Recursive Sparse Representation for Identifying Multiple Concurrent Occupants Using Floor Vibration Sensing

JONATHON FAGERT, Department of Engineering, Baldwin Wallace University, USA
MOSTAFA MIRSHEKARI, Searchable.ai, USA
PEI ZHANG, Department of Electrical Engineering and Computer Science, University of Michigan, USA
HAE YOUNG NOH, Department of Civil and Environmental Engineering, Stanford University, USA

In this paper, we present a multiple concurrent occupant identification approach through footstep-induced floor vibration sensing. Identification of human occupants is useful in a variety of indoor smart structure scenarios, with applications in building security, space allocation, and healthcare. Existing approaches leverage sensing modalities such as vision, acoustic, RF, and wearables, but are limited due to deployment constraints such as line-of-sight requirements, sensitivity to noise, dense sensor deployment, and requiring each walker to wear/carry a device. To overcome these restrictions, we use footstep-induced structural vibration sensing. Footstep-induced signals contain information about the occupants' unique gait characteristics, and propagate through the structural medium, which enables sparse and passive identification of indoor occupants. The primary research challenge is that multiple-person footstep-induced vibration responses are a mixture of structurally-codependent overlapping individual responses with unknown timing, spectral content, and mixing ratios. As such, it is difficult to determine which part of the signal corresponds to each occupant. We overcome this challenge through a recursive sparse representation approach based on cosine distance that identifies each occupant in a footstep event in the order that their signals are generated, reconstructs their portion of the signal, and removes it from the mixed response. By leveraging sparse representation, our approach can simultaneously identify and separate mixed/overlapping responses, and the use of the cosine distance error function reduces the influence of structural codependency on the multiple walkers' signals. In this way, we isolate and identify each of the multiple occupants' footstep responses. We evaluate our approach by conducting real-world walking experiments with three concurrent walkers and achieve an average F1 score for identifying all persons of 0.89 (1.3× baseline improvement), and with a 10-person "hybrid" dataset (simulated combination of single-walker real-world data), we identify 2, 3, and 4 concurrent walkers with a trace-level accuracy of 100%, 93%, and 73%, respectively, and observe as much as a 2.9× error reduction over a naive baseline approach.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**; • **Computing methodologies → Machine learning approaches**.

Additional Key Words and Phrases: Structural Vibrations, Occupant Identification, Signal Separation, Sparse Representation, Smart Structures

ACM Reference Format:
Jonathon Fagert, Mostafa Mirshekari, Pei Zhang, and Hae Young Noh. 2022. Recursive Sparse Representation for Identifying Multiple Concurrent Occupants Using Floor Vibration Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 1, Article 10 (March 2022), 33 pages. https://doi.org/10.1145/3517229

Authors' addresses: Jonathon Fagert, jfagert@bw.edu, Department of Engineering, Baldwin Wallace University, 275 Eastland Road, Berea, Ohio, USA, 44017; Mostafa Mirshekari, mostafa@searchable.ai, Searchable.ai, San Francisco, California, USA, 94105; Pei Zhang, peizhang@umich.edu, Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, USA, 48109; Hae Young Noh, noh@stanford.edu, Department of Civil and Environmental Engineering, Stanford University, Stanford, California, USA, 94305.

2474-9567/2022/3-ART10 $15.00
https://doi.org/10.1145/3517229

1 INTRODUCTION

Identification (ID) of indoor occupants is an important component of smart infrastructure. Timely and accurate identification of indoor occupants can assist with building security, space allocation/utilization, personalized services, as well as healthcare monitoring. Various approaches exist in the literature for identifying indoor occupants including WiFi/Radio Frequency (RF) [2, 8, 34, 35, 84, 85], vision [7, 12, 39, 56, 60, 77, 81], mobile [28, 45, 48, 57, 79], and acoustic [29, 88]. These approaches are state-of-the-art techniques for person identification, but are limited in their deployment due to restrictions of line-of-site (vision), dense sensor deployment (WiFi/RF), ambient noise (acoustic), and requiring persons to carry a device at all times (mobile).

To overcome these limitations, prior works have utilized structural floor vibration sensing due to its passive and sparse sensing capability and "device-free" nature [65, 67]. The primary physical insight for these systems is that human gait patterns are a type of biometric and are unique [11]. The human gait is a complex activity that involves many skeletal and muscle groups within the human body. As such, physical characteristics of individuals (e.g., height, weight, posture, etc.) all influence how someone walks. The use of gait as a biometric has a rich history and many people can relate to the concept of identifying persons they know based on their gait (e.g., that looks like a friend walking in the crowd ahead of me). In fact, gait as a biometric is reliable enough to be used in forensic analysis [11]. For example, a tall person may have longer strides, a heavier person may have a higher footstep ground reaction force, and some people may walk more slowly/quickly (which affects both center of mass position and foot contact angle). When these individuals walk in a building strucutre, their footsteps impart a dynamic force on the structure itself, which induces a vibration response that can be measured using vibration sensors. Therefore, these footstep-induced vibration responses will also be unique and can be used for identification. Prior approaches using footstep-induced floor vibrations focus on person identification scenarios where only one person is walking in the sensing area [65, 67]. In real-world scenarios, however, people tend to walk in groups of two or more concurrent walkers. As such, these prior approaches are limited in many real-world scenarios.

In our work, we present a person identification approach for "multiple walker" scenarios (i.e., when two or more persons are walking concurrently) using footstep-induced structural vibration sensing. The challenges with this approach are twofold: 1) the footstep responses from multiple concurrent walkers are an *overlapping mixture* of each individual's response, with the timing of each step as well as the contribution of that step to the overall response being unknown. As a result, it is difficult to separate the components of the signal that correspond to each individual, making it difficult to accurately identify each walker. Figure 1 shows a conceptual example of this challenge; and 2) vibration responses from different persons have a *codependency* on the properties of the underlying structure, and these structural components dominate the overall response. As such, it is difficult to extract the unique signal components necessary for identifying each occupant.

This type of problem can be viewed as a "signal/source separation" problem. These types of problems are highly studied area in the signal processing domain. These works are characterized by their goal of isolating/separating individual responses from mixed signals, typically without any prior knowledge of the input/individual responses. The most prominent works in this field fall under the category of "Blind Source Separation" (BSS). Common BSS approaches include ones that use Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Non-Negative Matrix Factorization (NMF) [5, 33, 40, 41, 54, 72, 73, 90].

In the case of footstep-induced vibration responses, however, these approaches have limitations that result in decreased performance. PCA and ICA-based approaches have a forced/assumed source independence requirement which is not satisfied with footstep-induced structural vibrations because the response for each occupant depends on the vibration characteristics of the medium (i.e., the structure). Further, NMF, which does not rely on the independence of the source signals, imposes a non-negativity requirement, which can result in loss of source signal information and reduce the ability to uniquely identify occupants with separated signals [40].

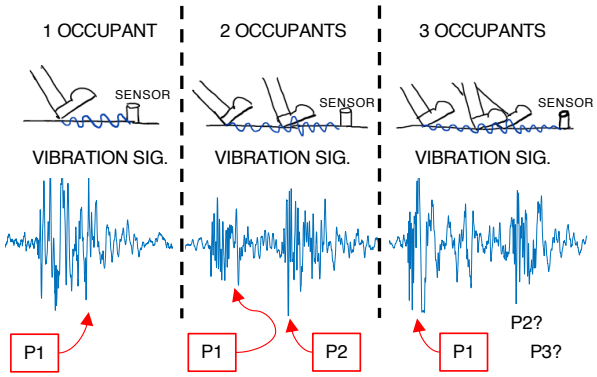


Fig. 1. Footstep-induced signals from multiple walkers may overlap. With overlapping signals, it is difficult to determine the number of walkers and their individual signal components.

To address these challenges, our approach leverages sparse representation to sequentially and recursively identify individual responses, reconstruct them, and remove them from the combined signal. This approach is built on the insight that the dynamic response of the structure is the superposition of the response from each person's footstep excitation. Further, individuals tend to walk similarly across different steps and prior steps when a person was walking independently can be used as basis functions to represent new steps from that same person. In this way, our approach leverages a sparse representation to sequentially and recursively identify which basis function (prior step) best represents each step within a detected footstep event (where a "footstep event" may contain one or more individual steps). To reduce the effect of the *structure codependency*, we find the sparse coefficients using a cosine distance-based error function. We observe that the frequency spectrum of footstep-induced vibrations is dominated by the "structural response" components (i.e., those due to the dynamic properties of the structure itself, such as resonant frequencies with large magnitude peaks). As such, a traditional error function such as the Euclidean distance would place more emphasis on the dominant peaks and have higher identification errors. The cosine distance, on the other hand, measures the cosine of the angle between the two vectors (e.g., the two frequency spectra in our problem) [32]. It is a normalized frequency spectrum comparison, which focuses on the similarity between the overall shapes of the two spectra regardless of their magnitudes and thus reduces the influence of the structure-dependent peaks. By using a cosine distance error function, our approach minimizes the difference between the overall shape of the frequency spectrum, enabling a better match between similar responses (i.e., the same person). The basis functions identified by the sparse representation are then used to reconstruct that individual's independent response, and it is subtracted from the combined response. The remaining signal then contains the next person's footstep response, which can be identified, reconstructed, and removed in the same way. This process is done sequentially and recursively for each footstep event until all individual steps are identified, reconstructed, and removed from the combined response. As a result, the multiple walker response is separated into individual components and each person is identified. In summary, our primary research contributions are:

(1) We present an approach for identifying multiple concurrent walkers using footstep-induced floor vibrations.
(2) We separate mixed vibration responses through a recursive sparse-representation-based occupant identification and signal reconstruction approach.

(3) We overcome the effects of *structure codependency* using a cosine distance-based error function which minimizes the difference between the normalized frequency spectra, thereby reducing the influence of dominant peaks in the frequency spectrum that correspond to the dynamic properties of the structure itself.
(4) We evaluate our approach with real-world walking experiments involving 3 concurrent walkers and hybrid real-world/simulation experiments with 10 total walkers.

The rest of the paper is organized as follows: First, we provide an overview of relevant related works (Section 2). Next, we describe the physical insights that enable our multiple person identification approach (Section 3). Following this, we describe our recursive sparse representation-based approach in detail (Section 4). Next, we evaluate the performance of our work in Section 5, and describe our future work (Section 6). Finally, we provide concluding remarks (Section 7).

2 RELATED WORK

In our work, we identify indoor occupants in "multiple walker" scenarios using footstep-induced floor vibrations. The relevant prior work in this area is largely from two primary categories: 1) approaches for identification of indoor occupants, and 2) signal/source separation approaches. The following sections presents each of these categories in detail and discusses how our approach fills the research gaps that exist in prior works.

2.1 Indoor Occupant Identification

Indoor occupant identification is a common application for various sensing modalities in the literature. In this section, we explore some of the most prominent works in this area which include: vision-based [7, 12, 39, 56, 60, 77, 81], radio-frequency(RF)/WiFi-based [2, 8, 34, 35, 84, 85], mobile/wearable-based [28, 45, 48, 57, 79], acoustic-based [29, 88], and vibration-based [65, 67].

Vision-based methods determine person identification from visual biometrics such as facial features [26, 39, 81] and body structure/characteristics (including gait mechanics) [7, 12, 56, 60, 77]. Accuracy with vision-based methods is typically quite high and for this reason they are commonly used in society (e.g., many mobile devices now use some form of face recognition for unlocking devices). The limitation of these approaches, however, is that they require a clear line of site to operate (i.e., no walls or objects obstructing the view of the camera). As a result, vision-based methods may have "blind spots" or areas where they do not function, and require additional cameras to cover areas, which is expensive at building-scale. Further, many vision-based approaches raise privacy concerns and people often do not like cameras in their homes or in areas that need additional privacy (e.g., bathrooms, bedrooms, etc.).

Radio frequency methods have been shown to successfully identify indoor occupants in a variety of indoor settings and typically rely on tracking body movement and gait patterns to identify individuals [2, 8, 34, 35, 43, 84, 85]. These approaches, however, typically require the persons to walk or move through the path between the transmitter and receiver, so they have limited identification range unless a large number of sensors is deployed.

Mobile/wearable-based systems, on the other hand, provide ubiquitous person identification due to their personalized information and ability to travel everywhere with occupants [28, 45, 48, 57, 79]. These types of approaches typically fall in two categories: 1) IMU-based systems that track gait and body movement to learn unique characteristics of each person [28, 45, 48], and 2) tag-based systems that remember unique IDs from devices given to each person and then use those devices/IDs to identify/track the individuals [57, 79]. The limitation with these approaches, however, is that they require individuals to always wear/carry the device, and maintain its charge throughout the entire day.

Acoustic-based techniques sense changes in vibration patterns through a medium to determine occupant identities. Acoustic-based methods use the air as a medium and determine occupant IDs through unique body

movements and/or gait [29, 88]. They are limited, however, by the presence of ambient noise in the environment - if there is a loud machine or other sources of noise, these approaches do not perform well.

Prior works have utilized structural floor vibration sensing to overcome many of the limitations of the sensing modalities discussed above due to its passive and sparse sensing capability and "device-free" nature [67]. Structural floor vibration sensing has been shown be successful with applications involving identification [65, 67], detection/tracking [18, 47, 49, 58, 61, 68], localization [3, 50–52, 69], and monitoring human gait health and activity levels [6, 17, 19–24, 42, 44, 59, 62, 64]. These prior approaches focus on scenarios where only one person is walking in the sensing area. In real-world scenarios, however, people tend to walk in groups of two or more concurrent walkers. As such, prior approaches using footstep-induced floor vibrations are limited in many real-world scenarios. Some preliminary work focused on multiple walker scenarios [25, 75] shows that this sensing modality has potential for expansion from the previously studied multiple-walker scenarios, but was limited to extracting step onset and/or small-scale signal characteristics.

In this work, we address the limitations of prior works using footstep-induced structural vibration sensing. Our approach leverages physical insights regarding the differences in human walking and floor vibration characteristics to identify multiple concurrent walkers through a recursive sparse representation-based approach.

2.2 Signal Separation

As discussed previously, our treatment of the multiple walker-induced structural vibrations can be viewed as a "Blind Source Separation" (BSS) problem. The most prominent BSS approaches are those which leverage Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Non-Negative Matrix Factorization (NMF) [5, 33, 40, 41, 54, 72, 73, 90].

Of these, the most common BSS approach is ICA, which relies on independence of each input "source" [10]. ICA has been successfully applied to speech recognition (i.e., the *"cocktail party problem"*) [9, 37], medical images/EEG signals [76, 83], financial data [31, 53], image processing/denoising [36, 46], and communication systems [14, 71]. For multiple walker footstep-induced vibrations, however, potential loss of signal information (NMF) and/or source codependency on the underlying structure cause many of these prior BSS approaches to fail to accurately separate the vibration responses from the multiple concurrent walkers, resulting in erroneous identification.

To overcome the limitations of these prior BSS approaches, our work introduces a recursive sparse representation-based signal separation approach which leverages a cosine distance error metric. In this way, our approach reduces the influence of the structure codependency and separates mixed/overlapping responses from multiple concurrent walkers. Section 4.2 explores our approach in more detail.

3 PHYSICAL INSIGHTS FOR MULTIPLE PERSON IDENTIFICATION

To identify occupants in indoor settings, we separate multiple-person footstep-induced structural vibration signals into the individual components corresponding to each walker. In this section, we discuss the physical insights regarding the dynamics of footstep-induced floor vibrations that enable our recursive sparse representation-based approach.

For building structures, the floor vibration response can be modeled as the convolution between the structure's frequency response function and the excitation forcing function. This convolution relationship is given by the following expression [78, 80]:

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

where $x(t)$ is the time history of the building vibration response, $h(t)$ is the structure's impulse response function, $f(t)$ is the vibration forcing function (i.e., the footstep force in our case), and $*$ is a symbol representing the

convolution integral. In the case of a multiple walker footstep-induced forcing function, $f(t)$ contains the mixture of the individual forcing functions from each individual person.

In a linear-elastic structure, dynamic responses (i.e., floor vibrations) can be modeled as the addition of each individual excitation source. This behavioral effect is typically referred to as the Principal of Superposition [1, 13, 86]. Superposition, therefore, enables us to rewrite Equation 1 as the summation of the responses due to each walker as follows:

$$x(t) = h(t) * (f_1(t) + f_2(t) + ...f_N(t)) \tag{2}$$

where $f_n(t)$ represents the individual footstep forcing function of person n and N is the total number of concurrent walkers in the sensing area. Further, the distributive properties of the convolution integral allow us to rewrite Equation 2 as the summation of each individual response convolution in the following way:

$$x(t) = h(t) * f_1(t) + h(t) * f_2(t) + ...h(t) * f_N(t) \tag{3}$$

From this expression we can make two key observations: 1) each of the individual footstep excitations (i.e., those from each walker) are not independent - they have a mutual dependence on the structure's frequency response function ($h(t)$). Based on this observation, we conclude that traditional approaches like ICA are not suitable for separating multiple walker's footstep-induced vibrations; and 2) building off the insight regarding gait as a biometric described above, if we assume that the same person walks similarly across different steps, we can use prior information (i.e., prior footstep responses) from each person to estimate their components of the total signal ($h(t) * f_n(t)$) and isolate/separate them. This enables us to uniquely identify individual walkers from a multiple walker mixed signal. In real-world settings, the response superposition shown in Equation 3 typically involves a few number of concurrent walkers. Further, if we assume only a few number of people's footsteps will completely overlap at a time, this superposition can be modeled as a sparse representation of the prior responses from each person (i.e., a sparse representation of a dictionary of prior step basis functions). We leverage these observations and insights to enable our recursive sparse representation-based multiple person identification approach. Further details regarding this approach are presented in Section 4.2.

4 MULTIPLE PEOPLE IDENTIFICATION APPROACH

Our approach for the identification of multiple concurrent walkers uses sparse representation to sequentially and recursively identify occupants and reconstruct their portions of the mixed footstep responses. This approach consists of four main modules: 1) vibration sensing and adaptive footstep event detection (Section 4.1), 2) sparse-representation-based recursive occupant identification (Section 4.2), 3) signal reconstruction for identified footstep removal (Section 4.3), and 4) trace-level dictionary updating (Section 4.4). An overview of our approach is shown in Figure 2.

4.1 Vibration Sensing and Adaptive Footstep Event Detection

The first module of approach collects structural vibration signals and detects when footsteps occur. To measure the floor vibration signals, we use geophone sensors, which are low-cost vibration sensors that measure the velocity of the floor vibrations and can be easily retrofitted into existing buildings. To increase the signal resolution, we amplify the signal with a variable gain operational amplifier (100-1000X gain), which can be calibrated based on observed responses in the sensing area.

We monitor the collected vibration signals for the presence of footstep-induced vibrations using an adaptation of the Chi-squared anomaly detection approach [51]. This approach detects footsteps as impulsive events that have a higher variation from the ambient noise levels and extracts them from the total signal into footstep event windows. Specifically, we compare the null hypothesis ($H_0 : \sigma_w^2 \leq \sigma_n^2$) and the alternative hypothesis

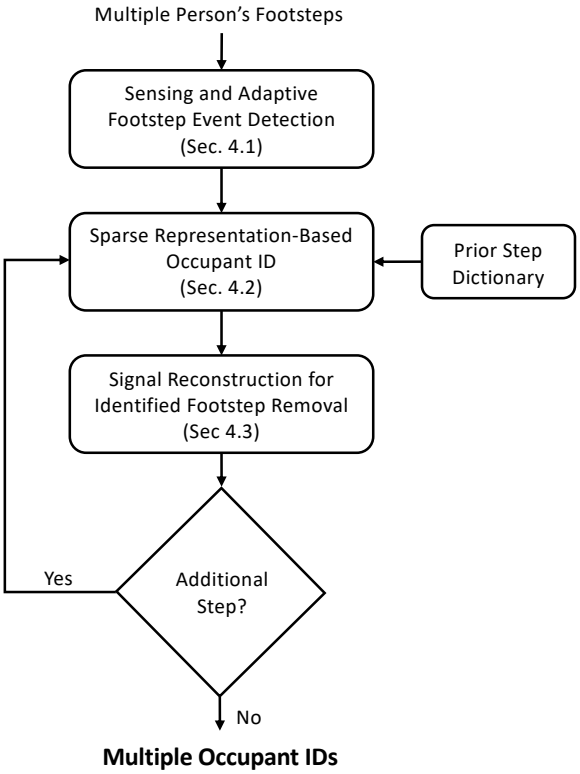
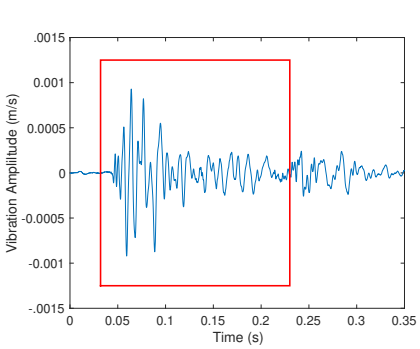


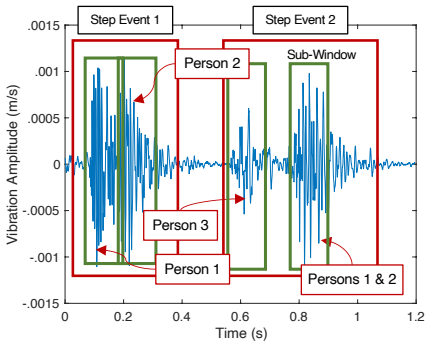
Fig. 2. Overview of our approach for separating and identifying multiple concurrent walkers. Once several step events have been identified at the step level, the predictions are used at the trace level to update/restrain the prior step dictionary and update the predictions to the most common persons in the area.

($H_1 : \sigma_w^2 > \sigma_n^2$). In these hypotheses σ_w^2 represents the sample variance of the current vibration signal window and σ_n^2 represents the sample variance of the ambient noise levels [4, 30, 49, 51]. When the null hypothesis is rejected, our system marks the signal window as an "impulsive event" and further classifies it as a "footstep" or "non-footstep" event using the process outlined in [49].

In real-world situations, there are typically multiple concurrent walkers. As such, impulsive events that are classified as footsteps may contain one or multiple overlapping steps from different walkers. Figure 3 shows examples of this challenge. In this figure, when there is only one walker, detected impulses can be classified as footsteps and used for determining the identity of the walker. However, in the case of multiple walkers, detected impulsive events may contain one or more footsteps from each of the walkers, and it is difficult to determine how many footstep responses are present when signals are mixed and overlapping. As such, our adaptive footstep event detection approach has a variable window length and determines both the onset and the termination of the footstep event. We determine the onset using the anomaly detection process described above, then continue monitoring the signal with a sliding window until the null hypothesis is no longer rejected (which is then used to mark the footstep event termination). In the case of many concurrent walkers (e.g., more than 4), it is possible that the individual steps from each person will continuously overlap. In this case, the null hypothesis would be



(a) Single walker adaptive footstep event detection.



(b) Multiple walker adaptive footstep event detection.

Fig. 3. An example of the output from our adaptive footstep event detection approach with (a) one walker and (b) multiple walkers. In (b), the first event contains two partially overlapping footsteps and the second event contains three footsteps (2 of which are overlapping). Red boxes indicate detected footstep event windows and green boxes indicate sub-windows.

rejected for many consecutive individual steps, and the likelihood of detecting multiple steps from the same person increases. To reduce this effect, we set an upper-bound on the footstep event length equal to the average step period of walkers in the area (which can be determined from prior information). In this way, our detected footstep events have a variable length and we can capture each of the footsteps that occur at approximately the same time as one, combined, footstep event.

The challenge with the adaptive footstep event detection employed in our method, however, is that the number of footsteps inside the window is unknown. To address this challenge, our approach conducts a trace-level, step-level, and sub-window-level classification to determine the number of occupants in the sensing area, number of steps in a "footstep event", and number of individual steps in a sub-window of the overall footstep event (i.e., if there are "perfectly overlapping" steps or not), respectively. We define a footstep trace as a series of consecutive footsteps in the sensing area, a footstep event as described above, and the sub-window represents a portion of the step event corresponding to each individual detected step within the overall footstep event window. The size of the sub-window can be determined empirically based on the average duration of footstep responses in the sensing area. For this work, we define the sub-window size as 4000 samples (approximately 0.15s) based on the observed average duration of a singular footstep response.

Figure 3 shows example outputs from our adaptive footstep event detection module. In Figure 3a, a single footstep is detected as a footstep event, while in Figure 3b, two footstep events are detected; in the first footstep event (the one to the left), there are two clear footstep responses, and our approach is able to detect those as one event and output the number of footsteps within the footstep event window (2). Then, for the second detected footstep event, there appear to be two footstep responses, but there are, in fact, three (Persons 1 and 2 have completely overlapping steps). In these cases, our system detects the footstep event, and outputs the correct number of footstep responses (3) within that window using the classification approach described above.

At the trace-level, we extract the entropy of the footstep trace for each sensor (defined by: $H(X) = -\sum_{i=1}^{n} \mathrm{P}(x_i) \log \mathrm{P}(x_i)$, where $H(X)$ is the entropy and $P(x_i)$ is the probability of the value of sample i [74]), and use this entropy as a feature for determining the number of walkers in the sensing area using a multi-class support vector machine (SVM) classifier [15, 27]. We choose to use an SVM classifier based on the insight that our dataset is small, and the feature distribution is unknown, and SVM classifiers are well-suited to distinguishing classes in those types of conditions. The maximum number of walkers (i.e., number of classes) for

this classifier can be determined empirically during initial system calibration based on anticipated building/sensing area occupancy. Alternatively, the trace-level number of walkers can be determined according to the procedures outlined in prior occupancy tracking works which utilize footstep-induced vibration sensing [63, 68].

Next, at a step-level, we extract the signal energy from each sensor (defined as sum of squares of signal sample values [70]), length of the footstep event window (number of samples between the detected signal onset and termination as described above), and entropy of the frequency domain [74] as features for each footstep event window and use a multi-class SVM classifier to determine the number of persons (footsteps) in the footstep event window. These features were chosen empirically based on the observation and intuition that the number of footstep impulses is proportional to the randomness in the signal (i.e., the entropy), that the step event duration is typically longer with an increasing number of consecutive footsteps, and also that the signal energy is proportional to the number of footstep impulses. In this classifier, our system outputs the number of footsteps in the footstep event window. If the predicted number of steps exceeds the trace-level number of walkers, the prediction is updated to the number of walkers (i.e., there cannot be more steps than total walkers). This is based on the assumption that no more than one step from each person is in any one footstep event (based on the upper-limit for the footstep event duration discussed above).

Finally, the same set of features used at the step-level is extracted from the sub-window signal to classify the number of individual footsteps within each sub-window with a separate multi-class SVM classifier. In this classifier, when the number of footsteps predicted exceeds 1, our system considers the footstep responses to be "completely overlapping", meaning that two (or more) persons step at approximately the exact same time. During training, the threshold for "completely overlapping" can be set empirically to allow for some partial overlapping to be considered "completely overlapping" for analysis purposes. This takes into consideration that, when signals are separated by a very small amount of time (i.e., less than a few thousand samples/less than 0.1s), most of their responses are completely mixed and difficult to distinguish. In this work, we assume a threshold of 3000 samples (0.11s) for this purpose.

The output of this model (trace-level, step-level, and window-level number of steps, and isolated footstep events) are then sent to the next module to be used in our recursive sparse representation-based approach.

4.2 Sparse Representation-Based Occupant Identification

Once footstep events have been detected using the procedure outlined in Section 4.1, our system next identifies the footsteps in the detected footstep event using a recursive sparse representation-based approach. As previously discussed, the insight behind this approach is that each person's footstep response is unique and similar across multiple footsteps because humans each have a unique walking style, and that each sub-window contains only a few number of persons' overlapping individual steps. As such, the response in each footstep event sub-window can be approximated using a sparse representation of a dictionary of basis functions, which are created using prior footstep responses from each person when he/she was walking solo. For this work, we use the magnitude spectrum of the Fourier transform of the vibration signals when conducting this sparse approximation. The magnitude spectrum is used based on prior work, which showed that the frequency domain signals best highlight the differences in responses between different people (and are, therefore, well-suited for use in identification) [67].

Sparse representation is a common technique in the fields of signal processing and data analysis to represent a measured/observed signal as the product of a "dictionary" of basis functions and a sparse coefficient vector [89]. The basic concept for sparse representation is founded on the principal of superposition outlined above, and, the insight that an individual footstep event sub-window can be represented using a few number of basis functions. Therefore, a sparse representation-based approach is well-suited for our multiple occupant problem. In these problems, the term "sparse" refers to the constraint that there are more zero-valued elements in the coefficient vector than non-zero ones. The general expression for sparse representation used in this work is the following [89]:

$$\mathbf{Y} = \mathbf{DX} \tag{4}$$

where $\mathbf{Y}_{n\times 1}$ is the magnitude of the Fourier Transform of the multiple person vibration response signal window (i.e., the "combined" response), $\mathbf{D}_{n\times m}$ is the "dictionary" of basis functions (which consists of prior steps from potential walkers in the area), $\mathbf{X}_{m\times 1}$ is the coefficient vector, n is the length of the current signal window, and m represents the number of basis functions in the dictionary. In our work, m is dependent on the prior footstep information available, and is equal to the summation of the number of basis functions (prior steps) available for each potential walker in the area.

A number of approaches exist in the literature for solving sparse representation problems [89]. In this work, we are using a sparse representation to concurrently identify an individual walker from a combined (i.e., multiple walker) response, and isolate/reconstruct that person's individual footstep response. As such, for our sparse representation, we are imposing a constraint that the ℓ_0-Norm of X is less than or equal to the number of estimated walkers in the current footstep event sub-window, n_w. That is to say, there is one basis function in the prior step dictionary that best represents each individuals' footstep in the current signal window. The "less than or equal to" in this formulation provides the algorithm flexibility in the event that the number of steps in the sub-window was misclassified. For example, if two steps are predicted in the current sub-window, but the algorithm identifies that one person's basis function best represents that sub-window, the algorithm will identify the step in the sub-window as the singular step only (instead of two, which may have been an incorrect estimation). Therefore, a traditional approach to solving the sparse representation with constraint on the ℓ_0-Norm of X is given by the following expression [89]:

$$\begin{aligned} \hat{X} &= \arg\min_{\mathrm{X}} ||Y - DX||_2^2, \\ s.t. &||X||_0 \leq n_w \end{aligned} \tag{5}$$

where $\hat{X}$ is the optimal sparse coefficient vector to represent Y using the prior step dictionary D, and n_w is the number of predicted steps in the current footstep event sub-window (as determined according to the process outlined in Section 4.1 above). This approach leverages a Euclidean distance-based error function to approximate the residual between the measured signal (Y) and the sparse approximation signal (DX). However, from our observations, the responses for each person are dominated by similar structure-based frequencies (i.e., the *structure codependency*). As a result, this Euclidean distance error function will concentrate on reducing the error of the highest magnitude frequencies (i.e., the resonant responses occurring at/around the natural frequency of the structure), leading to confusion between individuals.

To overcome this challenge and reduce the erroneous predictions from a Euclidean distance-based error function, our work leverages a cosine distance-based error function. The cosine distance is characterized as a metric for comparing the overall shape (direction) of two vectors in space. It differs from a Euclidean distance metric in that it is a normalized dot product of the two vectors being compared [32]. As such, by normalizing the comparison, the cosine distance is less influenced by the dominant peaks in the frequency spectrum and provides a better comparison of the overall shape of the measured signal and sparse approximation signal. Therefore, we replace the Euclidean distance error function in Equation 5 with a cosine distance error function to obtain the following expression for solving the sparse approximation:

$$\begin{aligned} \hat{X} &= \arg\min_{\mathrm{X}} \left[1 - \frac{\Sigma_{i=1}^{n} y_i (\mathbf{DX})_i}{\sqrt{\Sigma_{i=1}^{n} y_i^2} \sqrt{\Sigma_{i=1}^{n} (\mathbf{DX})_i^2}} \right], \\ s.t. &\|\mathbf{X}\|_0 = 1, \|\mathbf{X}\|_2 = 1 \end{aligned} \tag{6}$$

With this formulation, we impose the constraint of one nonzero coefficient in X (ℓ_0-Norm of X equals one) and utilize a grid search which guarantees the optimal solution, instead of $||X||_0 \leq n_w$ in Equation 5. To facilitate this change, we modify the dictionary to include additional basis functions when n_w is predicted to be greater than one, which are the combination of each persons' individual basis functions. Additional details regarding these cases are provided below. This change helps reduce the likelihood of misidentification and facilitates the identification of the footstep. Specifically, this constraint ensures that the algorithm uniquely identifies one walker (or one combination of walkers when $n_w > 1$) for each footstep event sub-window. Further, we impose the constraint that the ℓ_2-Norm of X equals one to ensure that a unique solution is obtained (since the cosine distance is a normalized comparison).

Using this approach, we construct a dictionary for sparse representation using prior footstep information from each potential walker in the sensing area when he/she was walking solo (i.e., not with multiple walkers). To construct the dictionary, we use the frequency spectrum of the individual responses, which are normalized by the energy of the window to reduce the effects of attenuation and dispersion. Our approach uses this prior footstep dictionary to identify the persons walking with the sparse representation approach described above. We first take a sub-window of the overall footstep event window, with a length empirically set to 0.15s based on our observed average duration of a single step response. With this sub-window, our approach identifies the first step in the footstep event.

When the number of footsteps in the sub-window is classified as being greater than one (like the mixed step in Figure 3b), we append the dictionary with linear combinations of the dictionary elements for each of the potential walkers (e.g., Person 1 + Person 2, etc.). In this work, we use all possible combinations of the potential walkers. However, at large scale (i.e., if the number of possible walkers is large (in the hundreds or more)), this number of combinations becomes computationally expensive. In these cases, the sparsity constraint in Equation 6 can be relaxed from $\|\mathbf{X}\|_0 = 1$, $\|\mathbf{X}\|_2 = 1$ to $\|\mathbf{X}\|_0 <= n_w$, $\|\mathbf{X}\|_2 <= n_w$ where n_w is the sub-window number of steps (i.e., the number of overlapping steps). In these cases, it would also be necessary to impose the constraint that only one element of $\mathbf{X}$ can be non-zero for each group of basis functions for each person (i.e., only one basis function per person identified). Finally, to identify the person(s) in the sub-window, we use the known label of the basis function corresponding to the nonzero coefficient in $\hat{X}$ after the sparse representation and take a plurality vote across all sensors in the sensing area.

4.3 Signal Reconstruction for Identified Footstep Removal

The next module of our approach involves sequentially and recursively reconstructing each identified footstep response in the footstep event and subtracting that reconstructed response from the combined (multiple person response). The remaining signal then represents the combined footstep responses from the remaining persons who generated the footstep event response.

As discussed in Section 3, the intuition behind our approach is that a vibration signal can be represented as the summation of the components corresponding to each excitation source (i.e., superposition). The challenge is that only the magnitude of the frequency spectrum is known from the sparse representation procedure described above. As a result, the timing (i.e., phase) of the signal is unknown. In addition, because of the constraint on the ℓ_2-Norm of X, the scaling of the reconstructed signal is also unknown. To address this, we leverage prior work [25] to reconstruct the overall signal by optimizing a scaled and time shifted version of the magnitude spectrum with the phase of the prior step dictionary signals corresponding to the index of the nonzero coefficient of $\hat{X}$.

We find this scale factor and time-shift using a sparse representation with a Euclidean distance error function where the new dictionary, D^*, contains time-shifted versions of the original prior step basis function corresponding to the index of the nonzero coefficient of $\hat{X}$. Based on the observed time shifts in our preliminary experiments,

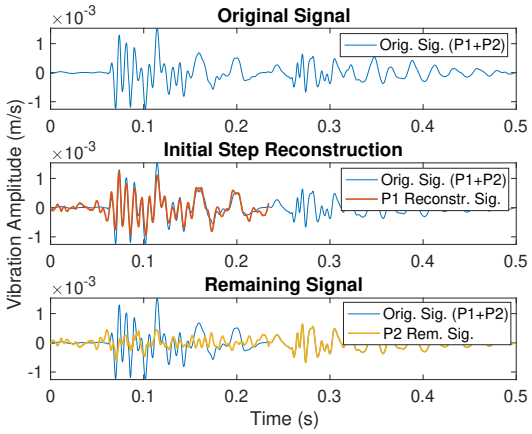


Fig. 4. Example of signal reconstruction and identified footstep removal

this new dictionary is bound by a shift of -2000 to 2000 samples (approximately +/- 0.08s). Therefore, the time shift and scale factor are determined with the following expression:

$$\hat{X^*} = \arg\min_{\mathrm{X}} ||Y^* - D^* X^*||_2^2, \quad s.t. ||X^*||_0 = 10 \tag{7}$$

where Y^* is the time series response for the current signal sub-window. For this sparse representation, we use a smoothed version of this time series response (100 sample smoothing window) to reduce prediction errors due to minor variations in the reconstructed signal. Additionally, for this sparse representation, we acknowledge that while prior steps from an individual may be *similar* to the current step from that same individual, they are not *exactly* the same. For example, there may be some minor variations within a persons gait pattern and/or changes in the environmental noise. As such, the basis function used to correctly identify the individual may not perfectly match the new signal during reconstruction. To address these differences, we relax the sparsity constraint in the coefficient vector X^* to 10 non-zero elements. This relaxation allows for our approach to find a more accurate time series reconstruction of the identified individuals response and results in fewer errors in identifying the next footstep response (i.e., if there are differences they can result in the system thinking another step exists at the same time as the already identified step).

We next use the output of the above sparse representation to reconstruct the identified signal as the product of the new dictionary, D^*, and the estimated sparse coefficient vector, $\hat{X^*}$. We denote this reconstructed signal as $\hat{Y_k^*}$, where k represents the kth signal sub-window. Our approach next removes the components of the identified and reconstructed step from original footstep event signal (i.e., the signal from multiple walkers) by subtracting $\hat{Y_k^*}$ from the original footstep event signal.

The resulting signal then contains the information for each other walker in the area. As such, the process outlined above (i.e., the sub-window number of step classification, sparse representation, and reconstructed step removal) is repeated for the next signal sub-window and each subsequent, sequential sub-window until all concurrent walkers' signals have been isolated/reconstructed, and those persons have been identified. Figure 4 shows an example of this process where we reconstruct the first step in the overall signal, remove it, and then use

Table 1. Recursive Sparse Representation Variables

Variable Name	Description
$\mathbf{S}$	Occupant Identity Index Vector
$\mathbf{D_j}$	Basis Function j (Prior Step)
$\mathbf{\Phi}$	Phase Matrix for Prior Steps
$\mathbf{Y}$	Mixed Vibration Response
m	Number of Basis Functions
n_s	Number of Steps
t_{n_s}	Step Start Time
τ	Time Shift

Algorithm 1 Recursive Sparse Representation-Based Multiple Person ID

$n \leftarrow size(\mathbf{D_1}, 1)$
$\mathbf{D_j} \leftarrow \frac{\mathcal{F}(\mathbf{D_j(t)})}{\sum_{i=1}^{n} D_{ij}^2}$
$\mathbf{D} \leftarrow [|\mathbf{D_1}|, |\mathbf{D_2}|, ...|\mathbf{D_m}|, |\mathbf{D_1}| + |\mathbf{D_2}|, |\mathbf{D_1}| + |\mathbf{D_3}|, ...|\mathbf{D_{m-1}}| + |\mathbf{D_m}|]$
for $k \leq n_s$ **do**
  $\mathbf{Y_k} \leftarrow \mathbf{Y}(\mathbf{t_{n_s}} : (\mathbf{t_{n_s}} + \mathbf{n} - \mathbf{1})$
  $\mathbf{Y_k} \leftarrow \mathbf{DX}$
  Estimate $\mathbf{X}$, *subject to* $||X||_0 = 1, ||X||_2 = 1$
  $\hat{\mathbf{X}} \leftarrow \arg\min_{\mathbf{X}} 1 - \frac{\sum_{i=1}^{n} y_{ki}(\mathbf{DX})_i}{\sqrt{\sum_{i=1}^{n} y_{ki}^2}\sqrt{\sum_{i=1}^{n}(\mathbf{DX})_i^2}}$
  $\mathbf{S(k)} \leftarrow l : \hat{\mathbf{X}}(\mathbf{l}) = 1$
  $\mathbf{\Phi} \leftarrow [Im[\mathcal{F}(\mathbf{D(t)}_{l,\tau_1})], Im[\mathcal{F}(\mathbf{D(t)}_{l,\tau_2}), ...Im[\mathcal{F}(\mathbf{D(t)}_{l,\tau_{n_t}})]]$
  $\mathbf{D}^* \leftarrow \mathcal{F}^{-1}(Re[\mathcal{F}(\mathbf{D(t)}_l] + 1i * \mathbf{\Phi})$
  Estimate $\mathbf{X}^*$, *subject to* $||\mathbf{X}^*||_0 = 10$
  $\hat{\mathbf{X}}^* \leftarrow \arg\min_{\mathbf{X}^*} ||\mathbf{Y}^* - \mathbf{D}^*\mathbf{X}^*||_2^2$
  $\mathbf{Y_k^*} \leftarrow \mathbf{D}^*\hat{\mathbf{X}}^*$
  $\mathbf{Y} \leftarrow \mathbf{Y} - \mathbf{Y_k^*}$
  Estimate t_{n_s} ▷ New Step Start Time
end for
return $\mathbf{S}$, the identities of each occupant in the footstep event window

the remaining signal to identify the next step. An overview of our approach algorithm is provided in Algorithm 1 below, with the relevant system variables described in Table 1.

Using our approach, our system overcomes the challenges of mixed/overlapping multiple person responses that are co-dependent on the underlying structure. Once each occupant's step has been identified in the footstep event window, we output the IDs of that window and continue to the next footstep event.

4.4 Trace-Level Dictionary Updating

Once our approach has identified a series of "step-level" predictions, we incorporate a "trace-level" updating step to the framework. By incorporating a "trace-level" updating step, our algorithm reduces the volatility of predictions from one step to the next. This stage of our algorithm is based on insight from normal human walking behavior. For example, if two or more individuals are walking down a hallway, the information contained in the

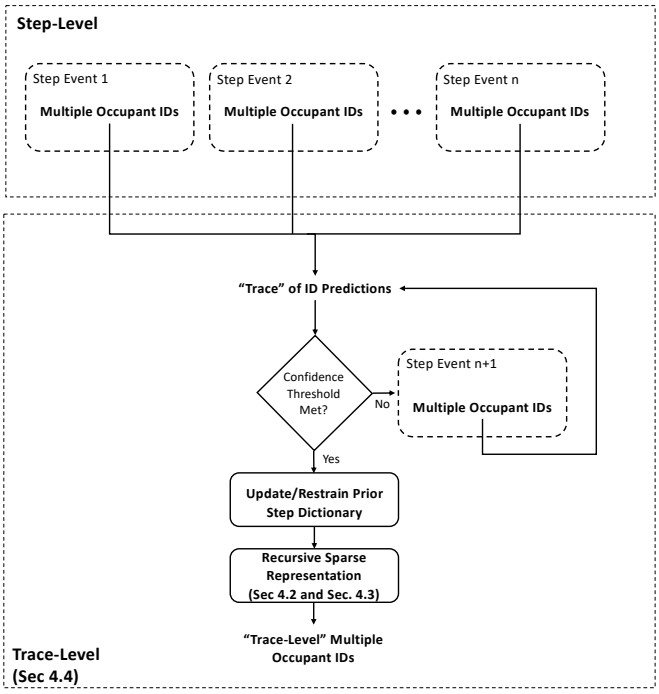


Fig. 5. Trace-level dictionary updating approach overview. Once each step has been identified at the step level, our system identifies the most frequently identified individuals over several consecutive steps and recursively updates step-level predictions.

series of footstep events should be limited to only those individuals walking, and an individual step within that "trace" of footsteps is unlikely to belong to a person who was not otherwise detected in the sensing area at that time.

In this step, we aggregate the predictions from several (typically 10 or more) consecutive footstep events (defined as "trace level") across each sensor in the sensing area. Using these aggregated predictions, we find the persons with the highest number of predicted steps and restrain our dictionary to only those persons. In doing so, we remove the potential for extraneous predictions and increase the confidence in accurately identifying each individual walker. For example, if 9 out of 10 consecutive step events predict an identity of Person 1 and Person 2, it is likely that the one event that is predicted as other persons (e.g., Person 7 and Person 8), actually contains steps from persons 1 and 2, so our model updates that step event's predictions to match the others in the trace.

To reduce the possibility of incorrectly updating the dictionary, we only update the dictionary after a confidence level has been reached. This confidence limit is empirically set such that the "n" walkers should each represent at least $60/n\%$ of the total predictions at the trace level and the "nth" most prominent prediction should represent at least $50/n\%$ of the total predictions. For example, in a 2 concurrent walker scenario, if, after 10 footstep events, 30% of the step-level estimates predict "P1", 40% predict "P7", and the remaining 30% are some combination of the other potential walkers, the dictionary updating would restrain the dictionary to those basis functions corresponding to Person 1 and Person 7. If the confidence threshold is not met after the initial trace of steps, it is re-evaluated with each new footstep event until a trace-level prediction can be made. In the event that the

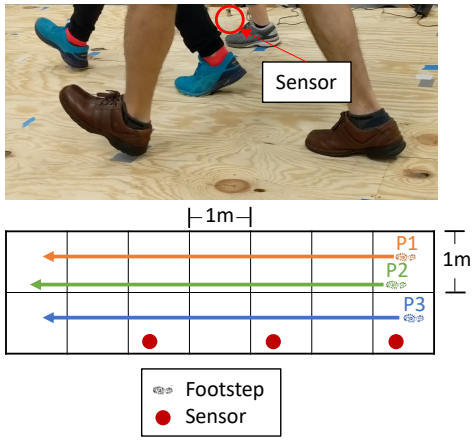


Fig. 6. Experimental Setup. Typical sensor layout and walking trajectory for the three walking participants.

confidence levels are never reached, the system updates the dictionary to the most prominent predictions across the entire set of footstep events in the sensing area once the walkers have left the area. This ensures that there is consistency to the predictions even if the model is not confident in those predictions across the entire sequence of footstep events.

In Algorithm 1, this "trace-level" updating results in a modified D, where the dictionary only contains the basis functions corresponding to the most-frequently identified persons in the walking trace. For this work, we restrain the dictionary updating such that the number of persons in the updated dictionary is equal to the estimated total number of walkers in the sensing area (using the approach outlined in Section 4.1). With this updated dictionary, the remainder of the recursive sparse representation is computed, and our system outputs the final "step-level" predictions for the occupant identities as well as their separated vibration responses. The following section provides an overview of our system performance through real-world walking experiments with multiple walkers.

5 MULTIPLE PERSON IDENTIFICATION EVALUATION

To validate the performance of our system, we have conducted a series of real-world experiments. These experiments are three-fold. First, we validate our model performance with experiments involving three concurrent walkers in a campus building (Section 5.1). Next, we evaluate the model performance using hybrid real-world and simulated data involving 10 total participants (Section 5.3). Finally, we validate our primary assumptions of response superposition and structure co-dependence in Section 5.2 and Section 5.4, respectively.

5.1 Real-World Walking Experiments

In our first experimental validation, we conducted a series of real-world walking experiments with three concurrent walkers in a campus building. We conducted this experimental evaluation in conjunction with our approved IRB protocols from Carnegie Mellon University (CMU) and Stanford University (CMU: 2015-00000125, Stanford: IRB-54912). In this section, we first present the experimental setup, then discuss the performance of our approach for uniquely identifying the three concurrent walkers.

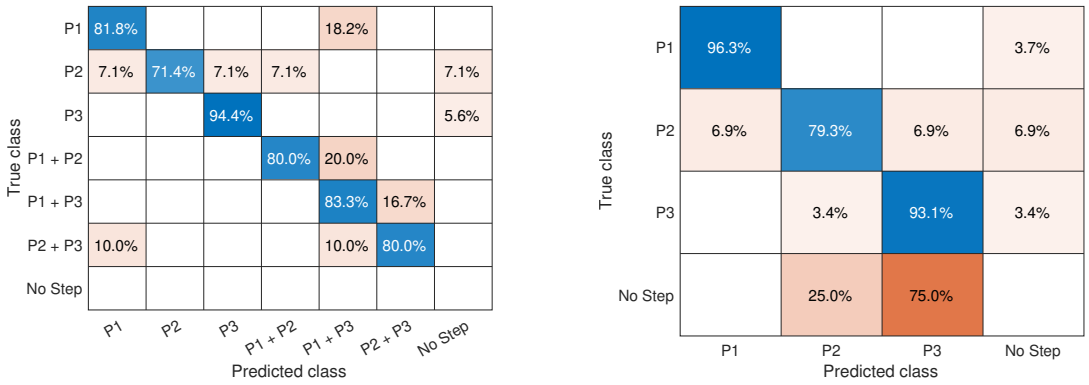
5.1.1 Experimental Setup. Our sensing system uses geophone sensors to measure footstep-induced floor vibrations. For our experiments, we placed three SM-24 geophone sensors on a wood floor test bed in our lab [38]. The three experimental participants were then asked to walk in a straight-line side-by-side for several consecutive steps, and then repeated this process 5 times (for approximately 90 total steps between all three people). To reduce inconsistencies with initiating walking, we removed the first step taken by each person from the experimental data set for a total of 85 steps in our evaluation dataset. In addition, each of the participants was asked to walk by themselves several times to provide data for constructing the prior foot-step dictionary. In each set of experiments, the participants were not required to wear any specific shoes, but were asked to use the same shoe for each walking iteration for consistency of data collection. Figure 6 shows an overview of the experimental setup.

5.1.2 Three Concurrent Walker Results. In this section, we evaluate the performance of our approach in identifying the three walking participants. The number of concurrent walkers (3) is representative of human social tendencies. In fact, in a study by Moussaid et al.[55], the authors studied human walking tendencies in real-world environments in two different settings. From these two scenarios, they observed that, for scenario A, walkers were in groups approximately 35% of the time, and, for scenario B, approximately 50% of the time. Further, a group size of 2 concurrent walkers accounted for approximately 30% of the total instances in A and approximately 40% in B. Higher groups of concurrent walkers (i.e., more than 3 concurrent walkers) were very small (less than 5% of all instances). As such, our selection of 3 concurrent walkers this evaluation is representative of (and perhaps more complex than) many real-world human social tendencies.

For this evaluation, we show our system performance in two ways: 1) we determine how well our approach uniquely identifies both solitary and partially overlapping steps (e.g., "P1", "P2", "P3"), as well as completely overlapping steps (e.g., "P1 + P2"). These results are shown in Figure 7a. Then, 2) we consider the completely overlapping cases in more detail by separately evaluating our approach performance for separately identifying each person (e.g., identify "P1" and "P2" from "P1 + P2"). This is combined with the solitary/partially overlapping performance and the results are shown in Figure 7b. In each confusion matrix, each accuracy percentage represents the number of correctly identified steps from each person/combination divided by the total number of steps for that person. In each confusion matrix, the "No Step" class represents instances when a footstep did not occur, but may have been predicted. For example, an erroneous "No Step" prediction represents situations where a step was not detected, but actually occurred, or where our system predicts fewer steps than are actually present (e.g., the system predicts 1 step in a footstep event, but there were actually 2).

From these confusion matrices, we can observe that our approach is able to accurately identify individuals when there are three concurrent walkers in the sensing area. In the first case, we identify individuals with as much as 94.4% accuracy. In addition, for cases of perfectly overlapping steps (e.g., "P1 + P2") our approach always identifies at least one of the two walkers and identifies both walkers with as high as 83% accuracy.

In the second case, we evaluate our model performance by computing the average F1 score across each of the classes/identities. The F1 score is a common approach for evaluating classification accuracy[82]. In this work we define the parameters of the F1 score as follows: $TruePositive_n$ represents the number of correctly identified steps for person n, $False\ Positive_n$ represents the number of steps belonging to Person n, but identified as a different person, and $FalseNegative_n$ represents the number of steps belonging to other persons, but identified as person n. In this way, we compute an average F1 score across all three persons of 0.89 for our approach. We compared this result with a baseline approach that uses a multi-class Support Vector Machine (SVM) for multiple step identification and found that our approach results in a 1.3× improvement over the baseline (0.68 F1 score). We also note that Persons "P1" and "P3" both have accuracy above 90%, while Person "P2" has a slightly lower accuracy by comparison. Based on our observations from prior work, we infer that the lower accuracy for "P2" is likely due to a prominent heel- and toe-strike for that individual when walking. As a result, there are often two impulsive signals for each step, and it is possible that each part of the overall footstep response is detected



(a) Performance with three concurrent walkers with overlapping steps.

(b) Total performance with three concurrent walkers.

Fig. 7. Confusion matrices showing the overall performance of our approach with three concurrent walkers. (a) shows the accuracy for individual step detection and instances of "completely overlapping" steps; (b) shows the accuracy for individual steps and separately evaluates each person within a "completely overlapping" step.

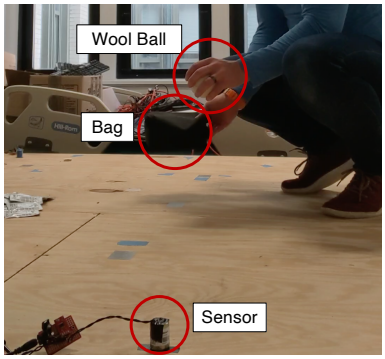
separately (i.e., the system thinks there are two steps when there is only one). Our future work plans to address these scenarios by characterizing the individual heel and toe components of footstep-induced vibration responses so that each can be accounted for in the overall response.

5.2 Superposition Validation

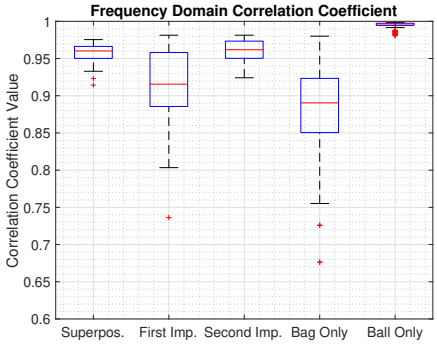
In this section, we validate our assumption of vibration response superposition through real-world experimentation. As discussed in Section 3, one of the primary assumptions of our approach is that the multiple walker vibration responses represent the superposition of individual responses from each walker. This assumption also relies on the assumption that the dynamic properties of the underlying structure are linear-elastic.

To validate this assumption we conducted a series of experiments on our test structure involving other sources of impuslive excitations: a ball drop and a bag drop. We chose these objects based on the insight that they produce more impulsive excitations than footstep responses, and their input excitations are more easily controlled (i.e., drop the items from the same height and same location) as compared to walking excitations from footsteps (people walk in different ways, with varying footstep ground reaction force, and in slightly different locations each time). Figure 8a shows the experimental setup for these experiments. A single geophone sensor was used in these experiments and a cloth bag and wool ball were dropped in quick sequence (bag first, then ball with a very short delay so that responses were not perfectly overlapping). We collected approximately 30 samples of these mixed/overlapping responses. Then, the cloth bag and wool ball were each dropped approximately 20 more times each by themselves (i.e., single responses, not overlapping/mixed).

For this evaluation, we modeled the combined bag and ball response as the superposition of the individual responses to determine how well the superposition matched the combined response. To determine the individual responses for the superposition, we used our recursive sparse representation approach. The combined responses were then compared to the modeled signal from the sparse representation reconstruction. In this way, we evaluate whether the combined signal for controlled excitations is able to be represented by the superposition of the individual excitations from each object.



(a) Experimental procedure for superposition evaluation.



(b) Boxplot showing performance of superposition compared to individual responses.

Fig. 8. Superposition Evaluation. (a) shows the typical setup for the bag and ball drop experiments; (b) shows a box plot of the results. Note that the superposition has similar correlation to individual responses.

To evaluate the similarity between the mixed responses and the superposition responses, we compared the frequency domain responses for each pair of signals (e.g., combined response 1 vs. superimposed response 1, combined response 2 vs. superimposed response 2, and so on) using a correlation coefficient measurement. In an ideal case, the correlation coefficient measurement would take a value of 1.0, and, if the signals are perfectly uncorrelated, would take a value of 0. Figure 8b shows a box plot of the distribution of correlation coefficient values in the experimental dataset. In this plot, the blue boxes represent the interquartile range of the correlation coefficients, the "whiskers" (dashed lines) represent the range of values, the red asterisks show outliers, and the red horizontal line indicates the median of the distribution. In this figure, we compare the correlation between the mixed signal and the superimposed signal (labeled "Superpos."), then separately considered the correlation between the first impulses in the superposition reconstruction (i.e., reconstructed bag vs. actual bag), then did the same analysis for the second impulse (i.e., reconstructed ball vs. actual ball), and, lastly, compared actual ball drop and bag responses with other ball and bag drop responses (for reference information).

From this analysis, we observe a very high correlation between the superposition results and the measured mixed response (median value of 0.96), and a similar distribution for the comparison of individual responses to their reconstructed counterparts. Due to the nature of the bag object, there is a wider distribution of responses (i.e., if it fell on the edge or a flat side, or if it was condensed when striking the floor), while the ball is significantly more uniform and has more consistent responses.

Through this evaluation, we have shown that the simultaneous dropping of a ball and weighted bag produces a comparable signal to the simulated superposition of each individual response. These results support the assumption of superposition in our approach, and that prior steps (basis functions) can be used to represent new steps by the same person. By using inanimate objects (the bag and ball), we further reduce the likelihood that the validity of superposition and the basis functions is only due to the general signal similarity of footstep responses. Lastly, the results of this evaluation support the use of the "hybrid" dataset for our 10-person experiments. Additional details of that evaluation are provided in the next section.

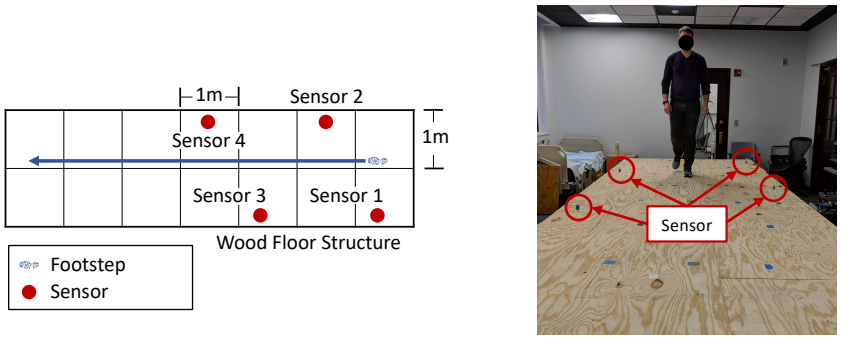


Fig. 9. Experimental Setup. Typical sensor layout and walking trajectory for the hybrid real-world and simulation data.

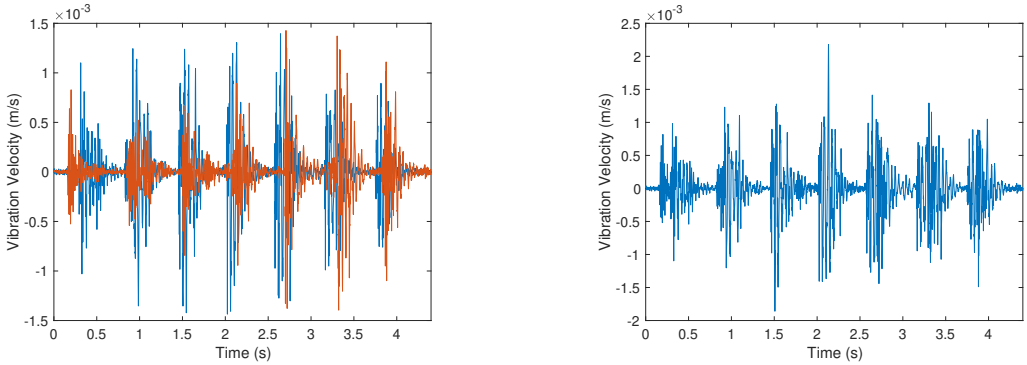
5.3 Hybrid Simulation and Real-World Validation

In this section, we evaluate the performance of our approach for identifying multiple concurrent walkers using a "hybrid" dataset of 10 different walking participants. As before, we conducted this experimental evaluation in conjunction with our approved IRBs (CMU: 2015-00000125, Stanford: IRB-54912), and an approved COVID-19 pandemic protocol. Due to the restrictions for human subject research during the COVID-19 pandemic, experiments that involved multiple side-by-side walkers were not permitted. Therefore, in order to demonstrate the performance of our approach, we conducted a series of experiments with 10 individuals walking by themselves in our laboratory space, and then simulated multiple sets of walkers walking concurrently by adding their signals together. In this section, we refer to this combined real-world and simulated data as a "hybrid" dataset.

We conducted a series of "hybrid" experiments with this dataset to showcase the overall accuracy and robustness of our approach. For these analyses, the "trace-level" number of concurrent walkers (from Section 4.1) was assumed to be known. This allows comparison of the sparse representation approach directly without error propagation due to incorrectly identifying the number of concurrent walkers, and also reduces errors due to the effect of additive noise when adding multiple sensor signals together. First, we describe the experimental setup for this set of data. Then, in Section 5.3.2, we evaluate the performance for each possible pair of persons walking (i.e., $P1P2$, $P1P3$, ... $P_{n-1}P_n$). Next, in Section 5.3.3, we explore the performance of our model with respect to the amount of signal overlap. Finally, in Section 5.3.4 we explore the model accuracy with respect to each sensor in the sensing area and compare results with additional numbers of concurrent walkers.

5.3.1 Hybrid Data Experimental Setup. As described in Section 5.1.1 above, our work uses SM-24 geophone sensors to collect footstep-induced floor vibration data. For the "hybrid" dataset, we collected walking data from 10 different experimental participants on an elevated wood-framed floor structure in our laboratory space. As previously discussed, due to COVID-19 restrictions, only one walker was permitted to walk in the experimentation area at a time. Each participant walked from one side of the floor structure to the other at a normal, natural walking pace. In each set of experiments, the participants were not required to wear any specific shoes, but were asked to use the same shoe for each walking iteration for consistency of data collection. Figure 9 shows an example of the typical walking path for the participants as well as the sensor layout. At least 12 walking repetitions were collected from each walker, for a total of approximately 840 footstep responses. These signals are then used in the hybrid analysis described below.

5.3.2 Performance for Two Concurrent Walkers. We first evaluate the performance of our model in the hybrid dataset for situations involving two concurrent walkers. To simulate the "two walkers" data, we combined



(a) Two different walkers' raw vibration signals.

(b) Simulated "mixed" signal from two walkers.

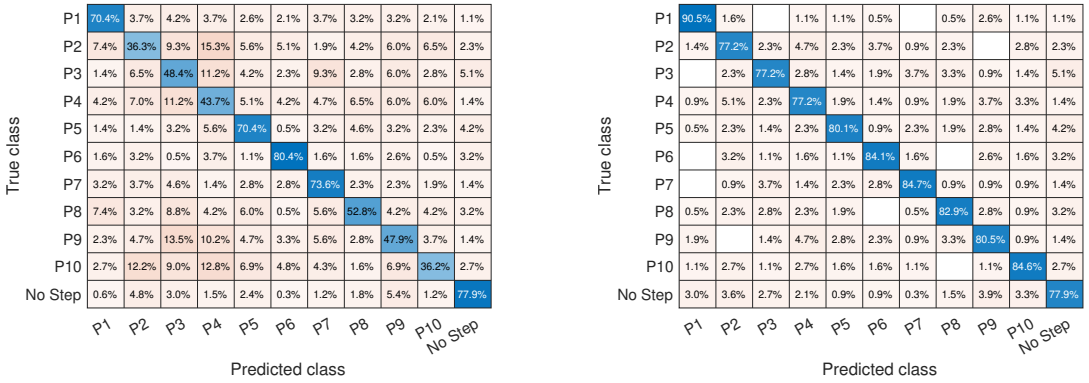
Fig. 10. Simulation Raw Data. a) shows the raw signal from one person in blue, and the additional raw signal from another person in orange; b) shows the resulting combined signal from the two signals from (a) are added together to create the 'hybrid' overlapping data.

the vibration responses from each possible combination of the 10 total participants (45 total pairs). In each combination, three of the walking repetitions from each person were randomly selected for test data, and the other nine were used for training data. To prevent each footstep response from perfectly overlapping the other person, an initial offset of the signals was applied (empirically set at 0.2s), and the order of the two signals was randomized (e.g., P1 first then P3, or P3 first then P1). Figure 10 shows an example of this procedure. In Figure 10a the blue signal represents one walkers series of footstep responses, while the orange signal shows a different person. Note the initial offset (separation between initial peaks). Despite this initial offset, many of the latter footstep responses have a large amount of overlap. Figure 10b shows the ensuing "simulated" signal which is used for further analysis.

To evaluate the performance of our approach, we aggregated the results of each combination of walkers. When there were instances of completely overlapping steps (e.g., P1+P2), we separately evaluate if our approach correctly identified Person 1 and Person 2. For example, if the system identified P1+P3, but the actual footstep event was generated by Person 1 and Person 2, our system would have correctly identified Person 1, but incorrectly identified Person 2. As described in Section 4.1, we empirically set the threshold for completely overlapping steps as 3000 samples (0.11s).

Figure 11 shows two confusion matrices summarizing the results. In Figure 11a we show the results using the full 10-person basis function dictionary, while in Figure 11b we show the results after the dictionary updating step (as described in Section 4.4). In this case, we are showing the results when the dictionary is updated to the two walkers identified at the trace level. In this hybrid dataset, a "trace" of footsteps was considered to be the entire test dataset (approximately 25 total footstep events, on average). The "trace level" predictions were then calculated using the process outlined in Section 4.4 above. Of the 45 test datasets, we note that 21 of them met the confidence threshold for trace-level updating within 10 total step events, 15 required just 6 footstep events, and the median number of step events was 12. This shows that our approach can accurately identify the two persons walking with a small number of footstep events. In each confusion matrix, the "No Step" class is the same as described in Section 5.1.2 above.

We compared the results of our model predictions to those obtained using a naive baseline approach. In this case, the baseline approach is one where the superposition/reconstruction step is used, but a multi-class



(a) Hybrid data model performance with 2 concurrent walkers without trace-level updating.

(b) Hybrid data model performance with 2 concurrent walkers after trace-level updating.

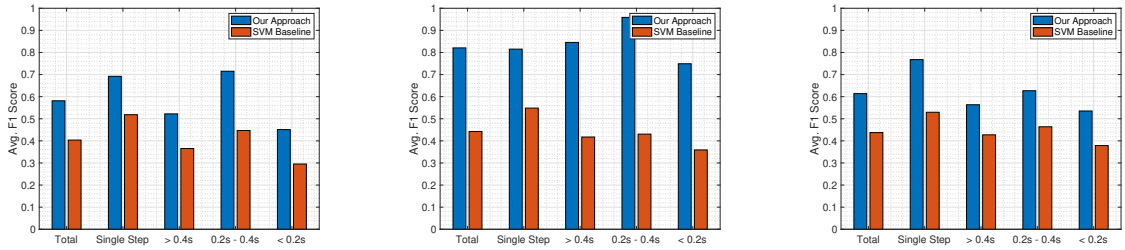
Fig. 11. Overall hybrid data results with 10 different walkers. (a) shows the step-level performance without trace-level dictionary updating; (b) shows the step-level performance after trace-level dictionary updating to 2 walkers.

support vector machine classifier is used to identify footstep responses in lieu of the cosine distance-based sparse representation proposed by our work. For the trace-level updating results, our approach achieved an average F1 score across all 10 persons of 0.81, which is a 2.9× reduction in error from the naive baseline approach (avg. F1 score of 0.44). Further, we note that our approach has similar performance for each of the 10 participants, indicating that it is robust to different persons and different walking styles.

5.3.3 Influence of Amount of Signal Overlap. In this section, we further investigate the performance of our approach with respect to the degree of signal overlap between concurrent walkers. As described above, the simulated data was generated by adding two different walkers signals together with some small amount of initial signal offset. This was done to best represent real world walking conditions, where it is unusual for individuals to begin walking at *exactly* the same time. Instead, it is common for persons to begin walking at approximately the same time, and they may continue walking asynchronously, or may begin to synchronize as they walk side-by-side. As such, the actual degree of signal overlap varies on a step-by-step basis.

We evaluate our model performance with respect to the amount of signal overlap for each step. We compute this amount of overlap by taking the difference (in time) between the signal onset for each walkers step. Based on the observed distribution of these offsets, the steps were then clustered into 4 distinct bins: 1) a 'single step' where there is enough separation between consecutive footsteps from each walker that the entire signal from each is clear and distinct; 2) when the amount of offset is greater than 0.4s; 3) when the amount of offset is between 0.2 and 0.4s; and 4) when the amount of offset is less than 0.2s.

Figure 12 shows a summary of these results for three different scenarios: 1) where there is no trace-level dictionary updating (Figure 12a); 2) where the dictionary is updated to be limited to the two most likely walkers (Figure 12b); and 3) where the dictionary is updated to be limited to the three most likely walkers (i.e., the total number of walkers plus one; Figure 12c). In each case, we compare the overall results with those for varying degrees of offset and compare our approach to the SVM-based naive baseline approach. Note that undetected steps (i.e., those that are considered "no step" in Section 5.3.2 above) are excluded from this analysis, as there is not any footstep present to determine the "offset". We observe that our approach has similar performance across each amount of offset in each scenario, and consistently outperforms the naive baseline approach. These



(a) Overlap Analysis without trace-level updating.

(b) Overlap Analysis with trace-level updating to 2 persons.

(c) Overlap Analysis with trace-level updating to 3 persons.

Fig. 12. Overlap Analysis. Our approach significantly outperforms the baseline approach across varying amounts of signal overlap for two concurrent walkers.

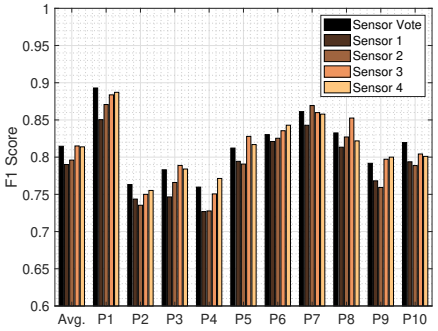
results show that our superposition-based sparse representation and signal reconstruction accurately isolates and identifies individual footstep responses even in cases where there is significant signal overlap due to multiple concurrent walkers.

5.3.4 Influence of Sensors and Step Breakdown Analysis. To further evaluate the performance of our approach with the hybrid dataset, we explored the impact of the sensor vote and the performance across each step inside of a footstep event (i.e. the "first" and "second" step if a footstep event contains multiple footstep responses. For this analysis, we are using the results when a 2 person trace-level dictionary updating is performed.

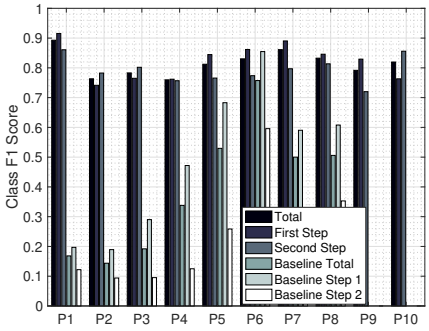
For the sensor comparison, we considered the accuracy of our approach (which uses a plurality vote across all four sensors) and compare that to the average and per-person F1 score for each individual sensor by itself. The results are shown in Figure 13a. We observe that the sensor vote has better or comparable accuracy for each person (and the average) when compared to the "best" sensor of the four. This indicates that the sensor vote helps improve model performance and reduce the likelihood of erroneous identification (i.e., performs better than the "worst" sensor in each instance). However, it is interesting to note that the "best" sensor has similar performance to the sensor vote in each case, and, most cases (e.g., P3, P4, P5, P6, P7, P8, P9), outperforms the sensor vote. This indicates that as few as one sensor can be used to identify walkers in multiple walker scenarios with high accuracy. However, when the "best" sensor outperforms the sensor vote, it is not the same sensor each time. As a result, the challenge with using one sensor is determining where to place the solitary sensor, and how to choose which one to use (in cases where multiple are in a sensing area). In our future work, we plan to explore these concepts of sensor selection/placement further. Some preliminary related work in this area suggests that sensing signal quality can be used to optimize sensor layout/placement in a given structure [87].

For the step-breakdown analysis, we compare the F1 score for our approach to the SVM-based naive baseline approach. The per-person F1 scores are broken down into separately evaluating the total results for each person, the results when that person's step was the first step in a footstep event window, and then when that person's step was the second step inside of a footstep event window. For example, if a footstep event contained two steps generated by Person 3 and Person 7, with Person 3's step being the first in the window and Person 7's step being second, that would be considered a "First Step" for Person 3 and a "Second Step" for Person 7.

The results of this analysis are shown in Figure 13b. From this figure, we observe that our approach achieves similar performance for the overall F1 score, the "First Step" F1 score, and the "Second Step" F1 score for each of the 10 experimental participants, and displays significant improvement over the naive baseline approach in each instance, in particular for the second step case. These results indicate that our recursive signal reconstruction for



(a) Sensor Vote compared to individual sensor performance.



(b) Per-person ID F1-score. Scores are broken down by first and second step in a footstep event window, and compared to a SVM-based baseline approach.

Fig. 13. Comparison of 2 concurrent walker results. (a) shows the sensor vote accuracy compared to each individual sensor, and (b) compares our approach with a naive baseline approach for per-person ID accuracy and with respect to first or second step in a footstep event window.

identified footstep removal step allows us to remove/reduce the influence of the initial step(s) in a window so that the latter step(s) can be more accurately identified, and that we do so with consistently high accuracy across each participant.

The individual results of Persons 6, 9, and 10 are particularly interesting and warrant further discussion. For Person 6, we observe that the "First Step" for P6 has similar accuracy for our approach (0.86 F1 score) and the baseline approach (0.85 F1 score). This likely is a result of the very high magnitude responses from P6 as compared to the other walking participants, which means that, when the steps from P6 are first, they dominate the overall response and are easily distinguished with either the sparse representation or a SVM classifier. Also of interest are the results from Persons 9 and 10, where we observe that the baseline approach does not correctly identify any steps from either person. This is due to the dictionary updating step, where the baseline approach does not ever identify Person 9 or Person 10 at the trace level, and, as a result, once the dictionary is updated there are no predictions for steps due to those persons. This likely occurs because the responses from Persons 9 and 10 are of relatively low magnitude, so they tend to be more difficult to distinguish from other steps, particularly when they are significantly overlapping. By using a cosine distance-based error function, our approach is able to overcome these differences in relative magnitude of the footstep responses in the combined signal, and produces more accurate predictions.

5.3.5 Robustness to Number of Concurrent Walkers. In the final evaluation with the "hybrid" dataset, we explored the performance of our system with respect to the number of concurrent walkers. For this analysis, we considered four scenarios: 1) one walker (for reference purposes); 2) two concurrent walkers; 3) three concurrent walkers; and 4) four concurrent walkers. To generate the datasets for scenarios 3) and 4), we randomly selected combinations of 3 and 4 of the 10 walkers in our dataset and combined their signals (in the same manner as described in Section 5.3.2). For initial signal offset, a similar approach was also taken to the one described above, with each signal being offset approximately 0.2s from the preceding one. In addition, similar to above, for testing data, 3 repetitions were randomly chosen from each concurrent walkers' data and combined to form one "trace" of test data, while the remaining 9 were used for training data (i.e., the basis function dictionary). For each of these

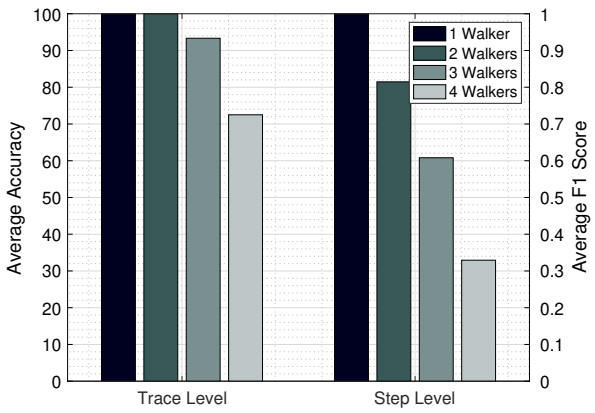
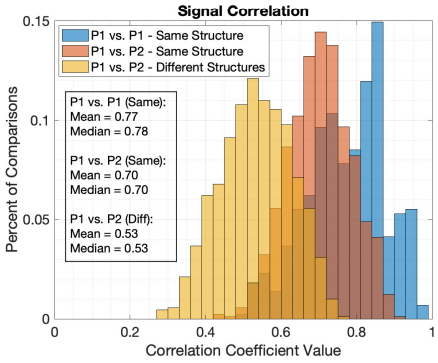


Fig. 14. Occupant ID results for different numbers of concurrent walkers. "Trace Level" shows accuracy for identifying the correct "n" walkers in the trace, and "Step Level" refers to the accuracy in identifying each individual step after trace-level dictionary updating.

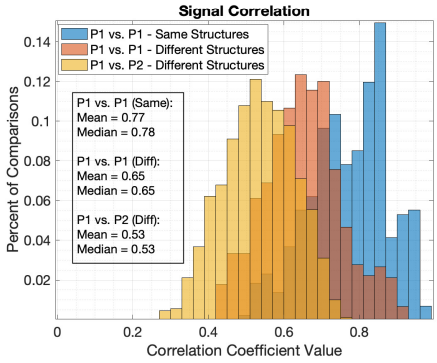
scenarios, we considered a total of 10 random combinations of different concurrent walkers, which provides at least 3 instances of each participant in the testing data.

We compared occupant identification results at the "trace level" and the "step level" after the dictionary updating step. At the trace level, we compute the average accuracy for identifying the correct "n" walkers out of 10 total walking participants. Then, at the step level, we compute the average F1 score for identifying each of the 10 walking participants. In each variation of the number of concurrent walkers, the dictionary updating step restricted the dictionary to the "n" most prominent predictions, where "n" is the number of concurrent walkers in the test dataset, following the procedure outlined in Section 4.4.

Figure 14 provides a summary of the evaluation results. From this figure, we observe that the model accuracy has high accuracy at the "trace level" with up to three concurrent walkers (100%, 100%, 93.3% avg. accuracy), and, even with four concurrent walkers, is able to accurately identify the correct 4 out 10 walkers with an average accuracy of 73%. At the step level, the model performance follows a consistent trend with increasing number of concurrent walkers, with a decrease in average F1 score of approximately 0.2 with each additional concurrent walker. This behavior follows the expected behavior; with each additional concurrent walker, there is an increasing likelihood of significant signal overlap, and instances where several steps occur at approximately the same time (i.e., multiple perfectly overlapping steps). In these cases, it is difficult to accurately extract each individual's components of the signal with high accuracy. In addition, the model performance for instances of several concurrent walkers for the hybrid dataset is likely worse than what would be observed in real-world conditions. This is due to the nature of the simulation, where multiple walker signals are generated by adding each individual signal. As more signals are added, this simulation process also increases the amount of noise (i.e., ambient noise is added "n" times to the combined signal). As such, the simulated signals are more adversely affected by the environmental noise than their real-world counterparts would be. Part of our future work will be to explore the limits of our multiple walker approach in real-world settings as there are increasingly more concurrent walkers in the sensing area. These results show that our approach is suitable for various scenarios of multiple concurrent walkers, and is able to correctly identify the persons walking in the sensing area (i.e., at the trace level) with high accuracy with as many as four concurrent walkers.



(a) Comparison of footstep responses with (1) the same person in the same structure; (2) two different persons when the walkers are in the same structure; and (3) two different persons when the walkers are in different structures.



(b) Comparison of footstep responses with (1) the same person in the same structure; (2) the same person in two different structures; and (3) two different persons when the walkers are in different structures.

Fig. 15. Structural Co-Dependency. (a) shows the correlation between responses across different persons; and (b) shows the correlation between responses for the same person across different structures. We note that the change in structure significantly reduces correlation between signals.

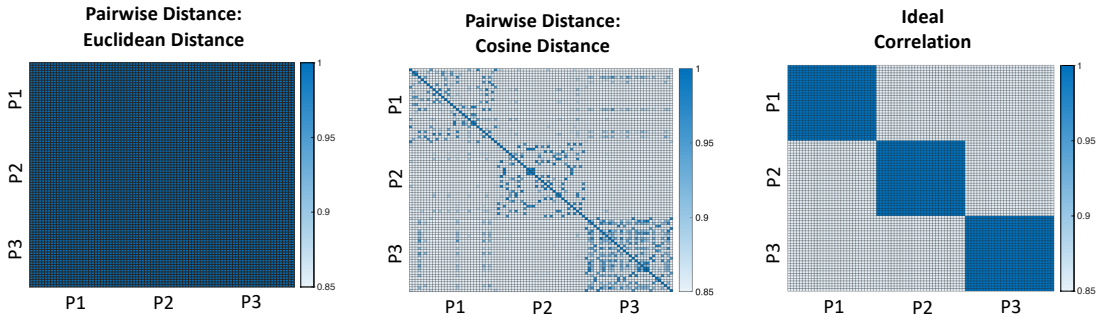
5.4 Structure Co-Dependence Evaluation

In the final evaluation of our approach, we consider the assumption and physical insight that individual vibration responses are co-dependent on the dynamic behavior of the underlying structure. As discussed above, this behavior results in failure to achieve independence of the excitation source-induced signals (which is a requirement for many Blind Source Separation approaches). Overcoming this challenge of structural co-dependence is one of the primary contributions of this work. In this section, we evaluate this assumption as well as our approach for reducing the effect of the structure codependency.

We first evaluate the assumption of codependency on the underlying structure. For this evaluation, we conducted a series of walking experiments with two different individuals across two different structures. One structure is the wood framed test structure (as described above), and the other is on the second floor of a steel framed building with metal deck and concrete topping slab for its floor structure. Each person walked at a normal, comfortable walking pace for several steps in both structures.

To understand the influence of structure codependency on the signal similarity, we computed the pairwise correlation coefficient between footstep responses. We computed these pairwise correlation coefficient values between different steps from Person 1, then did the same exercise for Person 1 compared to Person 2 when each person was in the same structure, and, lastly for Person 1 compared to Person 2 when Person 1 was in one structure, and Person 2 in the other. The resulting distribution of correlation coefficient values is shown in the histogram in Figure 15a. We observe a high correlation on average for both comparisons with persons in the same structure, and a significantly lower correlation when the two persons are in different structures. This reduction in correlation in different structures implies that the similarity between signals in the same structure is due to the codependency of responses on the dynamic properties of the structure itself.

We next compared the correlation coefficient values for Person 1 in the two different structures. First, we again considered different steps from Person 1 in the same structure, then compared steps from Person 1 in one



(a) Euclidean Distance (b) Cosine Distance (c) Ideal Correlation/Distance

Fig. 16. Correlation between three walker's responses using (a) Eucldean distance; and (b) Cosine Distance. Numbers represent (1 − *Distance*) for graphical representation purposes. Ideal measurements are shown in (c) and would be "0" for steps from the same person and "1" for steps from different person. Each block represents 1 step.

structure with steps from Person 1 in the other structure, and, finally, the comparison of Person 1 and Person 2 in different structures (for reference). The resulting distribution of correlation coefficient values is shown in Figure 15b. In this figure, we observe a drop in correlation for Person 1 in two different structures, and further observe that this drop is lower than the decrease in correlation for two different persons in the same structure (from Figure 15a). These results and the observation of the general decrease in correlation in different structures help support the validity of our insight regarding structural codependency, and our assumption of the dominant effect of the structure components of the signal. Part of our future work will focus on characterizing the structural component and footstep-induced component of signals, which will allow us to further study and validate these assumptions. We discuss this future work in more detail in Section 6.1.3.

In addition to evaluating the assumption of structure codependency, we evaluated the effectiveness of our cosine distance-based error function. To do this, we compared the similarity between footstep-induced vibration responses from three of our experimental participants. For this similarity comparison, we computed the pairwise distance between time series data for two different footstep responses using a Euclidean distance measurement and a cosine distance measurement. We repeated this for 30 total footsteps from each of the three persons for a total of 900 step comparisons.

Figure 16 summarizes the results of these comparisons. For visualization purposes, the values shown represent 1 − *Distance* so that a value of 1.0 represents a perfect correlation between responses (i.e., zero distance), and a value of 0 represents perfectly dissimilar responses (i.e. distance value of 1.0). For reference purposes, the ideal case is shown in Figure 16c. The ideal case would be for responses from the same person to be perfectly correlated with themselves (i.e., all of Person 1's steps are perfectly correlated with other steps from Person 1; distance of 0 and graphical value of 1.0), and responses between different people (e.g., Person 1 compared to Person 2) would have a value of 0 (i.e., perfectly dissimilar).

We can then compare this ideal case to the values obtained using a Euclidean distance measurement (Figure 16a) and a cosine distance measurement (Figure 16b). In this way, we observe that nearly all of the distance values using the Euclidean distance take a value of nearly 1.0, even when different pairs of participants are considered. This is likely the effect of the structural codependency previously discussed, and would result in confusion between the responses from different people when this distance metric is used for identification purposes. In contrast, the cosine distance metric used by our approach results in values much closer to the ideal case, and

there is a clear distinction between distance values for same- and different-participant comparisons. As a result, by using the cosine distance error function, our approach is able to reduce the influence of the codependency on the underlying structure and improve multiple walker identification accuracy.

6 DISCUSSION AND FUTURE WORK

In this paper we present a novel approach for separating footstep-induced vibration signals from multiple concurrent walkers, which enables unique identification of each walker. Through a real-world experimental evaluation we showed that our approach achieves high accuracy for identifying up to 3 concurrent walkers. Further, using a hybrid dataset, up to 10 different walkers were uniquely identified in multiple walker scenarios. Based on these results, our approach shows promise for use in signal separation and identification applications. In this section, we discuss relevant assumptions and limitations of our approach, as well as potential future research directions. These future directions can be categorized in two main directions: 1) System Scalability/Model Transfer; and 2) Newcomer Identification. The following sections explore each of these topics in more detail.

6.1 System Scalability

The subject of scalability in this work has three facets: 1) scalability of the sensing system, 2) scalability of the source separation algorithm and basis function (prior step) dictionary, and 3) scalability of the approach across different structures.

6.1.1 Scalabillity of the Sensing System. The first aspect of scalability (the sensing system) is primarily related to the deployment and maintenance costs. The geophone vibration sensors used for this work are low-cost [64], and have a sensing range of as much as 20 meters (depending on the structure [66]. Further, they can be easily installed in existing structures and require only a coupling with the floor structure (typically by way of wax). As such, the system can scale to building-wide sensing with relatively little effort and cost. In addition, we have developed a wireless version of our sensing system to reduce cabling and maintenance costs. To achieve the multiple walker identification discussed in this work, as little as one sensor is required in the walking area.

As discussed in Section 5.3.4, the challenge with using only one sensor, however, is how to chose the location for deploying the sensor, and/or how to choose which sensor's data to use (in cases where multiple sensors have already been deployed). Prior works using footstep-induced vibration sensing suggest that signal-to-noise ratio (SNR) alone may not be sufficient for determining which sensor to use and/or where to place it in the structure [87]. In our future work, we plan to explore sensor selection and placement for deploying our system in real-world environments so that the number of sensors required to achieve accurate indoor occupant monitoring (for applications such as identification) is optimized.

6.1.2 Scalabillity of the Source Separation Algorithm. The second aspect of scalability involves the sparse representation approach presented in this work. An important assumption in our work is that the current footstep response from one individual is similar to a prior response from that person and that this response is included in the basis function dictionary. For the majority of walking scenarios, this assumption is valid (e.g., when a person is walking "normally"). However, in some instances, individuals may be walking much more quickly or slowly, or they may have had some injury that changes how they walk. In these cases, our assumption of prior information in the basis function dictionary may not be valid.

Additionally, with regard to "completely overlapping" footstep responses (i.e., when 2 or more steps occur at the same time), this work addresses those situations by adding additional basis functions which are linear combinations of the dictionary elements for each of the potential walkers. This allows us to maintain the sparsity constraints of $\|\mathbf{X}\|_0 = 1, \|\mathbf{X}\|_2 = 1$ by selecting the overlapping step basis function. At large scale, this approach

becomes very resource intensive (i.e., the dictionary becomes very large and the computation time will increase significantly).

In our future work, we plan to overcome these limitations in three key ways. First, we plan to incorporate an online updating to the basis function dictionary, which updates basis functions for individuals over time to account for any changes in their walking patterns. This would be done by finding instances where those individuals are walking alone, and/or situations where our model has high confidence in the identity of that person's footstep response. This will allow expansion of the dictionary into varying responses due to walking speed variability, differing center of mass locations, and/or changes in walking behavior due to illness or injury. Additionally, by incorporating an online updating of the basis function dictionary, we can account for how the type of footwear an individual is wearing influences the vibration signal and obtain a system that is robust to shoe type.

In conjunction with this online updating of the dictionary, the second primary future direction involves establishing statistical models to understand "normal" walking behavior for each person and develop a distribution of their footstep responses. This will allow us to understand and track changes to that person's gait over time and assist with the online updating of their basis function dictionary.

Lastly, with regards to the overlapping steps, in the future we plan to explore relaxing the sparsity constraint to $\|\mathbf{X}\|_0 <= n_w$, $\|\mathbf{X}\|_2 <= n_w$ where n_w is the sub-window number of steps (i.e., the number of overlapping steps). In these cases, it would also be necessary to impose the constraint that only one basis function can be nonzero for each walker identified. Relaxing the sparsity constraint will provide maximum flexibility of the sparse representation algorithm in these "completely overlapping" scenarios and help reduce the overall dictionary size and computation time as the number of persons in the dictionary grows increasingly large.

6.1.3 Scalability across Different Structures. In this work, we assume that the underlying structure and sensing area are constant across both the prior footstep dictionary and the monitored walking. Prior work with footstep-induced structural vibration sensing has shown that it can accurately record human location, human gait, and differentiate between different types of impulses (i.e., balls/objects dropping, footsteps, doors closing, etc.) across the primary structural material types (wood, steel, concrete)[19, 49, 51]. However, this requires new models and/or additional prior footstep information in each new region of the structure, in different structures, or with each different structural material, which is often expensive and difficult to obtain. For our future work, we aim to reduce this training data requirement by developing an approach to transfer models across different structures/locations. Prior work in transferring structural vibration models across different structures shows promise for gait health monitoring and footstep event detection [21, 49]. We plan to leverage these prior works to develop a model transfer approach which learns the footstep-induced and structure-based signal components in one structure, transfers the footstep-induced components to a new structure, and re-learns the new structural components in an unsupervised manner. This will reduce or eliminate the need for retraining models in each new structure.

6.2 Newcomer Identification

One of the primary insights in this work is that the multiple walker vibration responses can be represented using basis functions generated by prior footstep responses from potential walkers in the area. This insight is based on the key assumption that there is prior information from the current walker in the basis function dictionary. In most cases, this is a reasonable assumption when the system is being used in an individuals' homes, an assisted living environment, or in an office environment. However, it is not always an appropriate assumption to make. For example, individuals may have guests visiting their homes, offices may have clients, new workers, delivery persons, etc., and places like assisted living and/or elder care facilities may have new residents or visitors. In these scenarios, our system would attempt to identify these "newcomers" (new individuals for whom we have no

prior information) using the current basis functions in the dictionary. As such, our system would not accurately identify those persons.

To overcome this limitation, our future work has two primary directions. First, our future work aims to directly address these "newcomer ID" situations directly by developing an approach which can determine if a person walking is one of those for whom we have prior information (and, therefore, can identify them), or, if it is a new person. Some of our preliminary work in this direction shows promise for newcomer ID. In a database of 6 walking participants, we were able to identify the 1 unknown person with approximately 91.7% accuracy (by iterating 5 of the 6 as known, 1 of the 6 as a "newcomer")[16]. Then, building off of these results, we plan to expand our use of the prior step dictionary to include an online dictionary updating component. In this way, when a person is identified as a "newcomer" we will revise our dictionary to include steps from that person. As the dictionary is being developed, our intent would be to find instances where the "newcomer" is walking independently, so that these responses can be used for the multiple walker ID approach presented in this paper.

7 CONCLUSIONS

In this paper, we present an approach for multiple person identification using footstep-induced vibration sensing. We overcome the primary research challenge of overlapping signals with unknown mixing ratios, scaling, and timing through a sparse representation where we sequentially and recursively identify multiple person's footsteps, reconstruct their portion of the signal, and remove it from the combined signal. Additionally, we reduce the influence of *structure codependency* on the separation, identification, and reconstruction of individual responses by leveraging a cosine distance-based error function in our sparse representation algorithm. Through real-world evaluations with three concurrent walkers, our system achieves an average F1 score for identifying all three walkers of 0.89, which is a 2.9× error reduction over a baseline multi-class SVM-based approach (0.68 avg. F1 score). Then, with a "hybrid" dataset involving simulated overlap of real-world walking data from 10 different participants, we show that our approach achieves an average F1 score of 0.81 across each of the 10 participants in 2-concurrent walker scenarios, which is a 2.9× reduction in error from the SVM-based baseline approach (0.44 avg. F1 score), additionally, we observe trace-level accuracies of 100%, 93.3%, and 73% for 2, 3, and 4 person combinations. These results indicate that our approach is effective at identifying and isolating the vibration responses from multiple concurrent walkers.

ACKNOWLEDGMENTS

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

REFERENCES

[1] William L. Hallauer Jr. . 2021. LTI Systems and ODEs. https://chem.libretexts.org/@go/page/7622 [Online; accessed 2021-03-31].
[2] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Frédo Durand. 2015. Capturing the human figure through a wall. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 1–13.
[3] 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.
[4] Michael Baron. 2013. *Probability and statistics for computer scientists.* CRC Press.
[5] Julian M Becker and Christian Rohlfing. 2014. Custom sized non-negative matrix factor deconvolution for sound source separation. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 2124–2128.
[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] Roberto Brunelli and Daniele Falavigna. 1995. Person identification using multiple cues. *IEEE transactions on pattern analysis and machine intelligence* 17, 10 (1995), 955–966.
[8] Yuanying Chen, Wei Dong, Yi Gao, Xue Liu, and Tao Gu. 2017. Rapid: A multimodal and device-free approach using noise estimation for robust person identification. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 1–27.

[9] Jen-Tzung Chien and Bo-Cheng Chen. 2006. A new independent component analysis for speech recognition and separation. *IEEE transactions on audio, speech, and language processing* 14, 4 (2006), 1245–1254.

[10] Pierre Comon and Christian Jutten. 2010. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press.

[11] Patrick Connor and Arun Ross. 2018. Biometric recognition by gait: A survey of modalities and features. *Computer Vision and Image Understanding* 167 (2018), 1–27.

[12] Etienne Corvee, Francois Bremond, Monique Thonnat, et al. 2010. Person re-identification using spatial covariance regions of human body parts. In *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 435–440.

[13] Roy R Craig Jr and Andrew J Kurdila. 2006. *Fundamentals of structural dynamics*. John Wiley & Sons.

[14] Razvan Cristescu, Tapani Ristaniemi, Jyrki Joutsensalo, and Juha Karhunen. 2000. Cdma delay estimation using fast ica algorithm. In *11th IEEE International Symposium on Personal Indoor and Mobile Radio Communications. PIMRC 2000. Proceedings (Cat. No. 00TH8525)*, Vol. 2. IEEE, 1117–1120.

[15] 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.

[16] Yiwen Dong, Jonathon Fagert, Pei Zhang, and Hae Young Noh. 2021. Non-parametric Bayesian Learning for Newcomer Detection using Footstep-Induced Floor Vibration. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2021)*. 404–405.

[17] 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.

[18] 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.

[19] 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.

[20] Jonathon Fagert, Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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).

[25] 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).

[26] Niall A Fox, Ralph Gross, Philip de Chazal, Jeffery F Cohn, and Richard B Reilly. 2003. Person identification using automatic integration of speech, lip, and face experts. In *Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications*. 25–32.

[27] Jerome Friedman, Trevor Hastie, Robert Tibshirani, et al. 2001. *The elements of statistical learning*. Vol. 1. Springer series in statistics New York.

[28] Davrondzhon Gafurov, Einar Snekkenes, and Patrick Bours. 2007. Gait authentication and identification using wearable accelerometer sensor. In *2007 IEEE workshop on automatic identification advanced technologies*. IEEE, 220–225.

[29] Jürgen T Geiger, Maximilian Kneißl, Björn W Schuller, and Gerhard Rigoll. 2014. Acoustic gait-based person identification using hidden Markov models. In *Proceedings of the 2014 Workshop on Mapping Personality Traits Challenge and Workshop*. 25–30.

[30] Paul Gingrich. 1992. *Introductory Statistics for the Social Sciences*. Dept. of Sociology and Social Sciences, University of Regina.

[31] Chonghui Guo, Hongfeng Jia, and Na Zhang. 2008. Time series clustering based on ICA for stock data analysis. In *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing*. IEEE, 1–4.

[32] Jiawei Han, Micheline Kamber, and Jian Pei. 2011. Data mining concepts and techniques third edition. *The Morgan Kaufmann Series in Data Management Systems* 5, 4 (2011), 83–124.

[33] Norsalina Hassan and Dzati Athiar Ramli. 2018. A comparative study of blind source separation for bioacoustics sounds based on fastica, pca and nmf. *Procedia Computer Science* 126 (2018), 363–372.

[34] Feng Hong, Xiang Wang, Yanni Yang, Yuan Zong, Yuliang Zhang, and Zhongwen Guo. 2016. WFID: Passive device-free human identification using WiFi signal. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*. 47–56.
[35] Chen-Yu Hsu, Rumen Hristov, Guang-He Lee, Mingmin Zhao, and Dina Katabi. 2019. Enabling identification and behavioral sensing in homes using radio reflections. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[36] Aapo Hyvärinen, Patrik Hoyer, and Erkki Oja. 1999. Image denoising by sparse code shrinkage. In *Intelligent Signal Processing*. Citeseer.
[37] Aapo Hyvärinen and Erkki Oja. 2000. Independent component analysis: algorithms and applications. *Neural networks* 13, 4-5 (2000), 411–430.
[38] I/O Sensor Nederland bv 2006. *SM-24 Geophone Element*. I/O Sensor Nederland bv. P/N 1004117.
[39] Rabia Jafri and Hamid R Arabnia. 2009. A survey of face recognition techniques. *journal of information processing systems* 5, 2 (2009), 41–68.
[40] Hirokazu Kameoka, Nobutaka Ono, Kunio Kashino, and Shigeki Sagayama. 2009. Complex NMF: A new sparse representation for acoustic signals. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 3437–3440.
[41] Gaëtan Kerschen, Fabien Poncelet, and J-C Golinval. 2007. Physical interpretation of independent component analysis in structural dynamics. *Mechanical Systems and Signal Processing* 21, 4 (2007), 1561–1575.
[42] 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.
[43] Nacer Khalil, Driss Benhaddou, Omprakash Gnawali, and Jaspal Subhlok. 2016. Nonintrusive occupant identification by sensing body shape and movement. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. 1–10.
[44] 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.
[45] Hongyin Lau and Kaiyu Tong. 2008. The reliability of using accelerometer and gyroscope for gait event identification on persons with dropped foot. *Gait & posture* 27, 2 (2008), 248–257.
[46] Te-Won Lee and Michael S Lewicki. 2002. Unsupervised image classification, segmentation, and enhancement using ICA mixture models. *IEEE Transactions on Image Processing* 11, 3 (2002), 270–279.
[47] 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.
[48] Jani Mantyjarvi, Mikko Lindholm, Elena Vildjiounaite, S-M Makela, and HA Ailisto. 2005. Identifying users of portable devices from gait pattern with accelerometers. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, Vol. 2. IEEE, Philadelphia, PA, USA, ii–973.
[49] 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.
[50] 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
[51] 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.
[52] 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.
[53] PY Mok, KP Lam, and HS Ng. 2004. An ICA design of intraday stock prediction models with automatic variable selection. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, Vol. 3. IEEE, 2135–2140.
[54] Morten Morup, Kristoffer H Madsen, and Lars K Hansen. 2007. Shifted non-negative matrix factorization. In *2007 IEEE Workshop on Machine Learning for Signal Processing*. IEEE, 139–144.
[55] Mehdi Moussaïd, Niriaska Perozo, Simon Garnier, Dirk Helbing, and Guy Theraulaz. 2010. The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PloS one* 5, 4 (2010), e10047.
[56] Brent C Munsell, Andrew Temlyakov, Chengzheng Qu, and Song Wang. 2012. Person identification using full-body motion and anthropometric biometrics from kinect videos. In *European Conference on Computer Vision*. Springer, 91–100.
[57] Le T Nguyen, Yu Seung Kim, Patrick Tague, and Joy Zhang. 2014. IdentityLink: user-device linking through visual and RF-signal cues. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, Seattle, WA, USA, 529–539.
[58] 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.
[59] 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.

[60] Shijia Pan, An Chen, and Pei Zhang. 2013. Securitas: user identification through rgb-nir camera pair on mobile devices. In *Proceedings of the Third ACM workshop on Security and privacy in smartphones & mobile devices*. ACM, ACM, Berlin, Germany, 99–104.
[61] 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.
[62] 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.
[63] Shijia Pan, Mostafa Mirshekari, Jonathon Fagert, Carlos Ruiz, Hae Young Noh, and Pei Zhang. 2019. Area occupancy counting through sparse structural vibration sensing. *IEEE Pervasive Computing* 18, 1 (2019), 28–37.
[64] 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.
[65] 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.
[66] Shijia Pan, Susu Xu, Mostafa Mirshekari, Pei Zhang, and Hae Young Noh. 2017. Collaboratively adaptive vibration sensing system for high-fidelity monitoring of structural responses induced by pedestrians. *Frontiers in Built Environment* 3 (2017), 28.
[67] 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.
[68] 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.
[69] 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.
[70] John G. Proakis and Dimitris G. Manolakis. 1996. *Digital Signal Processing (3rd Ed.): Principles, Algorithms, and Applications*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
[71] Tapani Ristaniemi and Jyrki Joutsensalo. 2002. Advanced ICA-based receivers for block fading DS-CDMA channels. *Signal Processing* 82, 3 (2002), 417–431.
[72] Mikkel N Schmidt and Morten Mørup. 2006. Nonnegative matrix factor 2-D deconvolution for blind single channel source separation. In *International Conference on Independent Component Analysis and Signal Separation*. Springer, 700–707.
[73] Mikkel N Schmidt and Rasmus K Olsson. 2006. Single-channel speech separation using sparse non-negative matrix factorization. In *Ninth International Conference on Spoken Language Processing*.
[74] Claude E Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal* 27, 3 (1948), 379–423.
[75] 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.
[76] Abdulhamit Subasi and M Ismail Gursoy. 2010. EEG signal classification using PCA, ICA, LDA and support vector machines. *Expert systems with applications* 37, 12 (2010), 8659–8666.
[77] Budi Sugandi, Hyoungseop Kim, Joo Kooi Tan, and Seiji Ishikawa. 2009. Real time tracking and identification of moving persons by using a camera in outdoor environment. *Int. J. Innov. Comput. Inf. Control* 5, 5 (2009), 1179–1188.
[78] 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.
[79] Thiago Teixeira, Deokwoo Jung, and Andreas Savvides. 2010. Tasking networked cctv cameras and mobile phones to identify and localize multiple people. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*. ACM, Copenhagen, Denmark, 213–222.
[80] W.T. Thomson and M.D. Dahleh. 1998. *Theory of Vibration with Applications*. Prentice Hall.
[81] Hsin-Chun Tsai, Wen-Chi Wang, Jia-Ching Wang, and Jhing-Fa Wang. 2009. Long distance person identification using height measurement and face recognition. In *TENCON 2009-2009 IEEE Region 10 Conference*. IEEE, 1–4.
[82] CJ van Rijsbergen. 1979. Information Retrieval, 2nd edButterworths.
[83] Filipa Campos Viola, Stefan Debener, Jeremy Thorne, and Till R Schneider. 2010. Using ICA for the analysis of multi-channel EEG data. *Simultaneous EEG and fMRI: Recording, Analysis, and Application: Recording, Analysis, and Application* (2010), 121–133.
[84] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 363–373.
[85] Yunze Zeng, Parth H Pathak, and Prasant Mohapatra. 2016. WiWho: wifi-based person identification in smart spaces. In *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*. IEEE Press, Vienna, Austria, 4.

[86] Ming Zhang and Anxue Zhang. 2019. The Superposition Principle of Linear Time-Invariant Systems [Lecture Notes]. *IEEE Signal Processing Magazine* 36, 6 (2019), 153–156.

[87] Yue Zhang, Lin Zhang, Hae Young Noh, Pei Zhang, and Shijia Pan. 2019. A signal quality assessment metrics for vibration-based human sensing data acquisition. In *Proceedings of the 2nd Workshop on Data Acquisition To Analysis*. 29–33.

[88] Zhaonian Zhang and Andreas G Andreou. 2008. Human identification experiments using acoustic micro-Doppler signatures. In *2008 Argentine School of Micro-Nanoelectronics, Technology and Applications*. IEEE, 81–86.

[89] Zheng Zhang, Yong Xu, Jian Yang, Xuelong Li, and David Zhang. 2015. A survey of sparse representation: algorithms and applications. *IEEE access* 3 (2015), 490–530.

[90] Andrei Zinovyev, Ulykbek Kairov, Tatyana Karpenyuk, and Erlan Ramanculov. 2013. Blind source separation methods for deconvolution of complex signals in cancer biology. *Biochemical and biophysical research communications* 430, 3 (2013), 1182–1187.