(2) We extract unique signal components corresponding to each hand washing-associated activity using a cepstrum-based feature extraction to reduce signal similarity and improve monitoring robustness.
(3) We evaluate our system and approach with real-world hand washing experiments in four buildings and with four different experimental participants.

The remainder of the paper is organized as follows. First, in Section 2 we explore the relevant related work in the areas of hand washing monitoring and structural vibration sensing. Next, in Section 3 we explore the physical insights that enable our approach. In Section 4 we provide a detailed description of our hand washing monitoring approach. Then, in Section 5 we present our real-world experimental evaluation. Finally, in Section 7 we discuss future work and summarize our work.

2 RELATED WORK

In our work, we monitor hand washing activities using structural vibration sensing. This work is primarily related to two main areas of research: 1) hand washing monitoring approaches, and 2) approaches using cepstrum analysis for feature extraction. In this section we will explore relevant prior works in each of these areas and discuss the research gaps that our work addresses.

2.1 Hand Washing Monitoring Approaches

Hand washing monitoring is a very active research area in the medical domain, particularly in light of the COVID-19 pandemic. Traditionally, the most prominent approach for monitoring hand washing in medical settings is observation-based, where staff migrate through the hospital/care facility and mark instances of proper compliance with hand washing policies [9]. However, these approaches require designated staff and it is not practical or possible for these staff members to directly observe all instances where hand washing is required.

To address the limitations of observation-based approaches, numerous sensing-based approaches have emerged. Vision-based approaches rely on motion tracking to verify compliance with hand washing recommendations [4, 26], or are combined with ultraviolet (UV) lights to determine how well hands were cleaned [48]. These vision-based approaches, however, are limited by requiring a direct and unobstructed line-of-site at the hand washing station, and raise privacy concerns. Other sensing-based approaches utilize wearable sensing including smart-watches and RFID tags to both track hand washing-associated hand/wrist motion as well as proximity-based detection (i.e., determine when someone is at the hand washing station) [40, 47, 58, 61]. In real-world settings, these wearable-based approaches are limited due to the coarse-grained information (RFID) and/or the requirement that staff wear or carry a device at all times and keep the device charged. Lastly, acoustic-based systems have been introduced for monitoring activities of daily living (ADLs), with some showing promise for monitoring bathroom activities including hand washing[12]. This system, however, was limited to general "hand washing" and did not distinguish between different phases of hand washing (e.g., soap use, rinsing hands, etc.). Further, acoustic-based systems are sensitive to the presence of ambient noise such as people speaking, objects falling, and operation of machines/medical devices (e.g., beeping noises from monitoring machines, etc.). As such, in clinical settings, many of these acoustic-based sensing systems would be significantly affected by ambient noise and have lower performance.

To overcome these limitations, our approach leverages structural vibration sensing to enable passive and continuous monitoring of hand washing activities. Structural vibration sensing has been shown to accurately monitor indoor human occupants' activities [6, 51, 53, 54], identity [55, 56], location/presence [2, 16, 39, 43–46, 50, 52, 59, 60], and gait health [14, 17–21, 33, 36]. These prior works, however, focus on general occupant activities and information and do not account for similarity in signals due to hand washing activities, nor do they account for signal characteristics with concurrent activities. In this work, we address these research gaps

through a cepstrum analysis-based approach which enables robust hand washing monitoring in a variety of indoor settings and across different persons.

2.2 Cepstrum Analysis-based Approaches

Cepstrum analysis is a common approach for analyzing the periodicity of time series signals. It is a prominent approach in the fields of speech/natural language processing (NLP), earthquake/seismic analysis, and in the medical field for digital signals such as Electromyography (EMGs) [62]. In natural language processing, works have used cepstrum-based techniques for differentiating between languages [27, 41, 73], speech analysis [70], and emotion recognition [63]. In the earthquake/seismic analysis domain, researchers have shown the potential for cepstrum coefficients to be used as features for determining the locations and characteristics of seismic events, as well as distinguishing them from other high intensity events such as quarry blasts [3, 8, 28, 72]. In the medical domain, cepstrum analysis is typically associated with processing imaging and/or time series signals, and has shown promise for applications in using EMGs for activity monitoring [32], neurological signal monitoring [75], detecting neuromuscular diseases [15], and to assist with analyzing ultrasound images [68, 71].

In this paper, we leverage cepstrum analysis to differentiate between structural vibration signals due to various hand washing activities. This represents a new domain and application for cepstrum analysis. The challenge with this approach is that many of the hand washing activities generate similar vibration responses, and may overlap each other. By extracting cepstrum-based signal features, our work is able to isolate signal components that differ in each activity and detect/classify hand washing activity with high accuracy.

3 BACKGROUND AND PHYSICAL INSIGHTS

To enable our hand washing activity monitoring system, our work combines data driven approaches with physical insights. In this section we explore these underlying physical insights and discuss how they assist with addressing the primary research challenges of our work. First, we describe the structural vibration sensing system used in our approach and the physical insights that enable us to extract hand washing activity information from the vibration signals (Section 3.1). Then, in Section 3.2, we present an overview of cepstrum analysis and the physical insights related to how it can be used for differentiating between similar and overlapping hand washing activities.

3.1 Structural Vibration Sensing

As previously discussed, our hand washing monitoring system uses structural vibration sensing to detect and measure the duration of hand washing activities. The main intuition behind this approach is that, when individuals conduct hand washing activities, their movements excite the surrounding structure, causing it to displace. In a linear-elastic structure (which we assume to be the case for any sink structure), the structure then restores to its original state. Repeated cycles of displacement-restoration result in vibrations of the structural material.

In this work, we are using the insight that we can measure these vibration responses due to handwashing activities to detect and measure the duration of the activities themselves, in an *inverse* manner. The foundation of this insight builds off of the common convolution integral, which is used to describe the relationship between a structure's vibration response, the structure's properties, and the forcing/excitation function. This convolution integral is given by the following expression [67, 69]:

$$x(t) = h(t) * f(t) \tag{1}$$

where $x(t)$ is the time history of the structural vibration response, $h(t)$ is the structure's impulse response function (i.e., a characterization of its dynamic properties), $f(t)$ is the vibration forcing/excitation function (i.e., the hand washing activities in our case), and $*$ is a symbol representing the convolution integral. From this expression, we can observe that, if the structure's impulse response function remains constant, we can record vibration responses

due to various hand washing activities, learn the differences in the ensuing signal from each (i.e., differences in $f(t)$), and generate a model which detects and classifies each hand washing activity given a structural vibration response signal.

3.2 Cepstrum Analysis

As discussed previously, cepstrum analysis extracts the periodicity of the frequency spectrum of a signal. In this section, we provide a brief overview of how cepstrum analysis can be used to extract signal features, and discuss how these cepstrum-based features enable our hand washing monitoring approach.

The concept of cepstrum analysis was first developed by Bogert as a way to study seismic signals. Cepstrum coefficients were defined as the power spectrum of the log of the power spectrum of a signal [5, 49, 62]. This was later revised by Oppenheim and Burgess to be the inverse Fourier transform of the log of the Fourier Transform of a signal [10, 62]. In this work, we use the latter definition and use the real-valued cepstral coefficients as signal features, as such, the cepstrum coefficients of a given vibration signal are obtained using the following expression:

$$C_r = \mathfrak{R}\left\{\mathcal{F}^{-1}\left\{log\left|\mathcal{F}\left[x(t)\right]\right|\right\}\right\} \tag{2}$$

where C_r are the real-valued cepstral coefficients for the vibration response $x(t)$ as a function of the time, t, $\mathcal{F}$ is a Fourier transform operator, and $\mathfrak{R}$ is an operator denoting the real values from the expression.

We combine the expressions from Equation 1 and Equation 2 by leveraging the properties of the convolution integral and Fourier transform operations, similar to the approach taken in [62]. For our work, we make an approximation of the full derivation by using the absolute values of the Fourier transform of the measured vibration response. As a result, we observe that real-valued cepstral coefficients of the measured vibration response $x(t)$ are a function of the structure's dynamic properties and the forcing/excitation functions ($f(t)$). In the case of hand washing activity-induced vibrations, we can then infer that the different activities will each have their own forcing function, and, therefore, the real-valued cepstral coefficients for each will differ.

In this work, we leverage this insight that different hand washing activities will generate different real-valued cepstral coefficients to train a model which uses cepstral coefficients as features and outputs the associated hand washing activity for a given time domain vibration signal. Further details for this approach are in Section 4.3.

4 HAND WASHING MONITORING APPROACH

Our approach for monitoring hand washing activities leverages cepstrum-based signal features to detect the presence of activities, classify them, and record their duration. This enables our system to determine if someone is washing their hands and then if they are adequately following the WHO and CDC recommendations. Our approach consists of three main modules: 1) a sink vibration sensing module (Section 4.1), 2) a hand washing event detection module (Section 4.2), and 3) a cepstrum-based hand washing activity classification module (Section 4.3). The second and third modules are each part of our hierarchical classification approach which contains three independent models for determining the presence of hand washing, and then which activity is occurring at any given time. Figure 2 shows an overview of our approach including each module and the hierarchical classifier. In this section, we explore each of these modules in more detail and describe our overall approach for monitoring hand washing activities using structural vibration sensing.

4.1 Sink Vibration Sensing

The first module of our approach measures the structural vibration responses due to hand washing activities. In this section, we provide an overview of the sensing modality and describe how we collect and process the structural vibration signals.

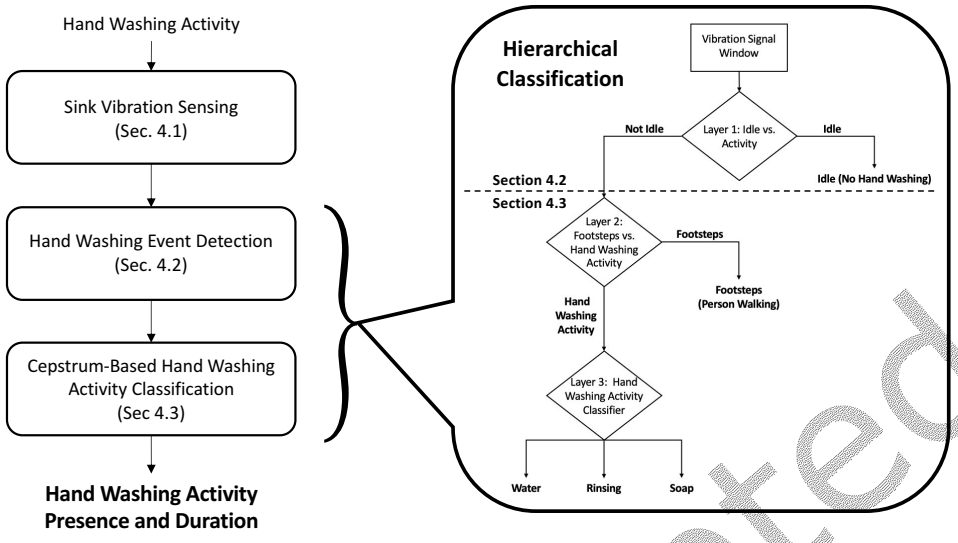


Fig. 2. Our cepstrum-based hierarchical hand washing monitoring approach. The left hand figure shows an approach overview and the right hand figure shows the hierarchical classification framework.

To measure these vibration responses, our system uses geophone sensors. Geophone sensors are low-cost vibration sensors which measure the vertical velocity of vibration. These sensors are mechanical sensors which rely on the vertical displacement of a suspended mass within the sensing element [29]. As such, they require a coupling with the structure. To accomplish this, we adhere the geophone sensors to the structure (i.e., sink) using bees wax; this ensures that the sensors will record all of the vertical vibrations that the sink structure undergoes from each of the hand washing activities. The benefit of this system is that the geophone sensors can be retrofitted onto any existing sink structure with little effort.

To increase the resolution of the vibration signals, we amplify them using a variable gain operational amplifier. The gain for the sensing system can be manually calibrated when initially deployed based on observed magnitudes of the vibration responses in that area. Ideally, the operational amplifier gain is maximized to provide the most signal resolution while also preventing clipping of signals (i.e., when signal values exceed the limit that the analog-to-digital converter can read). Typical values for amplification range from 100-1000X. Once the structural vibrations are induced and measured by the geophone sensors, they are converted to a digital signal and transmitted to a computer for further processing. An example of a typical sensor configuration is shown in Figure 3.

4.2 Hand Washing Event Detection

The first layer of our hierarchical classifier determines if there is any type of hand washing activity in the region. We describe this process as "hand washing event detection". In this section, we describe the process by which we distinguish hand washing activities from "idle" behavior (i.e., when no hand washing is occurring).

4.2.1 Vibration Signal Preprocessing. The signal-to-noise ratio (SNR) of hand washing activity-induced structural vibration signals has a large variance depending on the characteristics of the sink and surrounding structure. As such, it is often necessary to remove/reduce the ambient noise levels in the measured vibration signals. In particular, the activities that take place away from the sink (i.e., footsteps) typically have a low SNR because the

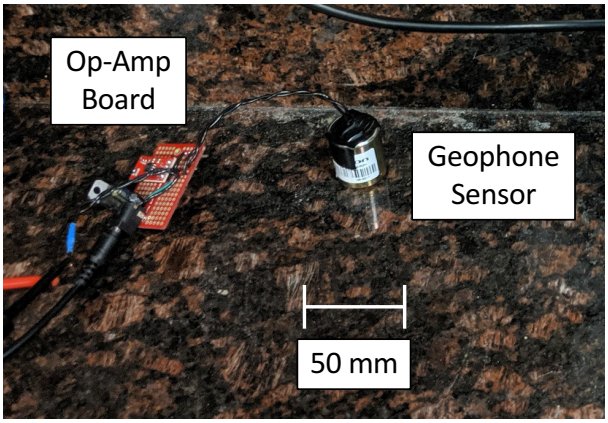


Fig. 3. Example of the geophone sensing system used in our approach.

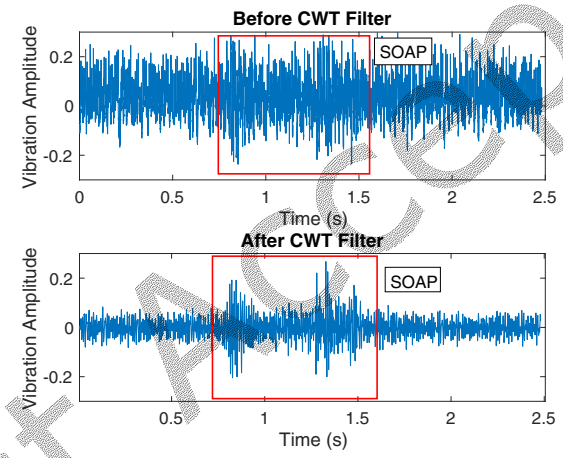


Fig. 4. Example of the wavelet-based filtering on hand washing-induced vibration signals. The top figure shows the raw vibration data due to a soap dispenser being used, and the bottom figure shows the same signal after being filtered. Note that after filtering, the soap activity is easily distinguishable from the ambient noise levels

vibration signals have to travel through multiple mediums to be recorded by the vibration sensors (e.g., from the floor to a wall to the sink). To increase the signal SNR, our approach uses a data preprocessing step where we reduce signal noise using a continuous wavelet transform (CWT)-based filtering approach.

The continuous wavelet transform is a time-frequency signal transformation that is well-suited to non-stationary signals such as those induced by hand washing activities [1, 30, 31]. In our preprocessing step, we first decompose the vibration signal window using a CWT with a Morse mother wavelet [37]. We choose a Morse wavelet based on the insight that Morse wavlets provide a broad range of wavelet shapes and are well suited to the varying excitation types encountered from hand washing activities (i.e., impulsive and continuous excitations) [37]. In this work, we use a sliding window with a size of 0.5s to extract features. This window size is

chosen empirically based on the observed duration of impulsive activities (e.g., footsteps and soap), and to ensure adequate data resolution for the continuous activities (e.g., water and rinsing).

Once the signal window has been decomposed, we filter it by selecting a frequency band and performing an inverse wavelet transform with only that frequency band. This frequency band can be chosen empirically during initial calibration and setup based on the observed vibration responses. For our work, we have selected a frequency band of 70 to 450Hz based on the observation that this band contains the most information about the activities of interest (i.e., footsteps, soap, rinsing, water running), and also removes/reduces the amount of ambient noise in the signal. An example of this CWT-based filtering is shown in Figure 4, where a raw vibration signal generated from a soap excitation is shown both before and after filtering. After filtering, we observe a much higher SNR and the soap activity is easily distinguishable from the ambient noise levels. Note that, for some structures, if the observed SNR is high, this filtering step may not be necessary and the raw signals can be used for the hierarchical learning approach below. The ensuing filtered signals are then used in the first two stages of our hierarchical learning to determine the presence of any hand washing activity, and then to distinguish footstep responses from other activities that occur on/around the sink structure. Figure 5 shows an example of the filtered vibration signal for each of the hand washing activities. Note the similarity between the "Footstep" and "Soap" activity and the "Rinsing" and "Water" activity. In the following sections, we describe how our cepstrum-based hierarchical learning approach overcomes the challenges associated with these similar vibration responses.

4.2.2 Idle Classification. To classify "idle" and "activity" events, our system uses a binary Support Vector Machine (SVM) classifier [13, 25]. We choose to use a SVM classifier based on the insight that SVM classifiers are well-suited to small datasets and those without a well-defined feature distribution. Based on our preliminary observations, we infer that our data fits these two descriptors well. To train this classifier, we extract features based on the standard deviation of the signal window. In this work, we have a limited amount of data, and, therefore, use the sample standard deviation of the window (as opposed to the population standard deviation) [35]. We then use the sample standard deviation from each sensor in the sink area as the feature values from each signal window and the system outputs a label of either "idle" or "activity". If the label is "idle", the system outputs this label as the final prediction for the window, otherwise, the data is sent to the next layer in the hierarchical classifier.

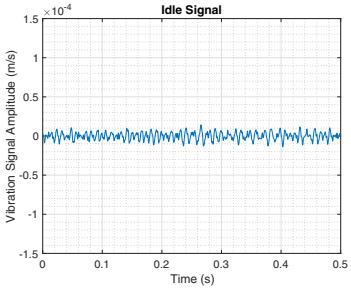
4.3 Cepstrum-Based Hand Washing Activity Classification

In the third module of our approach, our system uses the final two layers of the hierarchical classifier to determine if detected activities are footsteps, soap, water, or rinsing. In the following section, we describe this process in more detail.

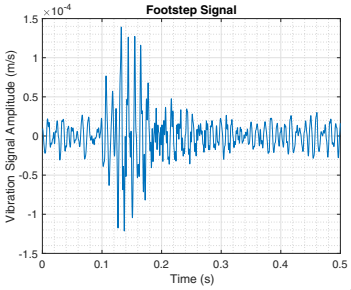
4.3.1 Footstep Classification. Once a signal window has passed through the first layer of our hierarchical classifier and did not get classified as an "idle" event, it is passed to the next layer, which focuses on distinguishing footstep activities from other hand washing activities. This layer, therefore, partially addresses the challenge related to activities on different surfaces/structures discussed above. By comparing footsteps (which occur on the ground) to other hand washing activities (which occur at/around the sink), we can better characterize the differences in their signals resulting from the different mediums.

Detecting footstep events is an important component of monitoring hand washing activity so that the system can tell if a person is walking to/from the sink/hand washing area. Additionally, in our prior work, we have used footstep-induced structural vibration responses to uniquely identify and track individuals [42, 56]. Therefore, we can combine this work and our prior work to track individuals and determine whether they have properly adhered to hand washing protocol (e.g., in hospital settings).

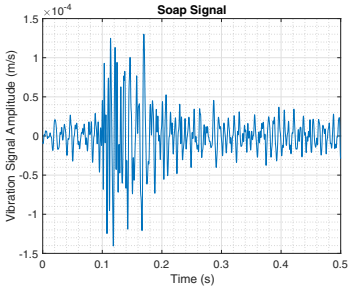
In this layer of our hierarchical learning approach, we classify vibration signal windows as footsteps or "hand washing activities" using a binary SVM model. For features, we extract cepstral coefficients from the filtered



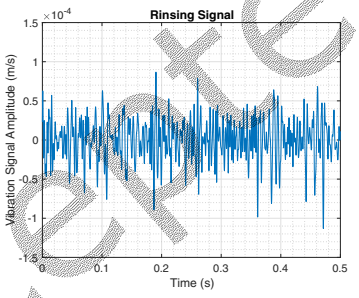
(a) Example of an "Idle" signal collected by our sensing system. In these cases, our system detects that there is not any hand washing activity.



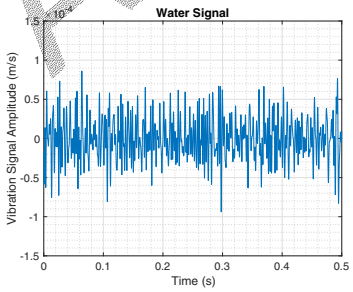
(b) Example of a "Footstep" signal collected by our sensing system.



(c) Example of a "Soap" signal collected by our sensing system.

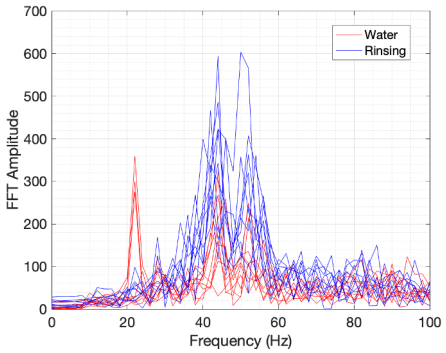


(d) Example of a "Rinsing" signal collected by our sensing system.

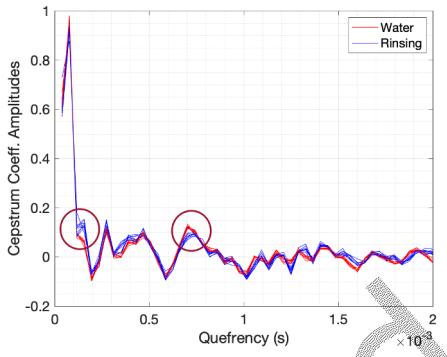


(e) Example of a "Water" signal collected by our sensing system.

Fig. 5. Example of vibration signals collected by our sensing system for each of the hand washing activities. Note the similarity between the impulsive signals (e.g., "Footsteps" and "Soap") and the continuous signals (e.g., "Rinsing" and "Water"). Our cepstrum-based features highlight the differences between these similar signals to enable accurate and robust detection and monitoring of hand washing activities.



(a) Fourier transform-based features for rinsing and water running activities.



(b) Cepstrum-based features for rinsing and water running activities. Note the clear separation between each activity in the two areas highlighted by the red circles.

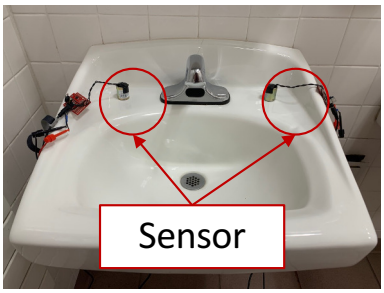
Fig. 6. Fourier transform- and Cepstrum-based features for rinsing and water running activities. We note that there is little or no separation between these activities using a Fourier transform alone, while cepstrum coefficients highlight the differences in periodicity of the two activity's signals, enabling accurate detection and classification of those activities.

signal in the current window. We use these filtered signals based on the insight that the wavelet-based filter helps to reduce the noise in the system, and provides more information about the signal variations due to each type of excitation. This is especially important for situations like the one shown in Figure 4, where the SNR is so low that activities are barely observable. We calculate the cepstral coefficients using the process outlined in Section 3.2 and use a linear kernel when training the SVM model. We choose a linear kernel based on the observation that the data is largely separable in this layer, and the linear kernel has lower model complexity as compared to higher order kernels (which reduces the risk of overfitting) [25]. Similar to the approach taken in the first layer, we combine the cepstral coefficient features for each sensor in the sink/sensing area for our training and predictions.

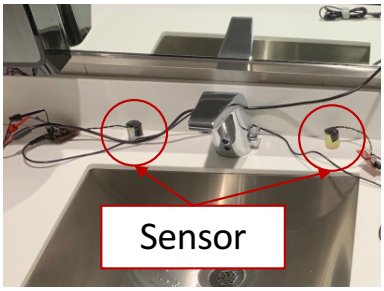
The output of this hierarchical model layer is the binary prediction of "footstep" or "hand washing activity". If the layer prediction is a footstep, then this is the final system output and the window is classified as a footstep. Otherwise, the system moves to the third and final layer of our approach to determine if the hand washing activity is best classified as "soap", "water", or "rinsing".

4.3.2 Hand Washing Activity Classification. The final layer of our hierarchical learning approach classifies hand washing activities as "water", "rinsing", and "soap". These activities are defined by water running directly into the sink, water running over a person's hands to rinse them, and a person pumping a soap dispenser. As previously discussed, there is a possibility that individuals may leave the water running while using the soap dispenser - in these cases our primary interest is in detecting the usage of soap. As such, these activities are labeled as "soap" for the purpose of model training and prediction.

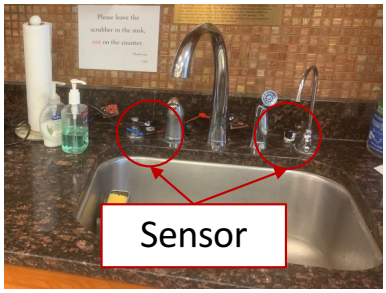
We distinguish between each of the aforementioned hand washing activities using a multi-class SVM classifier. Similar to the previous layer of our approach, we use cepstrum-based features in this layer by calculating the cepstrum coefficients for the signal window using Equation 2. By using cepstrum-based features, our approach is better able to distinguish between similar vibration responses (i.e., rinsing and water). Figure 6 shows an example of the cepstrum-based feature extraction for several "rinsing" and "water" signals. In Figure 6a, we observe that the Fourier transform fails to find unique signal components for each activity (there is a spike for "water" around



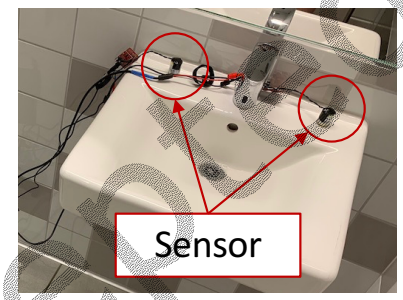
(a) Porter Bathroom experimental setup.



(b) Doherty Hall experimental setup.



(c) Porter Hall kitchen experimental setup.



(d) Tepper Hall experimental setup.

Fig. 7. Experimental setup for each of the four experimental locations. In each location, vibration sensors were mounted on the sink structure.

20 Hz for some signals, but many others do not have this spike, so it cannot be used for training a classifier). In contrast, with cepstrum-based features, our approach is able to find consistent separation between the two similar activities. Figure 6b shows an example of these features, with the red circles highlighting instances where there is clear separation between each activity.

For the SVM model, we use a Gaussian kernel based on the observation that our feature values are not linearly separable in the original feature space and due to the insight that a Gaussian kernel has lower model complexity than polynomial or non-parametric kernels, therefore reducing the risk of overfitting [38]. With this model, our system outputs the final window label of "water", "rinsing", or "water". Using the approach described above, our system is able to detect and differentiate hand washing activities using structural vibration sensing. Through a sliding window, we can then monitor the duration of each activity (i.e., count the number of consecutive windows with "rinsing" or "water". This allows our system to record the presence of hand washing as well as its duration to ensure that individuals follow the guidelines for proper hand hygiene set by the Centers for Disease Control and the World Health Organization [24, 34, 57].

5 EXPERIMENTAL EVALUATION

To validate the performance of our hand washing monitoring approach, we conducted real-world hand washing experiments with 4 total participants and across 4 different experimental locations. In this section, we discuss

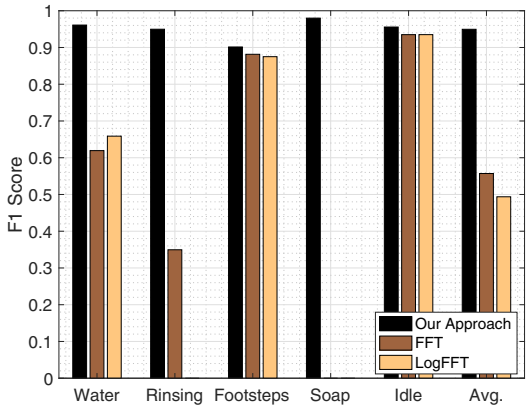


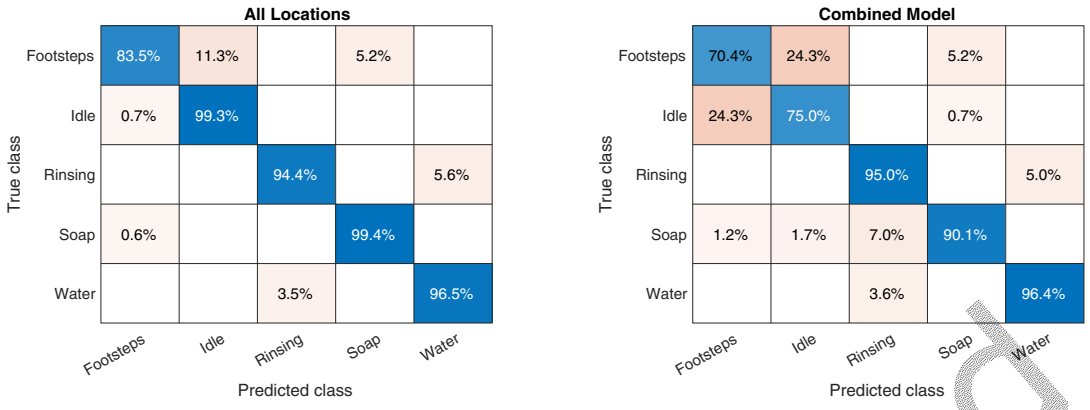
Fig. 8. Total performance of our approach compared to the baseline approaches. We observed an 8.8X and 10.2X error reduction respectively over the FFT and logFFT approaches. The baselines effectively detected footsteps, but struggled to detect classes that may overlap or have a wide range of vibration response, such as soap, water, and rinsing.

the overall performance of our approach (Section 5.2), then with regards to different structures (Section 5.3), different amounts of training data (Section 5.4), and through an uncontrolled hand washing experiment involving 3 different individuals (Section 5.5).

5.1 Experimental Setup

As described above, we evaluated our system in four experimental locations. For each location, vibration signals were collected using two sink- or counter-mounted SM24 geophone sensors [29]. In this work, geophone sensors are chosen over other vibration-based sensing modalities (such as acoustic sensors and accelerometers) due to their low installation cost, and based on observations that they are sensitive to the frequency bands that are typically excited by human activity [54]. The operational amplifier gain was separately calibrated for each sensor and for each location to maximize the signal resolution while also preventing clipping (typical range of amplification was 100-1000X). For data collection, we selected a sampling frequency of 25600 Hz. This sampling frequency was selected to ensure adequate time and frequency resolution for monitoring hand washing activities, and so that, in future work, this approach can be combined with our prior work for occupant localization and identification [42, 56], which requires a higher sampling frequency.

The four experimental locations were: 1) a bathroom in the Porter Hall building at Carnegie Mellon University (Figure 7a), which consists of a wall-mounted ceramic sink and separate wall-mounted soap dispenser, 2) a bathroom in the Doherty Hall building (Figure 7b), which consists of a counter-mounted stainless steel sink and separate wall-mounted soap dispenser, 3) a department kitchen area in Porter Hall (Figure 7c), which consists of a cabinet counter-mounted stainless steel sink and counter-mounted soap dispenser, and 4) a bathroom in the Tepper Quad building (Figure 7d), which consists of a wall-mounted ceramic sink and separate wall-mounted soap dispenser. At each location we collected an average of 12 repetitions of data for each hand washing activity (with approximately 10 seconds of data in each repetition) for an average total of approximately 120s of data per activity per location that we used for training and testing our approach performance. The following sections provide a summary of the evaluation results with respect to varying performance factors.



(a) Scenario 1 achieved high accuracy for all classes. Footsteps, which had the lowest accuracy, occurred furthest from the sensors and had the highest SNR.

(b) Scenario 2 resulted in reduced performance, which we attribute to different structural attributes causing different data distributions for each location.

Fig. 9. Confusion Matrices showing the prediction accuracy for all locations, with separate models trained independently for each location in 9a (Scenario 1), and with a shared model trained on the combined location data in 9b (Scenario 2).

5.2 Hand washing Monitoring Performance

We first consider the performance of our approach across all experimental locations. In this evaluation, we combine the results from each of the four experimental locations when: 1) independent models are trained and tested for each location; and 2) when all of the training data from each location is combined to create one model for testing. In each scenario, we are using the controlled experimental data collected with one experimental participant. Additionally, we evaluate each scenario using a 5-fold cross validation of the available data, where the data is randomly partitioned into 5 train/test splits, with each split using 80% of the data for training, and 20% of the data for testing. In this way, all of the available data is both used for training, and for testing.

5.2.1 Independent Models Performance. For Scenario 1, we independently train and test our approach for each of the four experimental locations and compute the per-class accuracy for each of the hand washing activities of interest (footsteps, soap, water, rinsing) as well as how well our model distinguishes these activities from "idle" states. In addition, we compare our results to two different baseline approaches: "FFT" and "logFFT". These baseline approaches use the same detection layer of our hierarchical learning (i.e., with data standard deviation as a feature), but for the "FFT" baseline, we compute the model feature values for the other two layers as the Fourier Transform amplitudes at each discrete frequency (frequency resolution of approximately 2Hz for a 0.5s data window). Then, for the "logFFT" baseline, we instead compute the feature values as the log of the Fourier Transform amplitudes at each discrete frequency. These two baseline approaches allow us to compare the cepstrum features from our approach with similar features and show that the cepstrum-based features are more suitable for differentiating hand washing activities.

When comparing to the baseline approaches (Figure 8), we observe a significant performance increase with our approach over each of the two baseline approaches for each class, and on average across each class. In this figure, we are showing the per-class F1 scores and an average F1 score. The F1 score is a common metric for evaluating classification performance and is defined by the following expressions:

$$F1_n = 2 * \frac{Precision_n * Recall_n}{Precision_n + Recall_n} \tag{3}$$

$$Precision_n = \frac{TruePositive_n}{TruePositive_n + FalsePositive_n} \tag{4}$$

$$Recall_n = \frac{TruePositive_n}{TruePositive_n + FalseNegative_n} \tag{5}$$

where $TruePositive_n$ represents the number of correctly identified data windows for activity n, $False\ Positive_n$ represents the number of data windows belonging to activity n, but classified as a different activity, and $FalseNegative_n$ represents the number of data windows belonging to a different activity, but identified as activity n.

In this way, we compute F1 scores for each activity and observe that our approach achieves F1 scores of 0.96 (Water), 0.95 (Rinsing), 0.90 (Footsteps), 0.98 (Soap), and 0.96 (Idle). Then, on average, we observe an F1 score of 0.95, which is a 8.8X error reduction over the "FFT" baseline (0.56 avg. F1 score), and a 10.2X error reduction over the "logFFT" baseline approach (0.49 avg. F1 score). These results show that our approach is able to accurately monitor hand washing activities in a variety of settings, and that the cepstrum-based features used in our approach overcome the challenge of similar vibration responses from different hand washing activities.

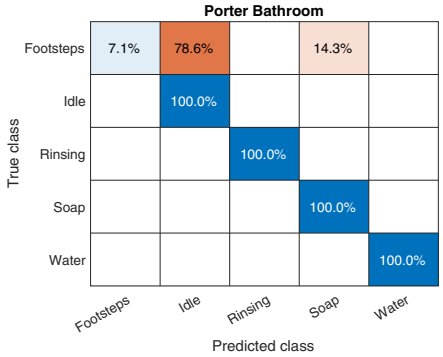
Figure 9a shows a confusion matrix summarizing the total performance for the independent models across each of the 5 classes. From this figure, we observe that our approach achieves a very high accuracy for each class associated with hand washing (i.e., soap, water, rinsing), and a high accuracy for distinguishing footstep responses as well.

5.2.2 Combined Model Performance. For the second performance scenario, we evaluated the accuracy of our model if the data from each of the four experimental locations was combined to create one unified hand washing event detection/classification model. Figure 9b provides a summary of the model performance. From this analysis, we observe that our model achieves a high accuracy for the sink area activities (95.0% rinsing, 96.4% water, 90.1% soap), but a lower accuracy for detecting footsteps and distinguishing them from "idle" data windows (70.4% and 75.0%, respectively). In addition, there is more confusion between "soap" and "rinsing" compared to independently trained models discussed above. It is likely these decreases in model performance are due to the differences in data/feature distribution for each location. When a different sink/structural material or configuration is present, it changes the dynamic properties of that structure, and, therefore, the ensuing vibration signals. As a result, a "footstep" or "soap" response in one structure may be significantly different than one in a different structure. In our future work, we plan to explore ways to transfer models across different structures, which will reduce this effect and improve model performance. We discuss this in more detail in Section 6.1.

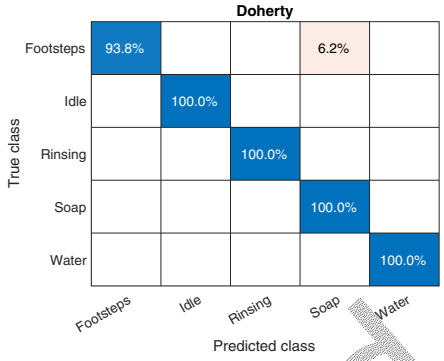
5.3 Robustness to Different Structures

In this section, we take a detailed look at the performance of our approach in each of the four experimental locations. Similar to the approach taken in Section 5.2.1, we independently trained and tested for each experimental location. Also, for each model, we perform a 5-fold cross validation to iteratively train and test on all of the data. Additionally, we compared our results with the same two baseline approaches, "FFT" and "logFFT".

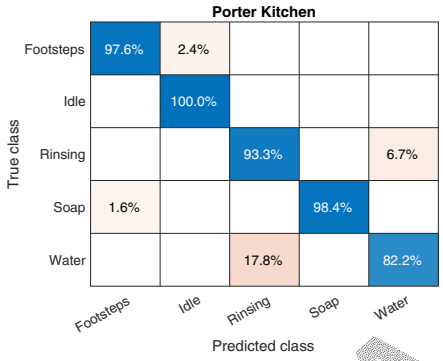
Figure 10 provides a confusion matrix for each structure summarizing the results of this evaluation. Of all the locations, the Doherty Hall location had the best overall performance, with 100% accuracy for every activity except "footsteps". In this structure, footsteps were occasionally confused with "soap", which is likely due to the fact that each of these represents a more impulsive excitation, and, therefore, has a similar vibration response. Despite these similar responses, we are still able to identify both "footsteps" and "soap" with high accuracy. In



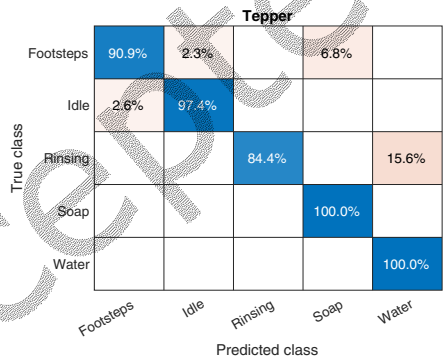
(a) Our approach performance at the Porter Hall bathroom location.



(b) Our approach performance at the Doherty Hall bathroom location.



(c) Our approach performance at the Porter Hall kitchen location.



(d) Our approach performance at the Tepper Quad bathroom location.

Fig. 10. Performance of our cepstrum-based hierarchical classification approach for detecting and classifying hand washing activities in four different structures.

both the Porter Hall Kitchen and Tepper Quad locations, we observe some confusion between the "rinsing" and "water" activities, which likely resulted from the similar nature of their responses as well. In particular, the flow rate of the water and the amount of hand motion while rinsing can cause each of these activities to exhibit a wide range of vibration responses and confusion between the two activities. We discuss our plans to further characterize characteristics such as water flow rate in Section 6.2.

Of all the four experimental locations, the Porter Hall Bathroom location displays the worst performance of detecting and classifying footstep responses and soap activities. In this location, we note that the signal-to-noise ratio (SNR) is very low, to the point where footstep and soap activities are not visible in the raw signals alone. As such, it is difficult for the model to accurately model them and distinguish them from other activities. In particular, due to the very low SNR, the footstep activities are effectively always classified as "idle". To reduce/eliminate these types of classification errors from our system, part of our future work will explore combining this work

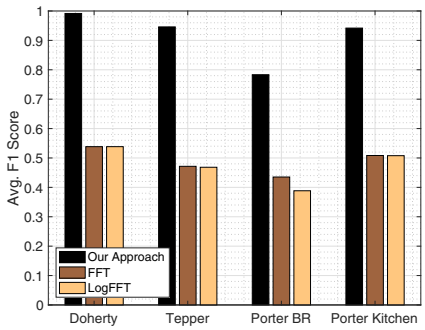


Fig. 11. Our approach compared to the two baseline approaches in each of the four experimental structures. In each structure, our approach significantly outperforms the baseline approaches for average classification performance.

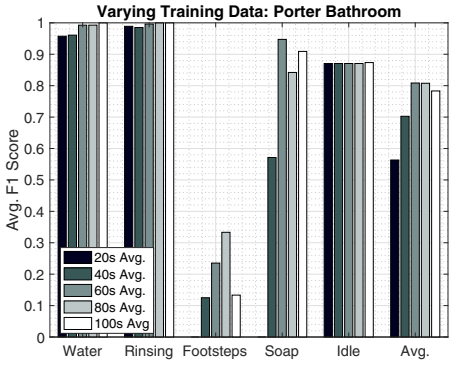
with our prior work in footstep detection/classification using floor-mounted vibration sensing [43]. For this work, we have focused on placing sensors only on the sink structure, but, if this system were deployed throughout a building, the sink-mounted sensors could be networked with floor-mounted sensors to improve the detection and classification of footstep activities.

In addition, we compared the average per-class F1 score for our approach in each experimental location with the average per-class F1 score obtained using each of the two baseline approaches. Figure 11 shows a summary of this comparison. We observe that, despite the lower performance in the Porter Hall Bathroom for footsteps, and Porter Kitchen/Tepper Quad for distinguishing rinsing and water, our approach significantly improves over the two baseline approaches. We observe average error reductions over the two baseline approaches of 46X (Doherty), 10.6X (Tepper), 2.7X (Porter Bathroom), and 8.2X (Porter Kitchen), which indicates that our cepstrum-based approach is effective for detecting and distinguishing between hand washing and robust to different sink/structural environments.

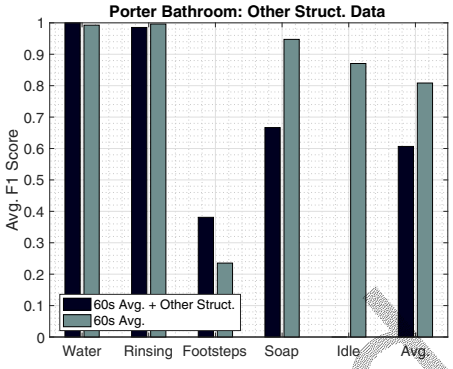
5.4 Sensitivity to Training Data Availability

In this section, we explore the sensitivity of our model performance to the amount of available training data. As previously described, we train our models using 80% of the available data, and test with 20% using a 5-fold cross validation. This equates to an average of approximately 100s of training data for each class (idle, footsteps, soap, water, rinsing). To understand the sensitivity of the model performance to the amount of training data, we randomly select subsets of this training data at increasing ratios. We consider 20% (20s average/class), 40% (40s average/class), 60% (60s average/class), and 80% (80s average/class), and compare the 5-fold cross validation performance of each with the performance when the entire set of training data is used. Additionally, we recognize that the sensitivity to the amount of training data may vary in different structures/locations; as such, we perform this analysis separately for each of the four experimental locations.

Figure 12a, Figure 13a, Figure 14a, and Figure 15a summarize the results of this analysis for each of the four experimental locations by showing the per-class average F1 scores for each level of training data, as well as the average F1 score across all classes. As expected, the model performance generally increases with each additional amount of training data. However, for the Porter Hall bathroom location, we note that the overall (average) performance is slighly lower for the total data (100s) than for the lower amounts of data. In particular, the model performance for “footsteps” decreases from 80% of the data to 100% of the data. Additionally, the “soap” performance with 60% of the data is higher than with 100%. This indicates that some of the training data is

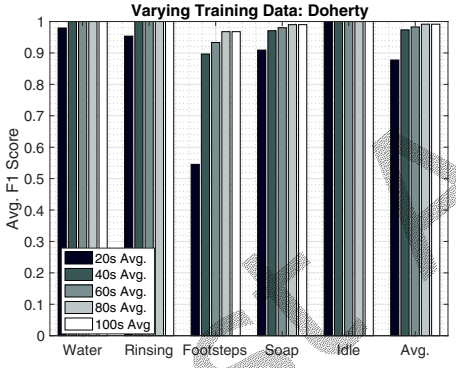


(a) Performance in the Porter Hall bathroom location with varying amounts of training data.

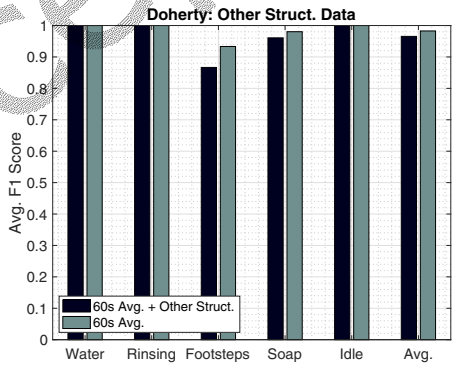


(b) Performance in the Porter Hall bathroom location with 60% of training data and with the addition of training data from other locations.

Fig. 12. Performance in the Porter Hall bathroom location with respect to the amount of training data used to train the hand washing monitoring model.



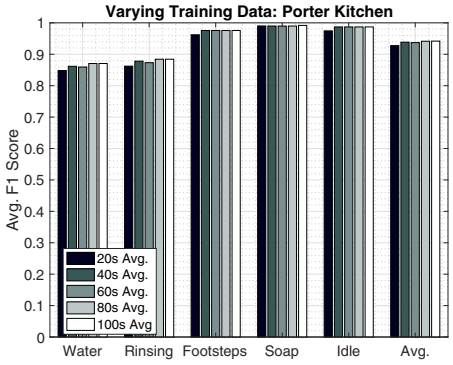
(a) Performance in the Doherty Hall bathroom location with varying amounts of training data.



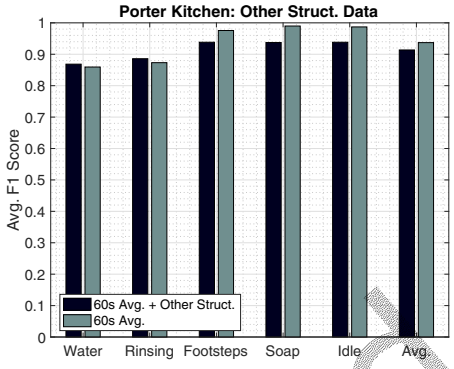
(b) Performance in the Doherty Hall bathroom location with 60% of training data and with the addition of training data from other locations.

Fig. 13. Performance in the Doherty Hall bathroom location with respect to the amount of training data used to train the hand washing monitoring model.

decreasing the model performance. This is likely a result of the comparatively low SNR for footstep responses in this location. If a footstep response has a particularly low SNR, it may contain much of the same information as the "idle" data (i.e., the ambient vibrations), which causes the model to have a different decision boundary between the two classes, resulting in increasing prediction error. There are a number of potential solutions for



(a) Performance in the Porter Hall kitchen location with varying amounts of training data.



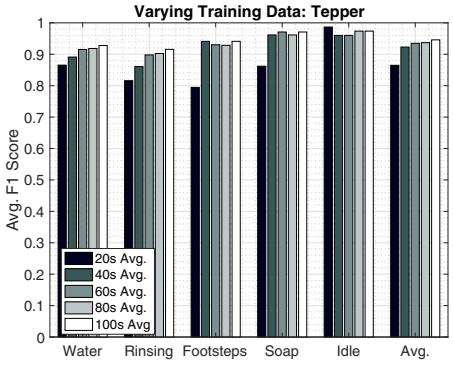
(b) Performance in the Porter Hall kitchen location with 60% of training data and with the addition of training data from other locations.

Fig. 14. Performance in the Porter Hall kitchen location with respect to the amount of training data used to train the hand washing monitoring model.

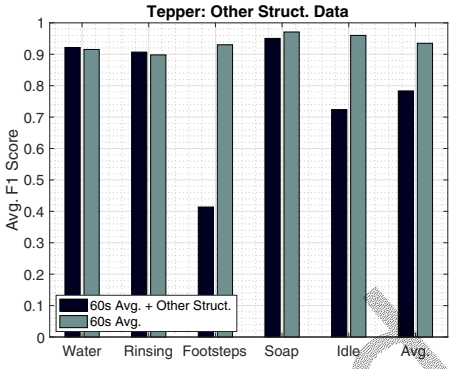
overcoming this challenge of "bad" training data, including the naive approach of introducing additional floor mounted sensors (as described previously), and also to leverage some active learning approaches, such as ones described in our prior work [65, 66], which involve incrementally choosing training data that best describes the class distribution and improves model performance. In our future work, we plan to explore these types of approaches to improve model performance and decrease the training data requirements.

Of additional interest is the observation that, with approximately 60% of the total training data (60s/class avg.), the models in each structure achieve similar accuracy as the models trained with the full set of training data. This observation suggests that as little as 60s/1 min of training data could be collected for each hand washing activity during initial calibration/deployment, and the system can achieve high accuracy for detecting and monitoring hand washing activities with that limited amount of training data. This observation suggests that our system can be easily deployed and scaled in real-world structures with little cost associated with initial calibration/training. We discuss the scalability of our system further in Section 6.1.

We additionally explored the model performance if training data from other structures was added to the training data from the test structure. With this analysis, we explore if the amount of training data can be reduced if additional data from other structures is used for training the hand washing activity models. Figure 12b, Figure 13b, Figure 14b, and Figure 15b show a summary of these results by providing the average F1 scores for each activity, as well as the average across all activities for each model. In this analysis, we considered the 60% model from the previous analysis (given that this achieved nearly the same accuracy as the 100% model), and added training data from each of the other three structures; we then compared the 60% model, and the 60% plus other structures model. From these results, we observe that, in each structure with the exception of the Porter Hall bathroom, the introduction of training data from other structures decreases model performance. This result is consistent with the observations in Section 5.2.2, where the combined model had decreased performance over the independent models. This indicates that the feature data from different structures has a different distribution, and, therefore, does not help with defining the model decision boundaries for the test structure. However, in some instances (e.g., water and rinsing for Tepper and Porter Kitchen), the addition of training data from other locations increases

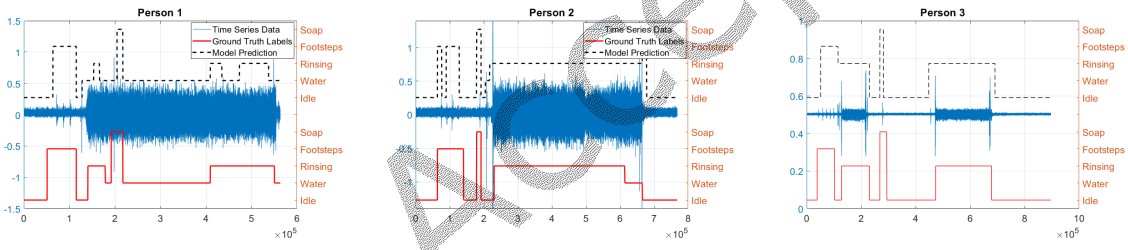


(a) Performance in the Tepper Quad bathroom location with varying amounts of training data.



(b) Performance in the Tepper Quad bathroom location with 60% of training data and with the addition of training data from other locations.

Fig. 15. Performance in the Tepper Quad bathroom location with respect to the amount of training data used to train the hand washing monitoring model.



(a) First person's predictions.

(b) Second person's predictions.

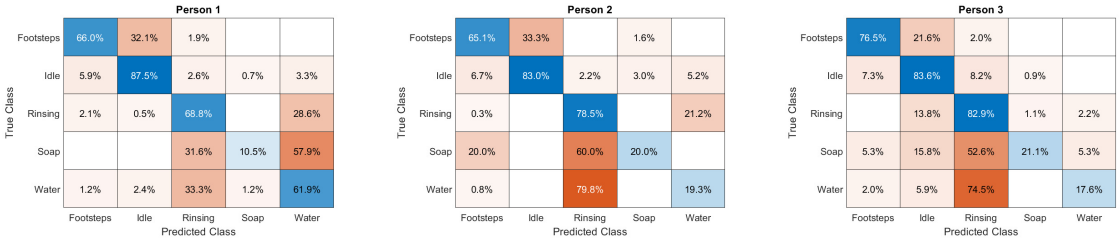
(c) Third person's predictions.

Fig. 16. Examples of time series predictions of our hand washing monitoring approach for each of the three experimental participants. Our approach is able to detect and monitor the duration of hand washing activities from each person.

performance. This indicates that there may be some benefit of additional data from other structures if their data can be transformed into one unified feature space that is transferable across structures/locations (similar to the approach taken in [43]). As discussed above, part of our future work will be to explore approaches to transfer models across structures.

5.5 Robustness to Different People

Our final experimental evaluation involves uncontrolled hand washing experiments with three individuals in our Porter Hall kitchen experimental location. These experiments were conducted in accordance with our approved Internal Review Board (IRB) study (STUDY2018_00000515). Each person was instructed to walk across the floor for several steps to the sink area, wash their hands, and then walk away. This process was completed 10 times by



(a) Overall performance: Person 1. (b) Overall performance: Person 2. (c) Overall performance: Person 3.

Fig. 17. Overall model performance for the time series predictions of the three experimental participants. We note that, in these instances, there is some overfitting of the model for rinsing or soap (depending on the person). This is likely due to the limited training data for each individual.

each person. We then applied our approach to the ensuing time series data from each person to determine the accuracy in detecting and monitoring the hand washing activities. Each person's data was analyzed independently with a 10-fold cross validation where 9 of the 10 repetitions were used for training, and the last repetition was used for testing. This was repeated 10 times so that every repetition was used for both training and testing for each person. Ground truth information was collected using a camera and labeled in 0.5s increments (to match the window size used in our approach).

We compared our results with the time series predictions to the ground truth labels, and also compared the overall prediction accuracy using a confusion matrix for each person. Figure 16 shows an example of the time series data, ground truth labels, and model predictions for each of the three walkers. From these examples, we can observe that there is some difference in hand washing behavior for each person; for example, persons 1 and 2 used soap while the water was continuously running, while person 3 turned the water off to use the soap. Our model was able to adapt to these styles for accurate monitoring of hand washing activities. In addition, we computed the overall time series prediction performance for each person across all 10-fold predictions with respect to the ground truth labels. These results are summarized in the confusion matrices in Figure 17. Overall, the most confusion was with predicting "soap". This is likely due to the observation that, in most instances, the "soap" activity occurred while the water was running in the background. When this occurs, it is difficult to isolate the soap effects in the signal from those corresponding to the water itself. In this work, we address that challenge by labeling the "primary" activity of soap in these overlapping/concurrent activity windows. This approach works for scenarios when the soap is still visible i.e., with high SNR), but has reduced performance in scenarios with low SNR and/or when the water response dominates the signal. In our future work, we plan to address these "overlapping" activity scenarios by modeling the combined effects of concurrent activities to distinguish them from the independent ones and improve overall classification performance.

The best performance was observed with Person 3, where our model had high accuracy for "idle", "footsteps" and "rinsing". However, in each person's results, there appears to be overfitting in the last layer of the hierarchical classifier. As a result, the models tend to predict either "rinsing" or "water" for the majority of the cases of "rinsing", "soap", and "water". We expect that this deviation from the performance of our approach in different structures and with combining all structures is due to these experiments having less training and test data. For example, each individual used the soap dispenser just one time during each repetition (total of 10 data points per person). As a result, it is more difficult for the algorithm to determine an accurate decision boundary, and it tends to overfit to the classes with more training data (i.e., rinsing and/or water). These results also appear to conflict

with the time series predictions shown in Figure 16. This is likely due, in part, to the fact that we are predicting on small windows (0.5s), while the duration of some activities like rinsing and water last for several seconds. As a result, while our approach may generally detect both "rinsing" and "water", it may also alternate predictions for consecutive windows, resulting in a "lower performance". To address this, we plan to incorporate a time series smoothing/updating step to our approach that incorporates the likelihood of different activities at different times. This is discussed in more detail in Section 6.2.

6 DISCUSSION AND FUTURE WORK

In this paper, we presented a novel approach for monitoring hand washing activities using structural vibration sensing. Through real-world experimental validation, we showed that our approach is able to accurately detect, classify, and monitor hand washing activities in a variety of settings, and with multiple different persons. In this section, we discuss some of the assumptions and limitations of our approach, and how those relate to areas of future work. These areas of future work can be categorized into two main sections: 1) large-scale system deployment, and 2) indoor occupant activity monitoring.

6.1 Large-scale System Deployment

One of the primary considerations for a hand washing monitoring system is its ability to be deployed at building-scale in a variety of settings. In this section, we explore some of the assumptions and limitations of our approach with regard to the system scalability, and how our future work will address those limitations.

As discussed in the experimental evaluation, our approach performs best when the system is trained with data collected at each location independently. Additionally, when data from other structures was introduced, it reduced the overall performance of the model. This result implies that our system would require calibration at each new location when it is installed. At the scale of an entire office building or hospital, this calibration requirement may be time-consuming and costly. To address this, we showed that reduced amounts of training data (as low as 60s per activity) can be used to train the model and achieve similar performance. For large-scale deployments, this approach can be used to reduce the training cost.

In addition, our future work aims to improve the model performance for large-scale deployments in several ways: 1) *floor-mounted sensors*: as discussed in the evaluation section, we plan to leverage our prior work using floor-mounted vibration sensors to create a large-scale deployment with both floor-mounted and sink-mounted sensors. This combination will enable our model to achieve higher accuracy for detecting and classifying footstep responses, and can also use those footstep responses for tracking/localizing and identifying the individuals. This is particularly useful in hospital settings for monitoring each employee's adherence to proper hand hygiene practices. 2) *model transfer*: another approach for improving large-scale performance and reducing training requirements is to develop an approach for transferring models across structures/locations. This will enable a model to be trained in one location and then used for each new location without requiring extensive re-training. In our prior work, we have shown that this can be useful for detecting footstep-induced vibration signals and differentiating them from other impulsive signals (e.g., objects falling, doors closing, etc.) [43]. In our future work, we plan to adapt this or other, similar approaches to transfer hand washing activity monitoring models across structures/locations. 3) *active/online learning*: in this work, we assume that the amount of training/calibration data and the ensuing hand washing activity model is fixed at the time of initial deployment/calibration. In our future work, we plan to explore methods for active and/or online learning to update the hand washing activity models over time as more instances of hand washing occur at each location. This can be done by choosing new training data in an unsupervised manner based on highest prediction confidence, or in a semi-supervised manner (e.g., with active learning) by manually selecting the training data that improves overall model performance.

6.2 Indoor Occupant Activity Monitoring

In this work, we assume that the only types of human activities to be detected are those associated with hand washing. As such, any other activity that is recorded by the vibration sensors in the area will be classified either as "idle" or one of the hand washing activities. At large scale and in real-world scenarios, this assumption will likely result in false-positives of hand washing activities and decrease the overall performance of our hand washing monitoring system. Additionally, to address the challenge of overlapping/concurrent activities, we define "primary" and "secondary" activities in this work. As we observed in the evaluation with different persons, this treatment can successfully overcome the challenge of overlapping activities in some instances, but does not perform well when one activity dominates the other (i.e., if the "soap" signal is significantly lower than the "water" signal). Lastly, in real-world scenarios, there may be instances where multiple sinks are side-by-side, and hand washing may occur concurrently at each sink. These situations represent an additional form of overlapping/concurrent activities (i.e., multiple concurrent users), and make it difficult to separate and monitor each individual activity.

In our future work, we plan to address these system limitations and assumptions by leveraging prior works, which use structural vibration sensing to monitor indoor occupant activity and consider situations with multiple concurrent occupants [6, 7, 23, 64]. In these works, we characterize the differences in numerous human activities in indoor environments, and explore the effect of overlapping/concurrent activities. We plan to incorporate this into our hand washing monitoring system to reduce/eliminate the assumption of only hand washing activities occurring in the sensing area, and to help with characterizing overlapping/concurrent hand washing activities. For example, by incorporating behavioral models that describe the order and likelihood of different activities in different settings, we can eliminate the assumption of only hand washing activities without overwhelming our classifiers with all possible indoor activities, many of which may cause similar vibration responses. By expanding our hierarchical classification framework to include parallel models, we can detect multiple activities happening concurrently without having to train for every activity combination. In doing this, we will improve the real-world performance of our system and enable it to be applied in a variety of indoor environments.

Lastly, in our future work we plan to explore techniques for giving feedback to individuals who are washing their hands to educate them on proper practices and inform them when they are not adhering to the CDC and WHO guidelines. This feedback can come in many forms including audible reminders and/or visual cues. For example, incorporating a small light that turns green when the individual has satisfactorily washed their hands. The challenge with these feedback systems is how to implement them in a way that encourages good hand washing practices, and is not seen as a nuisance or detractor. This may involve pilot studies using different feedback systems in real-world settings to evaluate how effective each system is. In our future work, we plan to explore these systems and leverage our collaborations with nearby healthcare facilities to evaluate each.

7 CONCLUSIONS

In this paper, we present a novel approach for monitoring hand washing activities using structural vibration sensing. We overcome system challenges of similar and overlapping hand washing activities which occur on multiple different surfaces/structures through a cepstrum-based hierarchical learning approach. This approach extracts the differences in the signal periodicity from different hand washing activities to enable accurate detection and classification of a person walking to the hand washing station (footsteps), turning on the water, using a soap dispenser, and rinsing their hands. This approach enables non-intrusive and accurate monitoring of hand washing activities in a variety of settings. We evaluate our approach with real-world hand washing experiments across 4 structures, and with uncontrolled experiments involving 3 different experimental participants. Through these evaluations, we show that our approach achieves an average F1 score of 0.95 across all four structures,

which represents an error reduction of 8.8X and 10.2X from baseline approaches which use FFT and logFFT-based features.

ACKNOWLEDGMENTS

This research was partially supported by NSF Career (CMMI-2026699) and Highmark.

REFERENCES

[1] Paul S Addison. 2002. *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. CRC press.

[2] Sa'ed Alajlouni, Mohammad Albakri, and Pablo Tarazaga. 2018. Impact localization in dispersive waveguides based on energy-attenuation of waves with the traveled distance. *Mechanical Systems and Signal Processing* 105 (2018), 361–376.

[3] Wu Anxu. 2012. On quantitative identification of explosion earthquake based on cepstrum computation of HHT and statistical simulation of sub-cluster. In *Proceedings of the 31st Chinese Control Conference*. IEEE, 5311–5316.

[4] A. Ashraf and B. Taati. 2016. Automated Video Analysis of Handwashing Behavior as a Potential Marker of Cognitive Health in Older Adults. *IEEE Journal of Biomedical and Health Informatics* 20, 2 (March 2016), 682–690.

[5] Bruce P Bogert. 1963. The quefrency alanysis of time series for echoes; Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking. *Time series analysis* (1963), 209–243.

[6] Amelie Bonde, Shijia Pan, Mostafa Mirshekari, Carlos Ruiz, Hae Young Noh, and Pei Zhang. 2020. OAC: Overlapping Office Activity Classification through IoT-Sensed Structural Vibration. In *2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, 216–222.

[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] Jessie L Bonner, Delaine T Reiter, and Robert H Shumway. 2002. Application of a cepstral F statistic for improved depth estimation. *Bulletin of the seismological Society of America* 92, 5 (2002), 1675–1693.

[9] John M. Boyce. 2008. Hand hygiene compliance monitoring: current perspectives from the USA. *Journal of Hospital Infection* 70 (2008), 2 – 7.

[10] John C. Burgess. 1979. Applications of Digital Signal Processing, edited by Alan V. Oppenheim. *The Journal of the Acoustical Society of America* 65, 5 (1979), 1354–1354. https://doi.org/10.1121/1.382940

[11] Center for Disease Control and Prevention. 2002. Guideline for Hand Hygiene in Health-Care Settings: Recommendations of the Healthcare Infection Control Practices Advisory Committee and the HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. *Morbidity and Mortality Weekly Report* 51, RR-16 (2002), 1–48.

[12] Jianfeng Chen, Jianmin Zhang, Alvin Harvey Kam, and Louis Shue. 2005. An automatic acoustic bathroom monitoring system. In *2005 IEEE International Symposium on Circuits and Systems*. IEEE, 1750–1753.

[13] Nello Cristianini and John Shawe-Taylor. 2000. *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. Cambridge University Press, New York, NY, USA.

[14] Yiwen Dong, Joanna Jiaqi Zou, Jingxiao Liu, Jonathon Fagert, Mostafa Mirshekari, Linda Lowes, Megan Iammarino, Pei Zhang, and Hae Young Noh. 2020. MD-Vibe: physics-informed analysis of patient-induced structural vibration data for monitoring gait health in individuals with muscular dystrophy. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*. 525–531.

[15] ABMSU Doulah and SA Fattah. 2014. Neuromuscular disease classification based on mel frequency cepstrum of motor unit action potential. In *2014 International Conference on Electrical Engineering and Information & Communication Technology*. IEEE, 1–4.

[16] Slah Drira, Yves Reuland, Nils FH Olsen, Sai GS Pai, and Ian FC Smith. 2019. Occupant-detection strategy using footstep-induced floor vibrations. In *Proceedings of the 1st ACM International Workshop on Device-Free Human Sensing*. 31–34.

[17] Jonathon Fagert, Mostafa Mirshekari, Shijia Pan, Linda Lowes, Megan Iammarino, Pei Zhang, and Hae Young Noh. 2021. Structure- and Sampling-Adaptive Gait Balance Symmetry Estimation Using Footstep-Induced Structural Floor Vibrations. *Journal of Engineering Mechanics* 147, 2 (2021), 04020151.

[18] J. Fagert, M. Mirshekari, S. Pan, P. Zhang, and H.Y. Noh. 2017. Characterizing left-right gait balance using footstep-induced structural vibrations. In *SPIE 10168, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, Vol. 10168. 10168 – 10168 – 9.

[19] Jonathon Fagert, Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. 2019. Characterizing Structural Changes to Estimate Walking Gait Balance. In *Dynamics of Civil Structures, Volume 2*. Springer, 333–335.

[20] Jonathon Fagert, Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. 2019. Gait health monitoring through footstep-induced floor vibrations. In *2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 319–320.

[21] Jonathon Fagert, Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. 2020. Structural Property Guided Gait Parameter Estimation Using Footstep-Induced Floor Vibrations. In *Dynamics of Civil Structures, Volume 2*, Shamim Pakzad (Ed.). Springer International Publishing, Cham, 191–194.
[22] Jonathon Fagert, Mostafa Mishekari, Shijia Pan, Pei Zhang, and Hae Young Noh. 2017. Monitoring Hand-Washing Practices using Structural Vibrations. *Structural Health Monitoring 2017* (2017).
[23] Jonathon Fagert, Mostafa Mishekari, 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).
[24] Centers for Disease Control. 2020. How to Protect Yourself & Others. https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html.
[25] Jerome Friedman, Trevor Hastie, Robert Tibshirani, et al. 2001. *The elements of statistical learning*. Vol. 1. Springer series in statistics New York.
[26] Albert Haque, Michelle Guo, Alexandre Alahi, Serena Yeung, Zelun Luo, Alisha Rege, Jeffrey Jopling, Lance Downing, William Beninati, Amit Singh, et al. 2017. Towards Vision-Based Smart Hospitals: A System for Tracking and Monitoring Hand Hygiene Compliance. In *Machine Learning for Healthcare Conference*. 75–87.
[27] Timothy J Hazen and Victor W Zue. 1993. Automatic language identification using a segment-based approach. In *Third European Conference on Speech Communication and Technology*.
[28] Michael AH Hedlin, J Bernard Minster, and John A Orcutt. 1990. An automatic means to discriminate between earthquakes and quarry blasts. *Bulletin of the Seismological Society of America* 80, 6B (1990), 2143–2160.
[29] I/O Sensor Nederland bv 2006. *SM-24 Geophone Element*. I/O Sensor Nederland bv. P/N 1004117.
[30] Hyunjo Jeong and Young-Su Jang. 2000. Fracture source location in thin plates using the wavelet transform of dispersive waves. *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on* 47, 3 (2000), 612–619.
[31] Hyunjo Jeong and Young-Su Jang. 2000. Wavelet analysis of plate wave propagation in composite laminates. *Composite Structures* 49, 4 (2000), 443–450.
[32] Wen-Juh Kang, Jiue-Rou Shiu, Cheng-Kung Cheng, Jin-Shin Lai, Hen-Wai Tsao, and Te-Son Kuo. 1995. The application of cepstral coefficients and maximum likelihood method in EMG pattern recognition [movements classification]. *IEEE Transactions on Biomedical Engineering* 42, 8 (1995), 777–785.
[33] Ellis Kessler, Vijaya VN Sriram Malladi, and Pablo A Tarazaga. 2019. Vibration-based gait analysis via instrumented buildings. *International Journal of Distributed Sensor Networks* 15, 10 (2019), 1550147719881608.
[34] Clair Kilpatrick. 2009. Save Lives: Clean Your Hands. A global call for action at the point of care. *American Journal of Infection Control* 37, 4 (2009), 261–262.
[35] Andrew King and Robert Eckersley. 2019. *Statistics for biomedical engineers and scientists: How to visualize and analyze data*. Academic Press.
[36] Mike Lam, Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. 2016. Robust occupant detection through step-induced floor vibration by incorporating structural characteristics. In *Dynamics of Coupled Structures, Volume 4*. Springer, 357–367.
[37] Jonathan M Lilly and Sofia C Olhede. 2010. On the analytic wavelet transform. *IEEE transactions on information theory* 56, 8 (2010), 4135–4156.
[38] Lijuan Liu, Bo Shen, and Xing Wang. 2014. *Research on Kernel Function of Support Vector Machine*. Vol. 260. Springer Netherlands, Dordrecht, 827–834.
[39] Ramin Madarshahian, Juan M Caicedo, and Diego Arocha Zambrana. 2016. Benchmark problem for human activity identification using floor vibrations. *Expert Systems with Applications* 62 (2016), 263–272.
[40] YuriMalina Malina, Yuri Iseri, Mert Reiner, Sandra Hardman, Jori Rogers, Jill Vlasses, and Frances Vlasses. 2013. A Portable Trackable Hand Sanitation Device Increases Hand Hygiene. *American Journal of Infection Control* 41, 6 (2013), S37 – S38.
[41] Sergio Mendoza, Larry Gillick, Yoshiko Ito, Stephen Lowe, and Michael Newman. 1996. Automatic language identification using large vocabulary continuous speech recognition. In *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, Vol. 2. IEEE, 785–788.
[42] 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*. ACM, 1382–1391.
[43] Mostafa Mirshekari, Jonathon Fagert, Shijia Pan, Pei Zhang, and Hae Young Noh. 2020. Step-Level Occupant Detection across Different Structures through Footstep-Induced Floor Vibration Using Model Transfer. *Journal of Engineering Mechanics* 146, 3 (2020), 04019137.
[44] Mostafa Mirshekari, Jonathon Fagert, Shijia Pan, Pei Zhang, and Hae Young Noh. 2021. Obstruction-invariant occupant localization using footstep-induced structural vibrations. *Mechanical Systems and Signal Processing* 153 (2021), 107499. https://doi.org/10.1016/j.ymssp.2020.107499
[45] 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.

[46] Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. 2016. Characterizing wave propagation to improve indoor step-level person localization using floor vibration. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2016*, Vol. 9803. International Society for Optics and Photonics, 980305.

[47] Md Abu Sayeed Mondol and John A. Stankovic. 2015. Harmony: A Hand Wash Monitoring and Reminder System Using Smart Watches. In *Proceedings of the 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services on 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services* (Coimbra, Portugal). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, 11–20.

[48] F. Naim, R. Jaafar, N. W. Arshad, R. Hamid, and M. N. Razali. 2013. Unclean hand detection machine using vision sensor. In *2013 Saudi International Electronics, Communications and Photonics Conference*. 1–4.

[49] Alan V Oppenheim and Ronald W Schafer. 2004. From frequency to quefrency: A history of the cepstrum. *IEEE signal processing Magazine* 21, 5 (2004), 95–106.

[50] Sai GS Pai, Yves Reuland, Slah Drira, and Ian FC Smith. 2019. Is there a relationship between footstep-impact locations and measured signal characteristics?. In *Proceedings of the 1st ACM International Workshop on Device-Free Human Sensing*. 62–65.

[51] 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.

[52] Shijia Pan, Kent Lyons, Mostafa Mirshekari, Hae Young Noh, and Pei Zhang. 2016. Multiple Pedestrian Tracking through Ambient Structural Vibration Sensing. In *SenSys*. 366–367.

[53] Shijia Pan, Mostafa Mirshekari, Jonathon Fagert, Ceferino Gabriel Ramirez, Albert Jin Chung, Chih Chi Hu, John Paul Shen, Pei Zhang, and Hae Young Noh. 2018. Characterizing human activity induced impulse and slip-pulse excitations through structural vibration. *Journal of Sound and Vibration* 414 (2018), 61–80.

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

[55] Shijia Pan, Ningning Wang, Yuqiu Qian, Irem Velibeyoglu, Hae Young Noh, and Pei Zhang. 2015. Indoor person identification through footstep induced structural vibration. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*. 81–86.

[56] Shijia Pan, Tong Yu, Mostafa Mirshekari, Jonathon Fagert, Amelie Bonde, Ole J Mengshoel, Hae Young Noh, and Pei Zhang. 2017. FootprintID: Indoor Pedestrian Identification through Ambient Structural Vibration Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 89.

[57] Didier Pittet, Benedetta Allegranzi, and John Boyce. 2009. The World Health Organization Guidelines on Hand Hygiene in Health Care and Their Consensus Recommendations. *Infection Control & Hospital Epidemiology* 30, 7 (2009), 611–622.

[58] PhilipM. Polgreen, ChristopherS. Hlady, MonicaA. Severson, AlbertoM. Segre, and Ted Herman. 2010. Method for Automated Monitoring of Hand Hygiene Adherence without RadioFrequency Identification. *Infection Control and Hospital Epidemiology* 31, 12 (2010), 1294–1297.

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

[60] Jeffrey D Poston, R Michael Buehrer, and Pablo A Tarazaga. 2017. Indoor footstep localization from structural dynamics instrumentation. *Mechanical Systems and Signal Processing* 88 (2017), 224–239.

[61] M Pyrek. 2012. Hand hygiene monitoring goes high-tech. *Infection Control Today* (2012).

[62] Robert B Randall. 2017. A history of cepstrum analysis and its application to mechanical problems. *Mechanical Systems and Signal Processing* 97 (2017), 3–19.

[63] Nobuo Sato and Yasunari Obuchi. 2007. Emotion recognition using mel-frequency cepstral coefficients. *Information and Media Technologies* 2, 3 (2007), 835–848.

[64] 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.

[65] Asim Smailagic, Pedro Costa, Alex Gaudio, Kartik Khandelwal, Mostafa Mirshekari, Jonathon Fagert, Devesh Walawalkar, Susu Xu, Adrian Galdran, Pei Zhang, et al. 2020. O-MedAL: Online active deep learning for medical image analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10, 4 (2020), e1353.

[66] Asim Smailagic, Pedro Costa, Hae Young Noh, Devesh Walawalkar, Kartik Khandelwal, Adrian Galdran, Mostafa Mirshekari, Jonathon Fagert, Susu Xu, Pei Zhang, et al. 2018. Medal: Accurate and robust deep active learning for medical image analysis. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 481–488.

[67] CT Sun and JM Bai. 1995. Vibration of multi-degree-of-freedom systems with non-proportional viscous damping. *International journal of mechanical sciences* 37, 4 (1995), 441–455.

[68] Torfinn Taxt. 1997. Comparison of cepstrum-based methods for radial blind deconvolution of ultrasound images. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 44, 3 (1997), 666–674.

[69] W.T. Thomson and M.D. Dahleh. 1998. *Theory of Vibration with Applications*. Prentice Hall.
[70] Keiichi Tokuda, Takao Kobayashi, and Satoshi Imai. 1995. Adaptive cepstral analysis of speech. *IEEE Transactions on Speech and Audio Processing* 3, 6 (1995), 481–489.
[71] Keith A Wear, Robert F Wagner, Michael F Insana, and Timothy J Hall. 1993. Application of autoregressive spectral analysis to cepstral estimation of mean scatterer spacing. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* 40, 1 (1993), 50–58.
[72] Fu-sheng Wei and Ming Li. 2003. Cepstrum analysis of seismic source characteristics. *Acta Seismologica Sinica* 16, 1 (2003), 50–58.
[73] Eddie Wong and Sridha Sridharan. 2001. Comparison of linear prediction cepstrum coefficients and mel-frequency cepstrum coefficients for language identification. In *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing. ISIMP 2001 (IEEE Cat. No. 01EX489)*. IEEE, 95–98.
[74] World Health Organization. 2011. *Report on the burden of endemic health care-associated infection worldwide*. Technical Report.
[75] Erdem Yavuz and Can Eyupoglu. 2019. A cepstrum analysis-based classification method for hand movement surface EMG signals. *Medical & biological engineering & computing* 57, 10 (2019), 2179–2201.