Indoor Human Information Acquisition from Physical Vibrations

Submitted in partial fulfillment for the requirements for
the degreee of
Doctor of Philosophy
in
Electrical & Computer Engineering

Shijia Pan

B.E., Computer Science, University of Science and Technology of China

Carnegie Mellon University
Pittsburgh, PA

May 2018

To my family.

Abstract

With the growth of networked smart devices in indoor environments, human information acquisition becomes essential for these devices to make the environment smart and people's lives more convenient. These networked systems, which are often referred to as Cyber-Physical Systems (CPS), learn and make decisions collaboratively based on data input. The data could come from sensors that perceive various signals in the physical world, human input, etc. In this thesis, I will focus on information acquisition based on data from sensing the physical world.

The major challenges to accurately interpreting the information these systems perceive result from the complexity of the physical world. An extreme solution to this problem is to have a large number of sensors or sensing configurations that collect a large amount of data. Ideally, we could then have labeled data for each sensing condition and possible scenario in order to accurately model the world. However, in the real world, such solutions could be difficult if not impossible to achieve due to constraints on the hardware, computational power, and (labeled) dataset.

This thesis targets this problem and sets the goal of obtaining accurate indoor human information through limited system configurations and limited labeled data. A new concept of utilizing structures as sensors is presented as the foundation of the system. The intuition is that people induce ambient structures to vibrate all the time, and their activities and information can be inferred from this vibration. To achieve that with the aforementioned constraints, an understanding of the physical world (that has been studied for centuries in multiple disciplines) is used to assist the sensing and learning process for more accurate information acquisition from sensor data.

Acknowledgments

It has been a long journey, and there are many people that I need to thank.

First, I would like to thank my advisors Pei Zhang and Hae Young Noh. Thank you for tolerating and helping me with the writing skills that were a struggle for the first three years. I have no idea what I put you through but I'm glad you survived. You are my inspiration to stay in academia because I want to guide more lost souls like me 6 years ago as you did.

Thanks to all the other committee members. Thank you Jie Liu for insightful discussions every time we met at conferences. Thank you for reading my thesis and writing my letter. Thank you Anthony Rowe for lending your lab when I visited Pittsburgh campus. The unique structural characteristics of your lab inspired my first ever first author paper. Thank you all for your support during my PhD program.

Thank you John Paul Shen, Yanyong Zhang, Fred Jiang, Wendy Fong, An Chen, Kent Lyons, Jie Gao, Rong Zheng and Mooi Choo Chuah for providing me career-wise advices. Thank you Aveek Purohit and Zheng Sun for showing me how research is done when I first joined the group. You are my first co-authors and it was fun doing experiments with you. Thank you Frank Mokaya for all the latex templates you sent to me and all the wise words you said to me. You are the model senior student that I tried to match up to.

Thank you Mostafa Mirshekari and Jonathon Fagert for being wonderful collaborators on vibration sensing. I wish I can talk about structural properties with vocabularies like yours. Thank you Carlos Ruiz for showing me that it is ok to have weekends sometimes. Thank you Amelie Bonde, Helena Bonde, and Adeola Bannis for your help on my English writing and for your moral support when I went through a few life-changing events.

Thank you Jun Han, Xinlei Chen, Kaifei Chen, Susu Xu and Tong Yu, it was always a great pleasure writing papers with you. Thank you

Heng-Tze Cheng, Le Nguyen, Rahul Rajan, Senaka Buthpitiya, David Cohen, Xiao Wang, Yuan Tian, William Chan, Bing Liu, Guan-Lin Chao, Ming Zeng for all the help and fun at school.

I'd like to thank my parents for their love, support and understanding. Thank you cousin Zhuying for showing me strength and independence in life. Thank you Chihiro Suga for five Thanksgiving we spent together and countless adventures we went through. Thank you for attending my thesis defense. Thank you Wenhui Hu for being a wonderful friend, your friendship means a lot to my survival of the PhD program. Thank you Sareh Yousefzadeh, you are my wonder woman. Hanging out with you in Pittsburgh makes the weather tolerable. Thank you Ji Jia for always have faith in me. You have been calling me Dr. Pan for six years, and now I finally become one. Thank you Lisa Zhou and Margaret Lee for your kindness, caring and support.

To all that have crossed my path, thank you for the journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Ubiquitous human information acquisition in indoor environments enables various smart building applications. For example, occupant identification detection has building security applications and enables personalized services in smart environments. Indoor human flow count is useful for market research, such as advertisement location efficiency evaluation. In addition, general activity level estimation over the building provides space usage information that can be used for optimization of the structure-based space design. Pedestrian location information allows indoor tracking based applications (e.g., an advanced hospital/nursing home can track patient activity range, or a smart supermarket can track customer behavior trends). Generally speaking, this spatio-temporal indoor human information unlocks various high-level applications, therefore, it is important to obtain it accurately.

Various sensing systems are designed for indoor human information acquisition. Traditional sensing methods often fall into two categories: mobile-based sensing methods and infrastructure-based methods. Mobile-based sensing methods often rely on the mobile devices in people's lives (e.g., smartphone, tablet, laptop) to obtain sensing information. These systems assume that 1) each pedestrian carries a networked personal device all the time; and 2) their personal devices have a similar sensing ability to finish certain sensing tasks (i.e., collecting pedestrian spatio-temporal information). Advantages for such systems include 1) no extra infrastructure is required, 2) the identification information is actively provided by the device itself, and 3) mobile sensing allows crowd-sourcing, which is a powerful dynamic information acquisition method. On the other hand, the assumptions

listed above may not hold: in many network systems we cannot assume that every pedestrian has a capable smart device. For example, an office or shopping mall cannot assume all occupants carry such devices all the time. Another possible concern is the power consumption on mobile devices may shorten the device's usable time, and therefore limit sensing duration.

Infrastructure-based methods are systems that utilize installed devices designed for specific sensing purposes. The advantages of these systems are 1) no requirement on pedestrian interaction (e.g., install a specific mobile application and keep it on) and 2) this infrastructure is often designed to be powered from a power outlet, and therefore have fewer concerns on the power limitations. The drawback is also straightforward: they require certain well-designed sensing devices to be installed in the environment to conduct sensing tasks. The pros and cons for these two categories of systems compensate for each other, and the selection for the design for different types of systems are mainly determined by the sensing purpose and algorithm characteristics.

1.1 Problem Statement

This thesis focuses on the overall research problem of **acquiring indoor human information through non-intrusive ambient structural vibration sensing**. The sensing approach I took is referred as 'structures as sensors', where the system infers human movements via sensing of the structural vibrations induced by these movements. For example, when pedestrians walk on the floor, their feet striking the floor induces the structure to vibrate. Such vibration travels across the floor plane and can be picked up by vibration sensors placed on the floor. When multiple sensors pick up the footsteps-induced vibration sequence, the pedestrian's movement information can be inferred from it. In addition, different pedestrians walk with various habits/positions. Therefore, their footsteps induce vibrations differently, which can be used for identification or characterization purposes.

The system is non-intrusive in three ways 1) the structural vibration sensing approach does not require any human-system interaction, therefore it's non-intrusive in terms of the sensing target; 2) compared to vision- or acoustic-based methods, it is less intrusive in terms of user privacy; 3) the system is designed to be "plug-and-play"

– the sensing node installation is as easy as putting the sensing node on the surface of the structure in an arbitrary location, therefore it's non-intrusive also in terms of the sensing environment.

1.2 Research Questions

The 'structures as sensors' system consists of three components: sensing (signal acquisition), signal processing (characterization), and information learning. Each component faces unique challenges compared to the other sensing modalities. In this thesis, I discuss the details of these challenges and solutions in Chapter 4, 5, and 6 respectively.

1.2.1 'Good' Signal Acquisition

In general, for information acquisition, signal resolution is one important metric to measure the quality of the signal, which also indicates the amount of information contained in the signal. Therefore, obtaining high-resolution signals that represent the information accurately (i.e., low distortion) is important for the information acquisition. However, the hardware limitations, including sensor's dynamic sensing range, ADC resolution, number of sensors and computing power, often make it difficult to obtain 'good' signal. Therefore, the research questions defined in this aspect include the following.

- How is it possible to obtain 'high fidelity' human-induced structural vibration signals using sensors with limited dynamic range and ADC modules with limited resolution?

- How is it possible to extract human-induced vibration signals from the raw vibration signals with time-variant noise?

1.2.2 Accurate Signal Characterization

Once the target signal is acquired, it is analyzed with various signal processing methods to obtain features that describe the characteristics of the signal. The processing methods include but are not limited to normalization, aligning, extracting/detecting features, filtering, time/frequency domain analysis, pairwise Time Difference of

Arrival (TDoA) estimation. The characterized signals or features are then used as input of the learning or analytic model to further extract human information.

Therefore, accurately characterizing the signal will allow accurate learning in the learning module. The research questions in this module are then defined as follows.

- How is it possible to characterize signals of structural vibrations based on the wave properties?

- What features are useful for specific applications we investigate (i.e., identification, pedestrian traffic estimation, localization)?

- How can the system adapt to different structural factors (e.g., different floor types, different structure sizes)?

1.2.3 Machine Learning for Cyber Physical Systems

Once the signals are characterized, the final target information needs to be learned from the processed signals, and the learning problem can appear in various forms: supervised, unsupervised, numerical analysis, etc. For most of the data-driven approaches, the models are built only relying on the accessible data. Therefore, the accuracy of the models is determined by the completeness of the labeled data.

However, for real-world physical systems, there are many factors that affect the sensing signals. Therefore, it is difficult to obtain 'enough' data, especially the labeled data to cover all the possible factors. Hence, the research questions in this module are defined as follows.

- How to learn pedestrian information from the characteristics extracted from their footstep induced structural vibrations?

- How to learn accurate information with limited initial label data?

- How well does the system perform when different human factors are involved (e.g., different shoe types, different walking behaviors)?

1.3 Contributions

To solve the research questions listed above, this thesis presents a methodology that uses physical insights to guide the sensing and learning process to achieve more accurate human information acquisition.

1.3.1 Improve Signal Quality by Collaboratively Adapting Hardware Configuration

Human-induced structural vibration varies in amplitude and duration, so obtaining a high-resolution signal for different human interactions is challenging. Signals such as footstep induced floor vibration can be of very small amplitude and can vary rapidly when a person walks by the sensor. Therefore, without proper hardware setting the signal obtained will be either of low resolution or distorted (e.g., clipping). This problem is important because acquiring a high fidelity vibration signal is crucial for information extraction.

I design the data acquisition system to collaboratively adapt the hardware based on the software computation. The intuition is that when people are closer to a sensor, their footstep signal has a higher magnitude. I propose a hardware adjustment mechanism based on the signal magnitude prediction calculated from detected events and a heuristic model of human movement. The goal of the mechanism is to achieve high sensing signal resolution for the detected vibration signals of individual footsteps.

To evaluate this collaboratively adaptive hardware configuration for footstep induced vibration acquisition, both simulation and real-world experiments are conducted. The system shows an up to $2\times$ signal resolution improvement measured by the sufficient resolution rate, and $2\times$ and $3\times$ improvements when applied to two localization applications.

1.3.2 Improve Signal Characterization based on Physics Properties

Different human-structure interactions cause different types of vibration waves. Impact-induced vibration signals are common in everyday life, such as footsteps, knocking on a door, typing, tapping, etc. However, there are other types of excitations such as dragging an object on the floor, swiping a finger or pen through a countertop, etc. These actions perform a stick-slip form of motion along the surface and generate friction-induced vibration. Due to different causes of the vibration, the wave properties are different, which directly affects the propagation characteristics (e.g., decay, dispersion).

Specifically, in the application of tracking, I study the properties of these waves

and design corresponding tracking algorithms to handle them. For impulse-like Rayleigh-Lamb waves, a wavelet filter is applied to reduce the dispersion effect and the first peak of the signal is extracted to calculate Time Difference of Arrival (TDoA) for localization; while for friction induced body waves, the TDoA is calculated through cross-correlation of the windowed signal segments. Compared to the baseline (no signal characterization), the physics property guided signal characterization allows up to $6\times$ improvements on impulse signal localization and up to $3\times$ improvements on slip-pulse signal tracking length estimation.

1.3.3 Improve Learning Accuracy with Limited Initial Labeled Data

Changes in human behavior lead to widely varying distributions in the corresponding sensing data. The traditional supervised learning methods rely on labeled data to build models, however, the complexity in the real world scenarios makes it hard if not impossible to achieve high prediction accuracy by collecting a large-scale labeled dataset. Transductive transfer learning methods have been explored to handle limited data distribution changes by taking both labeled and unlabeled data into account.

I build on prior work in transductive transfer learning by presenting an iterative learning method. Each learning iteration is guided by physical measurements to achieve high learning accuracy with limited labeled data. I use the intuition that in the complicated physical world, data distribution changes are often caused by changes in the sensing target or condition, which can be measured. I propose a physical measurement guided iterative learning algorithm, in which the iteration (the order of learning) is controlled by these physical measurements to ensure high prediction accuracy. The algorithm is applied to pedestrian identification with my ambient vibration sensing system, FootprintID, which shows a $1.5\times$ improvement on identification accuracy on the overall data distribution.

1.4 Thesis Organization

This thesis is organized as follows. Chapter 2 first discuss the background and the challenges/research gaps in the field. Chapter 3 introduce the overall methodology of this thesis: use physical insights to guide the sensing and learning process. Then

Chapter 4, 5, and 6 introduce how this methodology is applied to the signal acquisition, signal characterization, and learning respectively. Finally, I summarize this thesis in Chapter 7.

Chapter 2

Background

The history of the modern computation technologies falls back to decades ago, where the invention of the computer introduces the automation as well as the digital world into people's life. In this new era, the digital world, or as some people refer to as the cyber world, plays a significant role in people's life. The cyber technology has come so far and yet the challenges remain...

This chapter of the thesis first introduces the cyber-physical system's role and applications (Section 2.1). Then the growth of the sensing systems will be discussed (Section 2.2). Next, the state-of-the-art sensing modalities that target indoor human sensing will be compared (Section 2.3). Among all these sensing modalities, I will then go into depth on vibration sensing and its history (Section 2.4). Next, the limitation of pure data-driven learning approaches applied on cyber physical systems, which is the requirement of large amount of labeled data, is discussed in Section 2.5. In addition, I will discuss what is missing based on the background introduced and how my work fills the gaps in each section.

2.1   Cyber Systems, the Physical World, and Human

Cyber-physical systems [68, 122] are defined as the integration of computation and physical processes. The concept of cyber-physical systems, from a system perspective, can be extended concepts of sensor networks, where the understanding of the physical world is taking into consideration of the computation. Compared

to the prior concepts of sensor networks [5], cyber-physical systems explore a more tightly connected relation between human and the physical world.

Figure 2.1 shows the relation between these three aspects: human, physical world (environment), and cyber systems. The traditional sensing systems directly measure the physical world or people through a particular sensor [11] by converting particular physical attributes (e.g., temperature, sound, motion) into voltage changes that can be mapped to changes of these attributes. People can control the cyber systems through the provided interfaces [131], which, in turn, actuates the physical world. The physical world (i.e., structures as the focus of this thesis) perceives the appearance of human and responses to the interaction from people [14].

This thesis, on the other hand, emphasizes on the understanding of human interacting with the physical world, as well as the perception and interpretation of this signal (red arrows in Figure 2.1). The complicated interaction between the physical world and the human were studied over centuries (e.g., human-induced wave properties [44, 147], human gait [111]), however, not from a cyber system perspective.
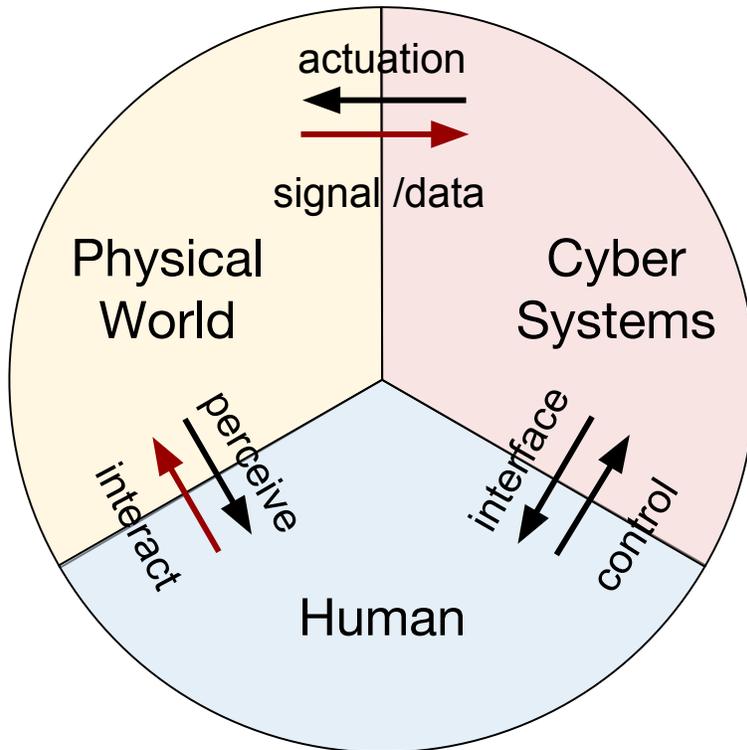
Figure 2.1: Cyber-Physical System (CPS) overview.

With this understanding, cyber systems can obtain human information both directly from human and indirectly from sensing the physical world (through the knowledge of this interaction). Therefore, in this thesis, I will focus on integrating prior studies on human and the physical world into cyber systems to acquire accurate human information indirectly.

2.2   Sensing Systems and Their History

Sensor networks have been studied for decades [5, 6]. The growth of the systems can always be traced back to the evolution of people's needs to monitor the world.

In the early years of sensing systems, sensor networks are deployed at specific locations to obtain designated types of information, mostly environmental [79]. For example, microphones are deployed to collect continuous data from an erupting volcano [153]. The challenges for these environmental sensing systems lie in the limitations of the computational power, communication, storage, etc. The research foci are on the resilience of the system.

Then, the challenge came when the sensing systems' target shifts from the environment to moving objects, bringing more challenges due to mobility [166]. For example, GPS equipped collars are installed on zebras to monitor the herd's social behavior [166]. The mobility of individual sensing nodes makes the problem divert into a new field mobile sensing, where the constraints on energy consumption, storage, as well as communication alter from the traditional sensor networks. These mobile sensor networks tend to collaboratively conduct sensing or communication tasks to enhance the system throughput [60]. Later, when the smartphone become ubiquitous in people's everyday life, the on-body mobile sensing raises more challenges in terms of the heterogeneity of the platforms [137], which will be explained in the next paragraph with the topic of human-centric sensing.

The next generation of the systems is more human-centric – either deployed on a targeted person (e.g., wearables [108]) or in the environment (e.g., vision based human activity recognition [114]). With the growth of this generation of systems, their intrusiveness is more and more to people's awareness in various sensing scenarios. For example, it is difficult to convince the elderly to carry certain devices all day and to remember to charge them at night. As for vision- or acoustic-based methods,

people tend to have privacy concerns.

At the time this thesis is written, the new generation of sensing systems has been integrated with new components – the understanding of the physical world – to achieve more ubiquitous and non-intrusive sensing. The cyber-physical systems often contain sensing components, which is like traditional sensor networks. The concept changing components is the rule played by the physical world in the computation, i.e., instead of from sensing signal to human information, the system infers human information from the physical world change represented by the sensing signals. Various applications have been explored by different sensing modalities, including radio-frequency- [155, 156], powerline- [110], vibration-based sensing [97, 104]. In the coming section, I will further compare these sensing modalities.

2.3 Indoor Human Sensing Systems

Many different sensing approaches have been proposed through the last decades [143], targeting information extraction about human in indoor environments, including presence, count, location, identity, and activity. They can be categorized by sensing methods as follows: vision-, RF-, mobile-, acoustic-, and load-based methods.

Vision-based methods track human and their interaction by recognizing their targeting tracking point (e.g., head, foot, hand, etc.) and localizing the point in the 3D perspective of the camera view [35, 61, 78, 81, 123]. The vision-based methods usually provide accurate information extraction (including tracking, interaction, and identification), if the light-of-sight between the target point and the camera is satisfied. In addition, vision-based methods often cause privacy concerns.

RF-based methods utilize the phenomena where human bodies between RF transmitters and receivers alter the signal characteristic to extract human information such as location or number of people [1, 3, 118, 119, 150, 161, 164]. Due to high sensitivity to environment setups, these methods either require dense sensor deployment or yield low resolution in terms of tracking accuracy.

Mobile devices, either smartphones or application-specific devices [23, 75, 80, 120, 169] that are equipped with inertial sensors can be used to extract human information, including interactions [23], locations [75, 169], identities based on gaits [80], and activities [120]. The inertial sensors embedded in the device can capture the motion

Techniques	Vision	RF	Mobile	Acoustic	Load	Vibration
Tracking Err. (m)	≤ 1 [35]	0.25 [119]	≤ 2 [75]	0.17 [140]	tile size [96]	0.21 [86]
Interaction Err. (mm)	10 [136]	NA	37 [150]	NA	.002[117]	30[103]
ID Accy.(%)	95[61]	80 [164]	85 [80]	90[168]	93 [96]	96 [104]
Detection Accy.(%)	98 [78]	100 [161]	87.4 [120]	97 [142]	100 [7]	99.6 [97]
Deploy Density	Low	High	NA	Low	Very High	Low
Carrying Device	No	No	Yes	No	No	No
Privacy Cond.	Poor	Good	Poor	Poor	Good	Good

Table 2.1: Performance summary of different sensing methods.

of the device usage, which are then used to infer the gesture or gait information. The Wi-Fi module on the phone can be used to estimate indoor locations, and the inertial sensors that detect significant movement changes (e.g., turns) can be used together with noisy location estimations to improve tracking quality. However, these methods rely on the assumption that the targets carry devices all the time, which as aforementioned, may not be realistic for scenarios such as elderly or kids monitoring.

In addition, acoustic-based methods [21, 45, 140, 142, 152, 168] rely on sound waves propagation to identify or localize sources. They often require the user to be equipped with special wearable devices or they are limited by sensing conditions (limited resolution and range). Load-based methods [7, 8, 96] achieve high accuracy on interaction points via high-density sensor deployment.

The performance, as well as pros and cons of these methods are summarized in Table 2.1. For the methods that do not provide enough resolution for direct interaction tracking, we put NA in the table. Compared to these prior methods, we focus on structural vibrations sensing which allows sparse sensing and high-resolution information acquisition, which will be discussed in details in Section 2.4.

2.4 Structural Vibration Sensing

Work has been done on structural vibration monitoring through sensor networks for various applications, including structural health monitoring [64, 91, 92, 93, 94, 160]. The insight on these works provides a good understanding of the wave propagation and characteristics of the human-induced structural vibration. Recent research also starts to utilize such signals for indoor human monitoring [67, 70, 85, 87, 97, 116]. In the rest of this section, I will first introduce the related work to the vibration sensors as well as the vibration signal acquisition (Section 2.4.1). Then I will present work on the human information inference through vibration sensing (Section 2.4.2).

2.4.1 Vibration Sensing: Signals and Sensors

Vibration sensing is a more specific term compared to acoustic sensing. Acoustics is the study of all types of mechanical waves, including vibration, sound, ultrasound, and infra-sound [145]. Sound signals are vibrations that propagate through air and of high frequency. Although acoustic signals have been very well studied in terms of its production, transmission, and effects, including biological and psychological effects [145], many challenges are not well understood for human sensing purpose. For example, the dispersion effects alter the waveform significantly when the signal propagates through solids.

From the sensing perspective, the mechanism that allow the signal to be converted to voltage representation is different over various sensors. Various conversion methods have been explored, including MEMS and piezo. MEMS stands for micro electro mechanical system. It applies to any sensor manufactured using microelectronic fabrication techniques. The MEMS sensors are often implemented as Variable Capacitive (MEMS VC) to obtain the vibration in terms of acceleration, which have the advantages including high resolution, long-term stability, sensitivity stability, and low thermal drift [36]. Piezo sensors, on the other hand, measures the vibration in terms of a change in acceleration leads either to a voltage difference (Piezoelectric effect) or a change in resistance (Piezoresistive effect) [36]. There are multiple conversion methods for piezo sensors, including dynamic, condenser, ribbon and crystal [82]. Piezoceramic sensors are based on the vibration of a crystal [82]. Piezo dynamic sensors have a 'moving coil' inside of them that captures the vibrations [84].

14

Piezo condenser sensors are based on a moving plate and a charged fixed plate and the capacitance that is generated between them [84]. Piezo ribbon sensors are based on a thin piece of aluminum or other types of material – the ribbon. The metal strip or ribbon is designed between two magnets in an electrical field that is disrupted as the ribbon moves back and forth within the magnetic field [84].

2.4.2 Vibration Sensing: Human Information Inference

Structural vibration sensing provides a sparse and non-intrusive solution for obtaining human information. The foundation of this type of sensing method lies in the excitation induced by human activities that interact with the structures. The sensing challenges of structural vibration monitoring for indirect human sensing include 1) noisy ambient noise [70], 2) multi-path of the vibration signal in an indoor environment [116], and 3) the distortion and low resolution of the sensing signal [101].

The information people extract mainly falls into the following categories: presence, count, identity, tracking, activity, status. When there is human-induced structural vibration detected, the presence of the people is detected. Work has been done on identifying the floor vibrations that are induced by human footstep from other impulses to achieve the presence detection [67, 127]. When multiple people interact with the physical surroundings, the vibration signal induced by different people will be mixed. Therefore, when the mixed signal is detected by multiple sensors at a different location, the features can be extracted across these sensors and then be used to estimate the number of people in the sensing environment [102, 138]. Since human gait is a unique biometric that can be applied to identify individuals, the footstep induced vibration can be used to infer such information [29, 39, 104]. The tracking of the excitation sources enables indoor human tracking through different approaches that enable different tracking resolutions [85, 88, 97]. As for activity, such as interaction with the structure, has also been explored through a different approach, including machine learning approaches [41, 48, 49] and localization approaches [103, 109, 113, 154] Recently, work has been done on the health monitoring through structural vibration sensing that targets at micro signals induced by human body such as heart rate [56].

I believe, the reason vibration sensing based methods have not been widely

explored over the last decades together with other sensing approaches is that the major work has been done on the vibration sensing systems are mainly pure data-driven approaches, which makes the labeling of human information impractical if not impossible. The pure data-driven sensing systems and their pros and cons will be discussed in the next section, together with the discussion on how the physical-guided sensing allows a more robust sensing and learning process.

2.5   Data-Driven Systems

Data-driven science (a.k.a. data science) utilizes scientific methods, processes, algorithms and systems to extract knowledge from data in various forms [30]. The opposite concept of the data-driven systems is the physics-driven systems. The former only relies on the obtained sensing data to learn the explicit relation between the target information and the sensing signal, which means the features used for learning may or may not has actual causality between the learned information and the signal. While the later is built upon the physics knowledge.

Most of the cyber-physical systems have been using pure data-driven approaches to compute the target information, such as machine learning, classification, cluster analysis, uncertainty quantification, computational science, data mining [50]. The pros of this approach are that it is flexible to the real world changes since the effect of the environment to the signal has been fully considered. However, the downside of the data-driven models, especially the pure data-driven model, is that they usually require a large amount of labeled data to train a model at high precision. On the other hand, the physics-driven models often require certain input based on the model people summarize, which may not be thorough enough to take various sensing condition changes into consideration.

The comparison and decision on which approach to use for particular systems have been discussed [10, 170]. The method introduced in this thesis – physical insight guided sensing – takes into consideration both the data-driven and the physics-driven models to achieve highly accurate learning with limited sensing data.

Chapter 3

Methodology

In this chapter, the methodology used to achieve works in this thesis will be introduced. The overall methodology applied can be summarized as physical guided sensing, which means utilizing measurable physical information to assist information acquisition. When applied to the sensing system, especially the 'structures as sensors' presented in this thesis, this methodology needs to be specified when applied at different phases of the information acquisition process.

3.1 'Structures as Sensors' Framework

A typical sensing system usually consists of three major modules: signal acquisition, signal processing (characterization), and information learning. The 'structures as sensors' framework enhanced the sensibility of the traditional sensing system by involving structures – part of the physical world – into the signal acquisition module as shown in Figure 3.1.

Many human activities of everyday life excite the structure around people, such as footstep, knock, finger touching, etc. These excitations induce the interacted structure, such as floor, table, door, etc., to vibrate. This vibration travels through the structure or associated structures and is obtained by the vibration sensors. The system acquires this vibration signal, then conducts signal characterization and information learning on the signal to infer the corresponding human information indirectly.
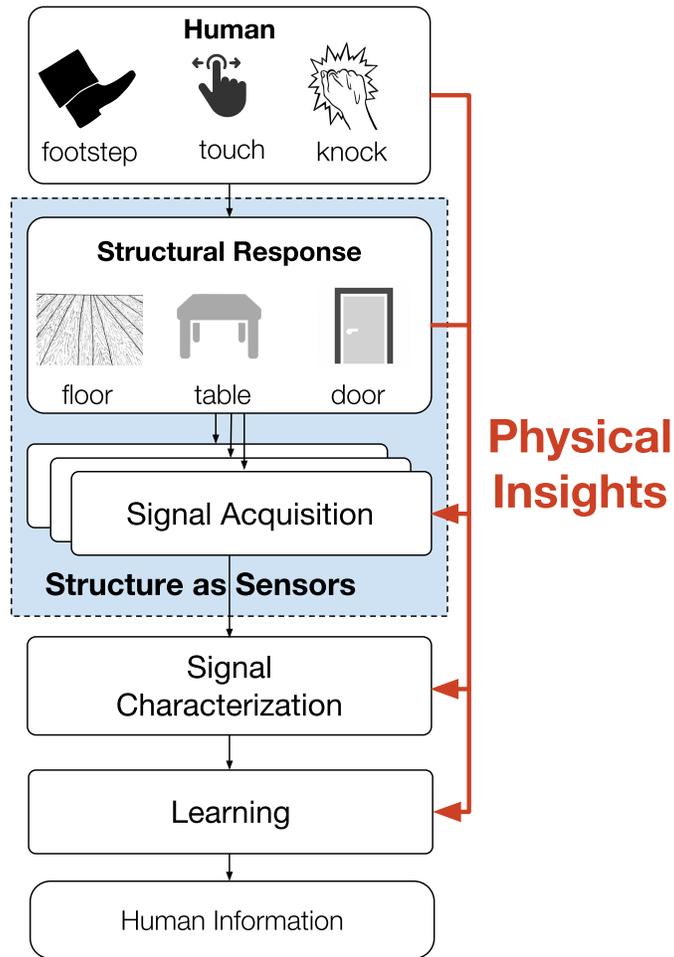
Figure 3.1: 'Structure as Sensors' framework. Structural and human physical insights or knowledge (e.g., wave attenuation, wave dispersion, gait analysis) are used to assist the process of signal acquisition, signal characterization and learning.

3.2 Physical Measurements Guided Signal Acquisition

Sensors in various sensing systems serve the purpose of obtaining physical measurements to allow further inference of other physical conditions. For example, geophone is a type of sensor that contains a coil and a metal core. When there is displacement of the surface where the sensor is placed on, the metal core moves slightly with the monitored surface, and this movement of the metal core causes the magnetic field

to change through the coil, hence changes the voltage between the two ends of the coil. This voltage change relates to the velocity of the target surface. In this way, the physical measurements of the surface vibration are achieved.

The voltage change between the ends of the coil is an analog signal, which cannot be directly analyzed by all types of algorithms. To achieve that, the signal needs to be converted from analog to digital and this digital signal can then be quantified for analysis. Therefore, from the signal being generated by a change in the physical world, to the signal is detected and analyzed, the chain of processing determines the quality of the signal, hence the accuracy of the information extracted from it. Because of this, the hardware configuration is critical for the physical measurement.

The method presented in this thesis to enhance this signal acquisition process is to design hardware configuration based on physics models and to adapt hardware configurations based on the physical measurements of the signal. To be more specific on human sensing through ambient structural vibration, the system is designed based on the signal resolution model derived from the surface wave attenuation model [97, 105]. In addition, the system interprets the current and historical physical measurements to predict the optimal hardware configuration [105]. This physical measurement is high-level information such as human movement and structural factors and used to assist signal/event predictions, which will be discussed in details in Chapter 4.

3.3   Physical Properties Guided Signal Characterization

Signal characterization identifies and extracts the designated form of the signal from the raw signal (e.g., decomposition, filtering) and is an important step before the information learning. Due to the complexity of the physical world, various condition changes may alter the signal obtained. This is extremely severe when the propagation medium of the signal is heterogeneous, for example, the structure where the vibration propagates can vary in material, density, stiffness. These propagation medium variations often lead to the variation of the excitation response and wave propagation. For example, the stiffness of the surface affects the vibration response of an impact on the surface, while the density of the material affects the attenuation rate of the

Lamb-Rayleigh wave, which is the main wave components of the impact-induced impulse-like waves.

The method presented in this thesis in terms of enhancing the signal characterization is to understand the physics properties of the wave, to link it to the measurable signal characteristics, and to utilize these properties to design specific algorithm for different types of signals. To be more specific on human sensing through structural vibration sensing, the system determines the types of waves based on the detected signal and then select the algorithm designed based on the corresponding signal properties to conduct further information learning [100, 103]. The system selects and extracts signal features for learning based on the wave superposition properties [99]. The contribution of this method is to refine types of signals obtained by the same type of sensors based on their physical properties (e.g., propagation, superposition) to enhance the information acquisition accuracy, which will be further discussed in Chapter 5.

3.4   Physical Attributes Guided Iterative Learning

To convert signals, or features extracted from the signals, to designated human information requires learning algorithms. Supervised learning is a commonly used way to obtain a specified type of information. It requires labeled data as a learning reference. However, for many real-world applications, obtaining labels is labor intensive and expensive. To tackle this problem, various semi-supervised learning and domain adaptation methods have been explored. They are able to lower the requirements of large-scaled data labeling. However, this mitigation is limited and can only handle limited data distribution changes.

The method presented in this thesis targets data distribution changes caused by sensible physical attribute changes. The assumption is that for some physical conditions, the data distribution change degree correlates to the physical attribute change degree. With this relationship between the physical attributes and data distributions, the learning order can be guided by the measured physical attributes to improve accuracy [106]. I will further discuss the details of this iterative learning guided by physical attributes in Chapter 6.

Chapter 4

Human-Induced Structural Vibration Signal Acquisition

The sensing component is important for the information acquisition system because the sensed signal quality directly determines the information contained in the signal. Most of the algorithms for information acquisition and learning can achieve high accuracy when the data is 'good', but what defines 'good data'? Different applications or sensing system have different criteria. In general, the data quality can be measured by the following metrics, including time resolutions (sampling rate), spatial resolution (deploy density), noise level, and signal resolution. In this section, the problem will be focusing on is the signal resolution, because it directly determines the information contained in the sampled data. For example, a footstep induced signal is of a high resolution of 16-bits contains more detailed information, such as the striking duration, striking sharpness, etc., than a 1-bits signal, which indicates there is or is not a footstep happening.

In this chapter, the overview of the structural vibration based sensing system will be introduced (Section 4.1). Once the raw vibration signal is obtained by the system, the segments induced by human are extracted for further analysis (Section 4.2). The analysis of the obtained signal, especially the signal magnitude, will be explained to assist the understanding of the sensing range of this sensing modality (Section 4.3). As the core methodology, physical insights are used to assist sensing system achieving higher signal quality by adapting the hardware configuration collaboratively and determine their collaborative networking group (Section 4.4).

The design is then evaluated with real-world experiments. Two examples of system configuration adaptation based on the physical measurements are presented with a shared goal of improving sensing signal and eventually enhancing the learning accuracy. (Section 4.5) 1) The measured target signals can be used to prediction the upcoming event and its optimal hardware settings, and the physical understanding of the human mobility and the wave propagation properties enable accurate prediction for adaptation; 2) the measured noise signals can be used to analyze the structural factors that affect the signal propagation to further enhance the prediction and achieve sensor grouping for accurate collaborative information learning.

4.1 Hardware Design of the Sensing Node

The hardware design sensing acquisition module consists of three parts: sensor, sensing configuration board, and controller board. The sensor used in this thesis is the geophone sensor. The geophone converts the surface displacement into voltage change in a manner of a piezo dynamic sensor. To be more specific, the SM-24 geophone [54] is used. The sensor allows full bandwidth at 2-ms sampling and provides the sensitivity of 28.8 V/m/s [54]. The rest of the system went through a few iterations due to different application specification, which will be discussed in Section 4.1.1 and Section 4.1.2.

4.1.1 Basic Version

The basic version of the sensing node is shown in Figure 4.1, and the design goal for this version of sensing node is to explore different hardware configuration parameters.

Since the target vibration is of small magnitudes (10^{-4} to 10^{-6} m/s), amplifiers are needed to enhance the signal representation voltage before they are digitized. In addition, since the target vibration is of high variation (10^{-4} to 10^{-6} m/s), multiple OpAmps with different gains are placed on the sensing configuration board and connected to the sensor to achieve different amplification gains. In the basic version of the sensing node, an amplifier board with three amplification gain configurations are connected to the geophone. The output pin of each OpAmp gain is then connected to an ADC pin on the controller board – a customized Arduino Due board.

The amplified analog signal is then converted to digital signals by the Analog-to-
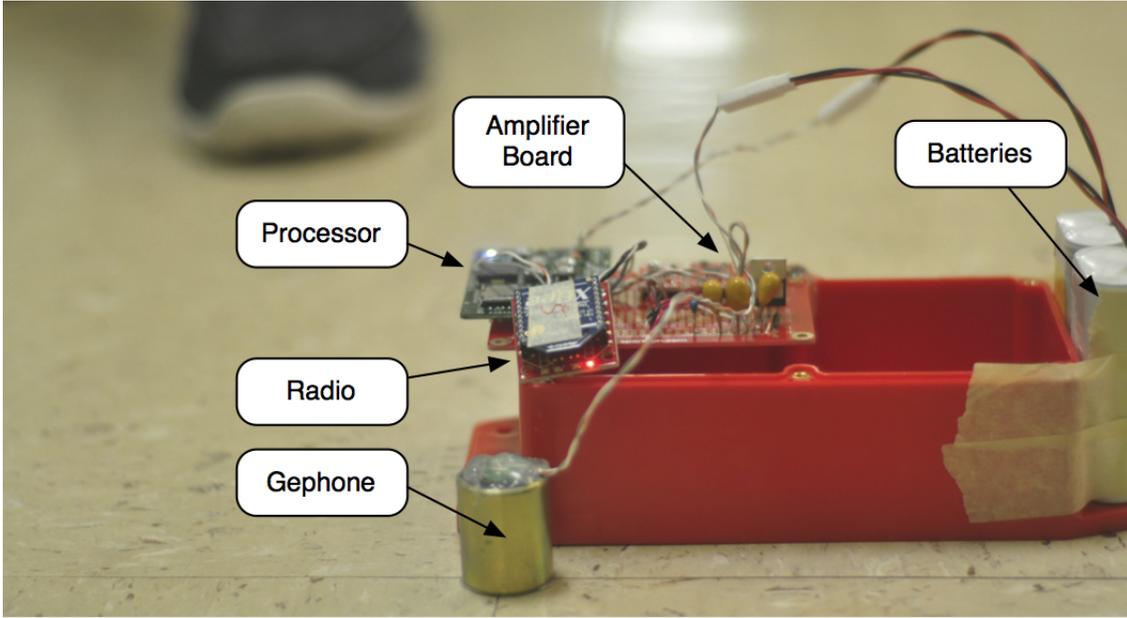
Figure 4.1: Photo of a sensing node placed on the floor [106].

Digital Converter (ADC). An n-bits ADC converts the signal of full scale voltage range E_{FSR} into 2^n values, and the voltage resolution of the signal is $E_{FSR}/2^n$. However, the actual resolution of a designated signal after ADC may not be as high as n-bits, since the signal voltage representative may be significantly lower than the full scale voltage range E_{FSR}, which will be further discussed in the next section 4.3.

On the controller board, the processor determines the time of sampling, data writing and wireless communication (networks). An XBee radio is used for information exchange between the sensing node and the central control node. Various versions are further developed to meet different application requirements, such as high-resolution signal synchronization, smaller size, communication stability.

4.1.2 Enhanced Version

The functionality of the sensing nodes are adjusted based on the application requirements. The iterations that I explored are listed as below:

- **Synchronization: Enable Step Level Localization**

    The basic version of the sensing system does not provide sample level synchronization, which became a problem when used for localization of the footstep

23
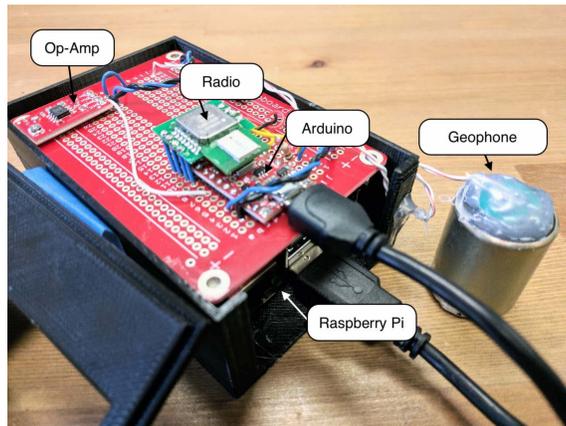
(a)
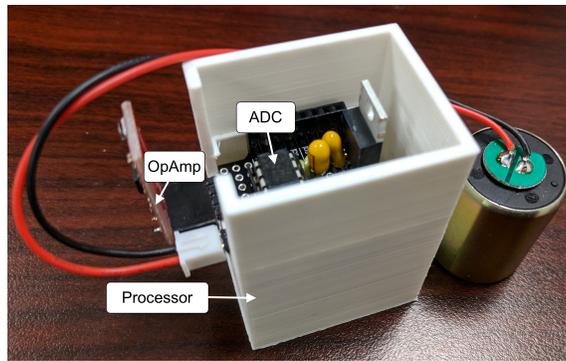
(b)

(c)

Figure 4.2: Different versions of sensors for various application requirements.

via Time Difference of Arrival between pairwise sensors. Therefore, upon the design of the first basic version, the second version, as shown in Figure 4.2 (a), utilized a high-resolution radio – Decawave DWM 1000 – to achieve sample level synchronization. Adjacent sensing nodes that cover the same area receives the broadcasted timestamps from the central control node. The physical layer implementation of DWM 1000 module allows high resolution of timestamp recording upon signal receiving [28].

- **Size: Fast and Easy Deployment**

  When deployed in various structures for different types of signals, it is important to be able to conduct characterization through temporary easy deployment. Therefore, the third version of the sensing node, as shown in Figure 4.2 (b), is implemented. Each sensing node is designed as a web-server and whichever device with a display can connect to the sensing node through WiFi connections and visualize the vibration signal in the browser.

- **Stability: Long Term Deployment**

  For long-term deployments in various environment, the wired data exchange, as well as control, shows higher stability. For example, when deployed in the pig farm with a metal bar covered floors, the wireless communication may fail due to the shielding. In addition, compared to the wireless communication, wired communication through Ethernet connection has a higher band and less package loss. Therefore, for the large-scale long-term deployment in the structure – underneath the floor or in the floor – the fourth version of the sensing nodes, as shown in Figure 4.2 (c), conducts transmission through wired communication through Ethernet cable.

For a particular deployment, the system requirement, as well as functionality priority, needs to be analyzed to make the final decision on which version of hardware to install. For example, when deployed in hospital or nursing home, wireless and 'invisibility' is of higher priority since the standard medical procedure cannot be interrupted by the sensing equipment. On the other hand, when deployed in a school building that has prior installed power line and sensor installation space, the stability for long-term monitoring becomes the design priority. Despite the variation between different applications, the sensing part (sensor + amplification board) design directly determines the signal quality for the later information extraction. In the rest of this
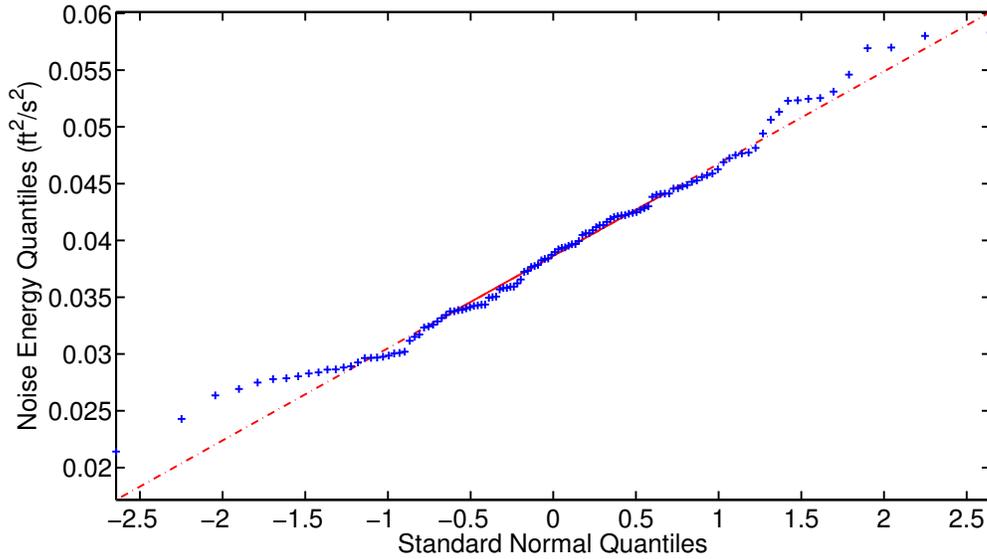
Figure 4.3: Q-Q plot of noise window energy sequence indicates a Gaussian distribution of the energy in noise windows. [97].

section, I will first introduce how to extract the target signal from the overall raw signal (Section 4.2), then discuss the factors that affect the signal quality and the methodologies to enhance it (Section 4.3.1).

4.2 Human-Induced Vibration Signal Extraction

The analysis, characterization, and learning on human-induced vibration signal rely on the extraction of relevant signal segments. For single footstep induced impulse signals, the Step Event (SE) Detection (Section 4.2.1) is applied. For non-impulsive or overlapping impulsive signals, the Signal of Interest (SoI) Extraction (Section 4.2.2) is applied.

4.2.1 Step Event Detection

The system detects distinctive signal segments induced by footstep impulses, which are referred to as *Step Events* (SEs) in the rest of this thesis. They are extracted from the vibration signals through anomaly detection based on a Gaussian model of the ambient noise (i.e., vibration signals detected when there is no excitation applied

26
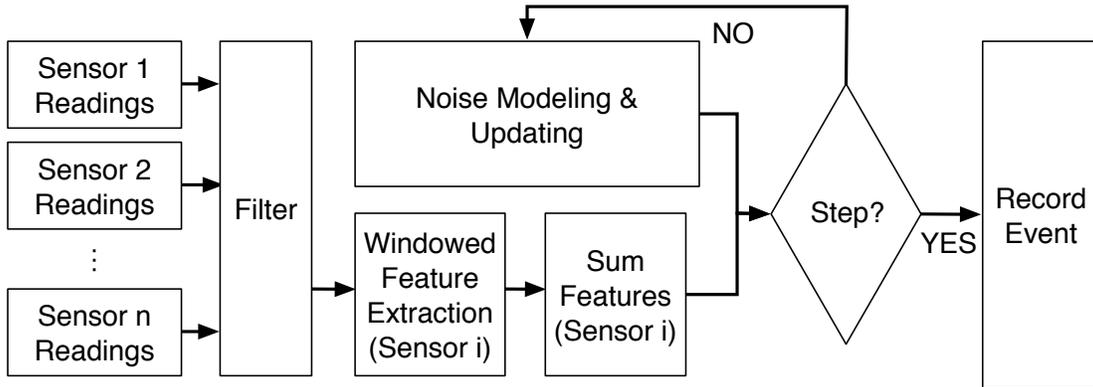
Figure 4.4: Event detection algorithm for sensors in a sensing area [97].

on the structure) [97].

The system calculates the signal energy for each windowed signal, and a series of noise signal window is then used to model the noise as the Gaussian distribution $\mathcal{N}(\mu, \sigma)$. The Gaussianity of the noise signal energy samples was tested using the quantile-quantile (Q-Q) plot in Figure 4.3. It compares the sample quantiles with the standard normal distribution quantiles. Perfect Gaussianity results in a straight line for the Q-Q plot. Figure 4.3 shows that most of the samples lie on a line. The samples fit the line better in the center (e.g., [-1, 1]) and deviate towards the tails. It validates the Gaussianity assumption and the characteristics of the tails may be caused by low sample size at the extremes.

Figure 4.4 shows the Step Event detection algorithm. First of all, a low pass filter is applied to the sensing data collected from different sensors [97]. Then a sliding window is used to segment signals. The system first collects a series of background noise signal segments and a Gaussian noise model is built based on the feature (signal energy) extracted.

When the new windowed signal is detected, its signal energy is calculated and compared to the Gaussian model. If the signal energy falls out of the threshold, e.g., 3 σ, this signal window is considered as an abnormal segment containing signal induced by people. The system detects peaks within the consecutive windows that contains detected signal segments and extracts a segment with designated length as a *Step Event*.

On the other hand, when a signal window is detected as noise, it is then included

27

in the series of prior noise signal windows. The system updates the Gaussian model for the ambient noise, so that the model adapts to varying noise levels through different time periods of the day.

4.2.2   Signal of Interest (SoI) Extraction

The Signal of Interest (SoI) is defined as the ambient structural vibration signal induced by people. On the other hand, the ambient structural vibration signals sensed by the system when there is no impulse applied to the structure within the system's sensing range is considered as noise.

The non-impulsive signals (e.g., swipe or drag an object on a surface) or overlapped impulsive signals (e.g., multiple people walk together and their footstep signals overlap) may not be covered by a segment with a fixed length. To extract SoIs from continuously detected ambient structural vibration signals, i.e., separate the SoIs from the noise, the anomaly detection is [67, 97, 99]. The anomaly detection models the noise signal as Gaussian noise and detects signals that fall out of three standard deviations (σ) from the mean (μ) as SoIs. Similar to the Step Event Detection algorithm introduced in Section 4.2.1, a sliding window is used to segment the raw signal. For each sliding window, the signal energy is calculated. Then the anomaly detection is conducted based on the signal energy of each sliding window. When the mixed impulsive signals are not occurring exactly at the same time, the energy of the sliding window may stay high for a longer period due to the signal mixture and variation of signal offsets. Similarity for the non-impulsive signals, the energy of the sliding window may stay high for an unknown period of time. Therefore, using the fixed-length window to extract signal [67, 97] may not contain the entire signal mixture. Therefore, the SoI is detected as the consecutive windows with high energy over a threshold [99, 103].

4.3   Signal Magnitude and Sensing Range

The output signal from the sensor is the voltage reading. I refer to the voltage range of the sensor output as the signal magnitude. The signal magnitude various due to varying factors, including structural material/size, signal strength (which is affected by a person's built), as well as the distance between the excitation and

the sensor. I refer to the distance between the sensor and the farthest distance a particular excitation's signal amplitude is still higher than the ambient vibration noise magnitude as the sensing range. The signal magnitude and the sensing range are two connected concepts: for a fixed sensing configuration, when the excitation induces a signal of high magnitude, the sensing range of this excitation is large.

In this section, I will first analyze the relationship between the perceived signal amplitudes and the hardware configuration of the amplification gains (Section 4.3.1), and then explain how to select the amplification gains based on this relation (Section 4.3.2).

4.3.1 Perceived Signal Amplitudes and Amplification Gains

For a particular amplification gain G and an excitation with a magnitude of M that is d away from the sensor, the perceived sensing signal amplitude A can be represented as $A = f(M, G, d)$. The sensing range and the perceived signal amplitude relation is as shown in Figure 4.5. The theoretical signal amplitude is $A_{theory} = k \times G \times \frac{M}{d}$, where k is the structural response factor.

When the excitation is within the clipping range, it means the voltage supply for the ADC E_{FSR} is lower than the theoretical signal amplitude A_{theory}, and the actually perceived voltage amplitude is E_{FSR} instead of A_{theory}. When the excitation is within the sensing range and outside the clipping range, the perceived signal

Clipping Range

Sufficient Resolution Range
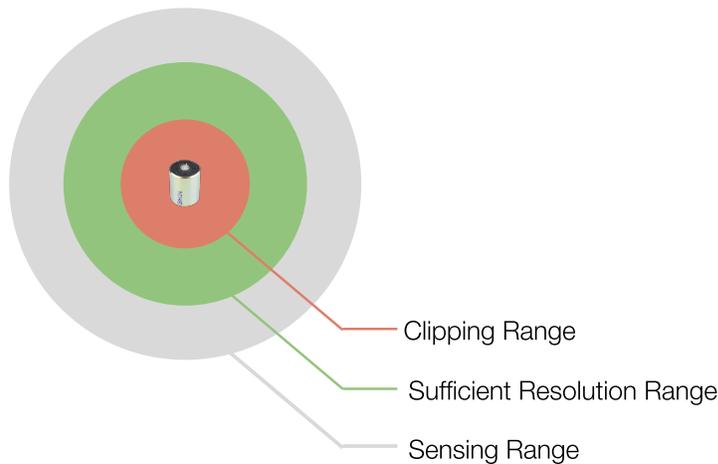
Sensing Range

Figure 4.5: Sensing range and signal resolution.

amplitude follows the A_{theory}. When the excitation is outside the sensing range, the perceived signal strength is lower than the ambient noise level, and cannot be detected anymore. Therefore, we consider the perceived signal amplitude outside the sensing range as A_{noise}.

For the excitation that falls into the sensing range and outside of the clipping range, we consider those, which are represented with a number of values larger than a threshold, as having 'sufficient resolution', which we marked as the green area in Figure 4.5. Different amplification gains will have a different resolution range as shown in Figure 4.5. Therefore different hardware configurations are needed to obtain vibration signal with sufficient resolution under different sensing conditions (e.g., different people, structural materials).

4.3.2 Amplification Gain Optimization and Implementation

The design goal here is to capture high fidelity structural vibration signals induced by indoor pedestrians using low-cost low-dynamic-range sensors. For a particular n-bits ADC module, the 'sufficient resolution range' is defined by application requirements or users, e.g., the signal that is represented by (n-1)-bits is of sufficient resolution. An optimization problem is formed to achieve the aforementioned goal with a minimum number of amplification gains, considering the limitation of the accessible ADC pins as well as to reduce the cost. In this section, I will first introduce the optimization problem (Section 4.3.2.1), and then present the implementation of the hardware based on the optimization calculation (Section 4.3.2.2).

4.3.2.1 Optimization Calculation

Given the perceived signal amplitude as discussed in Section 4.3.1, the cumulative distribution function (CDF) is derived as follows:

$$F_A(a) = P(A \le a) = \begin{cases} 0 & a \in [0, \dfrac{k}{\sqrt{R_{sensing}}}) \\ 1 - \dfrac{k^4}{R_{sensing}^2 a^4} & a \in [\dfrac{k}{\sqrt{R_{sensing}}}, A_{max}) \\ 1 & a \in [A_{max}, +\infty) \end{cases} \qquad (4.1)$$

Based on the CDF, the optimization problem is formed as follows:

$$\max_{g_1,\cdots,g_n} \sum_{i=0}^{n} F(\frac{A_{max}}{g_i}) - F(\frac{A_{sufficient}}{g_i}) \tag{4.2}$$

$$s.t. \quad 1 < g_i < g_{i+1} \quad \forall i \in \{1, \cdots, n-1\} \tag{4.3}$$

$$\frac{A_{sufficient}}{g_i} \geq \frac{k}{\sqrt{R}} \quad \forall i \in \{1, \cdots, n\} \tag{4.4}$$

$$\frac{A_{max}}{g_{i+1}} \leq \frac{A_{sufficient}}{g_i} \quad \forall i \in \{1, \cdots, n-1\} \tag{4.5}$$

where the probability of the signal falls into the sufficient resolution range without clipping is maximized for a series of amplification gains g_i. The solution of g_i is calculated as follows:

$$g_i = (\frac{A_{sufficient}}{A_{max}})^{n-i} \cdot \frac{\sqrt{R} \cdot A_{sufficient}}{k} \quad \forall i \in \{1, \cdots, n\}. \tag{4.6}$$

4.3.2.2 Amplification Board Implementation

Based on the optimization, the amplifier board as shown in Figure 4.1 is implemented as shown in Figure 4.6, where the $(OpAmp + Filter)_i$ is implemented with a gain of calculated g_i, and the k is calculated based on the average perceive excitation amplitude at distance of 1 meter [105]. The floor vibration velocity can vary from 10^{-6} m/s (e.g., concrete structure) to 10^{-4} m/s (e.g., wooden frame structure) [97], which corresponds to the voltage change between $\sim 3 \times 10^{-5}$ to $\sim 3 \times 10^{-3}$.
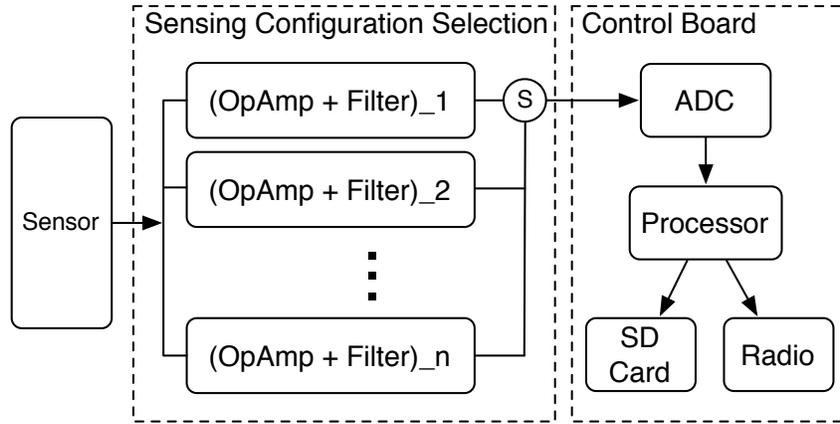
Figure 4.6: Hardware design of the sensing node [105].

The amplification gain needed is between 1k× [97] to 100k× [105] to achieve the input voltage for ADC of 3.3 volts, which allows full resolution range of signal representations. The maximum sensing range observed across different deployments is ~ 10 m radius.

4.4 Physical Measurements Guided Collaborative and Adaptive Hardware Configuration

To obtain accurate information from vibration signals requires 1) the signal waveform been preserved from various distortion, 2) enough resolution after digitization, and 3) right combination/placement of the sensing locations. There are many factors that affect the magnitude of signals obtained by sensors placed at specific locations, such as the excitation strength, distance to the sensor, and structural factors. To obtain high fidelity (i.e., high resolution and low distortion) signal for all excitation induced vibration signals is challenging due to the combination of these factors. On the other hand, due to the complicated structural factors in real buildings, the sensor placement, as well as grouping, affect the interpretation of the information from the signals.

When a person walks on the floor passing by the sensor, their footsteps induce the floor to vibrate. The system detects these footstep induced vibration and then further infer the human information (e.g., location, identity, physical conditions) from it. For a particular footstep happens at a particular location, the hardware configurations determine the quality of the perceived footstep induced vibration signal in terms of signal amplitude. After analog to digital converter, the signal amplitude in turn affects the digitized signal's resolution.

The causes of variation in detected human footstep strength mainly fall into two categories: human and environmental. The human variation includes two aspects: 1) the personal level as inconsistencies of individual footstep-to-sensor distance within a series of steps (we refer it as a *trace* in the rest of the paper), and 2) the interpersonal level as variations between individuals.

On the other hand, environmental variation occurs when the sensors are placed at different locations and also includes two aspects: 1) different impact point may have different impacts response due to material variations (e.g., stiffness, size, thickness),
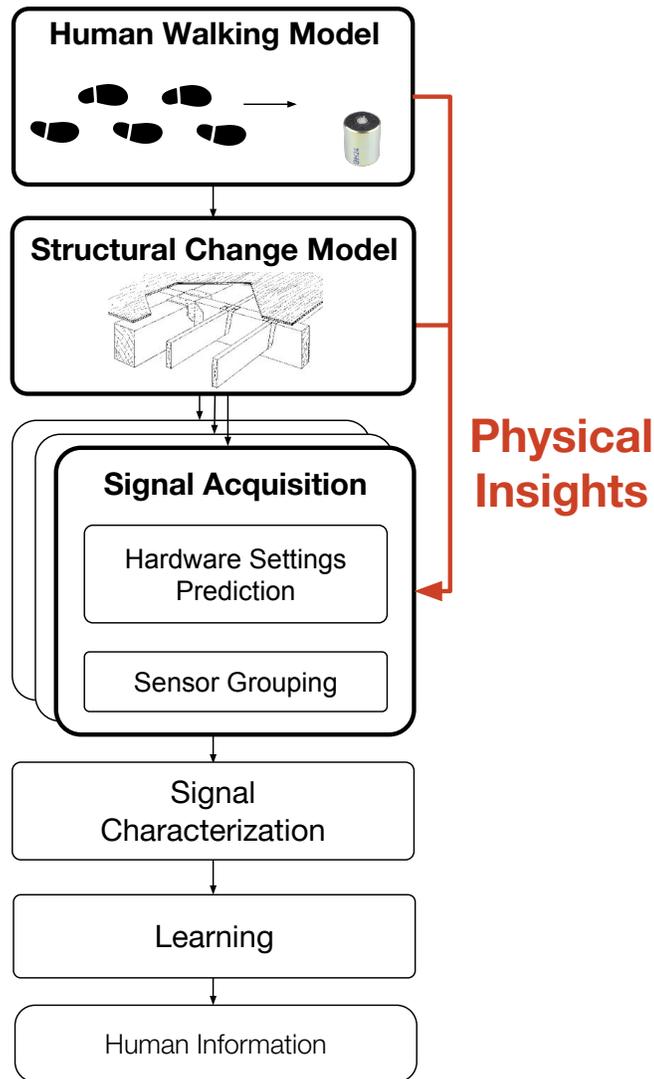
Figure 4.7: Signal acquisition module overview.

and 2) when the wave propagates through different structural factors like beams and partitions, the distortion along their path may vary.

To solve these raising problems, I'll introduce a collaborative and adaptive sensing solution that takes physical insights into account to predict the hardware adaptation and configuration. Figure 4.7 shows the overview of the framework used to acquire high fidelity sensing signals with the assistance of the physical insights. The physical insights from the human aspect are that the trajectory of a person walking by is continuous, i.e., the people has to take steps by steps to get to a new location instead

of telepathing. The physical insights from the structural aspect are that the sensors can detect structural factors (e.g., structural elements such as beams or partitions) through both the analysis of the human-induced vibration signals as well as the ambient vibration of the structure.

In this section, I will discuss in total three strategies to obtain high fidelity sensing signals and determine the group of sensors to use for collaborative information acquisition. First of all, as aforementioned, human trajectory, e.g., when a person approaches and then leaves the sensor, can be used to predict the perceived signal amplitude of the person's next footstep locally only based on the vibration detected by a particular sensor (Section 4.4.1). Then the hardware can be adapted based on this prediction. Second, since the sensors can collaboratively estimate the structural condition, e.g., determine if there is a structural factor between two sensors, this information can be further used to enhance the hardware configuration prediction accuracy (Section 4.4.2), as well as to guide the sensor grouping mechanism (Section 4.4.3).

4.4.1   Human Trajectory-Based Local Prediction

The goal of the human trajectory-based local prediction is to obtain footstep induced structural vibration signal with high-resolution and low distortion (without clipping). This prediction will be referred to as Local Profile Prediction (LPP) in the rest of the section, It predicts the optimal configuration for the next footstep signal that the sensing nodes will detect. To achieve this, the system first detects footstep-induced signals (Section 4.2.1). Then, it analyzes the detected signals' resolution condition (Section 4.4.1.1). Finally, based on the analysis, it makes a prediction on the next step's amplitude and changes the amplification configuration during a pedestrian approaching/leaving the sensor (Section 4.4.1.3).

4.4.1.1   Signal Resolution Analysis

For each detected Step Event, the system analyzes its resolution to predict the optimal configuration for the next Step Event. The Step Event resolution is deduced from the relation between the analog signal amplitude and resolution shown in Table 4.1. For an N-bit analog-to-digital converter (ADC), the $A_{sufficient}(v)$ and A_{max} are converted to a function of N as $DA_{sufficient(N)}$ and $DA_{max}(N)$. These thresholds are

Amplitude	Resolution
$(0, A_{sufficient})$	Insufficient
$[A_{sufficient}, A_{max}]$	Sufficient
$(A_{max}, +\infty)$	Clipping (distorted)

Table 4.1: Relationship between perceived signal amplitude and resolution.

applied to the detected Step Event range to determine the signal's resolution class based on the relation demonstrated in Table 4.1.

4.4.1.2   Physical Insight: Vibration Signal Decay

To obtain a model for the Step Event energy propagation, a characterization experiment is conducted in the hallway and the perceived signal energy of excitation at different distances is measured. Figure 4.8 shows the relationship between the measured energy of uniform ball drop impulse events and the distance of that event from the sensor. The impulse's distance from the sensor increases, the energy of the event decreases with a rate of roughly $\dfrac{1}{distance}$, which fits the description of the Rayleigh-Lamb wave [147].

4.4.1.3   Optimal Configuration Prediction

The optimal configuration for the upcoming Step Event is calculated by Algorithm shown in Figure 4.9: 1) predict the amplitude of the next Step Event and 2) calculate the amplification gain that allows maximum resolution without clipping.

When there are enough numbers ($> Th_{history}$) of Step Events within a walking trace, the system takes $Th_{history}$ number of prior step signals' into account to predict the amplitude of the next Step Event. When there are less than $Th_{history}$ number of steps detected in history, the decision is made based on the prior step signal. If the step history is almost linear, which is the most common step energy change behavior when the steps are far away, the system predicts the next Step Event amplitude Amp_{t+1} with the linear model estimated from the step history. On the other hand, if the step history is not linear when steps are near the sensor, the system predicts the next Step Event amplitude Amp_{t+1} with the $1/d$ model [97] estimated from the step history.

35
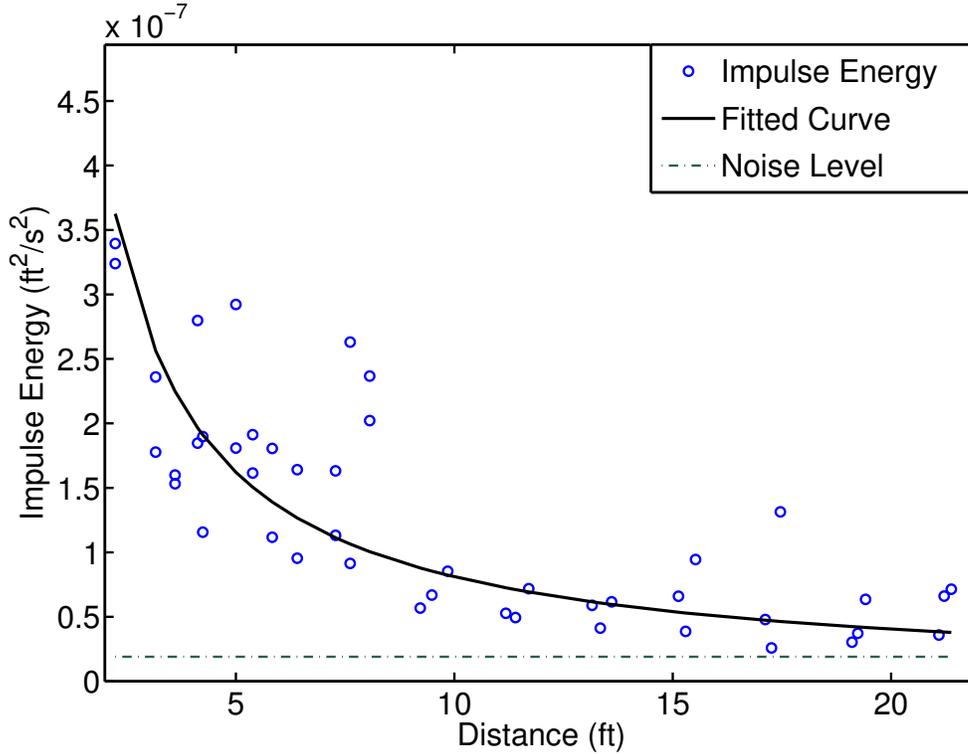
Figure 4.8: The relationship between impulse energy and impulse-sensor distance [97].

The system calculates the predicting amplification gain of the cases where the pedestrian approaches and leaves the sensing area separately.

- **Approaching.**

  When the pedestrian approaches, the system overestimates the predicted amplification by $Noise_{Amp}$. The algorithm finds the maximum level of amplification gain g_i that keeps $Amp_{t+1} \times g_i + Noise_{Amp}$ from clipping.

- **Leaving.**

  When the pedestrian leaves, the system underestimates the predicted amplification by $Noise_{Amp}$. The algorithm finds the maximum level of amplification gain g_i that keeps $Amp_{t+1} \times g_i - Noise_{Amp}$ from clipping.

Then the system adjusts the amplification gain based on this calculation for the next Step Event detection.
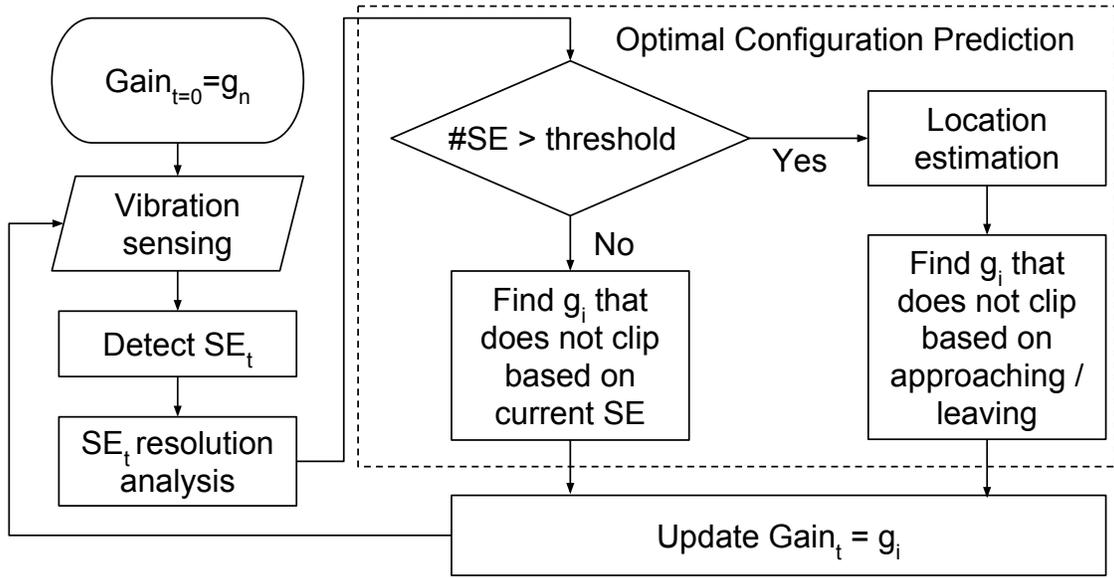
Figure 4.9: Local profile prediction algorithm (LPP) [105].

4.4.2 Structural-Based Collaborative Prediction

The goal of the structural-based collaborative prediction is to predict and adapt the hardware configuration to avoid Step Event signal distortion (e.g., clipping) for the high amplitude step signals. It is referred to as the Global Profile Prediction (GPP) in the rest of the section,

In real-world buildings, there are structural factors/changes such as beams and partitions, which affects the signal propagation. When the excitation and the sensor are on the same side of structural changes, the perceived signal amplitude change is based on the signal decay model that is determined by the sensor-excitation distance [97]. In this situation, the changes of the footstep induced signal amplitude can be predicted with LPP (Section 4.4.1). However, when the excitation and the sensor are on different sides of these changes, the perceived signal amplitude change does not follow the signal decay model anymore, which will lead to sudden changes in the perceived signal amplitude.

In addition, the location of the sensor (relative to the structural factors) also affects the perceived vibration signal amplitude. For example, if a sensor is placed near a beam, the detected footstep amplification is lower than that of a sensor located between two beams. This is similar to the phenomena where if a string has two fixed

ends, the center part has higher mobility than the parts that are closer to fixed ends. Because the stiffness of the structure increases at the locations that are closer to these structural factors.

To handle these structural variations, the communication between different sensing nodes in a sensor network is utilized. The information regarding structural variations is propagated through the network based on the pedestrian moving direction detection. This information propagation allows the system to conduct Global Profile Prediction (GPP) on sensors to improve their sensing resolution with historical information from other sensors.

GPP can either perform alone or be used with LPP to improve signal fidelity by taking structural variation into account. In this section, I will explain how the GPP works alone to achieve high-resolution signal acquisition for high signal-to-noise ratio step signals. Instead of processing on the Step Event level, GPP works on the Trace Event level. *Trace Event* is defined as the vibration signal induced by a person passing by the sensor, containing contiguous detected Step Events.

GPP first obtains the direction of the target trace (Section 4.4.2.1). Then, it predicts the pedestrian's trace (Section 4.4.2.3), i.e., to a specific neighbor sensing node, based on walking direction. Next, GPP propagates the pedestrian walking information towards neighboring sensing nodes that the pedestrian might pass based on their walking directions. These nodes rely on their location specifications (Section 4.4.2.3) and the pedestrian walking direction to make predictions.

4.4.2.1   Trace Event Direction Estimation

The Trace Event (TE) direction allows GPP to determine which neighboring sensing nodes a pedestrian approaches and which node they are heading away from. With this information, the system informs neighboring sensing nodes of possible structural anomalies that cause sudden signal amplitude changes, which we will detail in Section 4.4.2.3 later.

At least two sensing nodes are required to determine a person's walking direction based on the detected SE signal energy change trend [97]. Each sensing node detects the SE within a TE when the person is the closest to the sensor. When the pedestrian approaches then leave the sensor, the signal strength values of their SEs increase then decrease. The temporal information of the SE with the highest energy within a

TE detected by different sensors indicates the order in which the pedestrian passes sensors. Therefore the system determines the direction of a TE in terms of sensor location (i.e., from/to which sensor).

4.4.2.2 Trace Prediction

GPP propagates the SE information to neighboring sensing nodes. Therefore it is important for the system to be robust to direction ambiguity when a person continuously walk by a sensor. To predict which sensor the pedestrian is walking to, the system models all the deployed nodes as vertexes in a graph. If there is a physical route that a pedestrian can walk between two vertexes without passing a third vertex, there is an edge between these two vertexes. This graph is created heuristically at deployment time as a $k \times k$ binary table, where k is the node number, and the binary table entry value indicates if there is connectivity between two nodes. The binary design of the table is for computational search efficiency. When a pedestrian walks through sensing areas and their *stride direction* is detected, the system will notify all the other sensing nodes that share an edge with this node in the graph except the one that the person walked from.

4.4.2.3 Location Specification

Due to various structural factors such as beams and partitions, sensors may have different sensitivity to the same impulse (i.e., same strength and traveling distance). The goal for the GPP is to achieve high resolution for the high signal-to-noise ratio step signals by utilizing the historical information from neighboring sensing nodes. When multiple pedestrians walk by different sensors/locations, the system learns the different impulse response strength between sensors/locations.

When a pedestrian walks by one sensor and is detected, the system models their step energy change and sends it to neighboring sensing nodes that the pedestrian will pass by next. The neighboring sensing node then adjusts its own amplification configurations based on the historical data, which indicates the impulse response strength variation at these different locations. When the pedestrian approaches neighboring sensing nodes, the system detects the step signal with the highest energy through the structural variation profile as well as detected step signal strength from the last sensor.
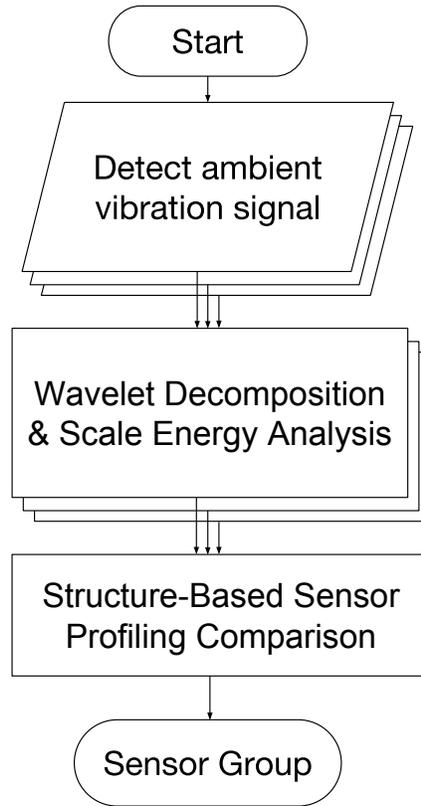
Figure 4.10: Structure-based sensor grouping algorithm.

4.4.3 Structure-Based Sensor Grouping

When the system relies on multiple sensors to make decisions or estimation, the selection of sensors (sensor grouping) for collaboratively computation becomes important. In this section, I will first introduce the physical insight on the structural changes and its effect on the perceived footstep-induced vibration signals (Section 4.4.3.1). Next, a structure-based sensor grouping algorithm will be presented, which takes discussed structural factors into account to achieve sensing grouping, as shown in Figure 4.10.

The algorithm first gathers the ambient structural vibration signals from different sensing nodes (Section 4.4.3.2). The ambient structural vibration signal is the vibration signal detected when there is no excitation around (e.g., pedestrians walking, vending machine, etc.). Then these ambient vibration signal are sent to the wavelet decomposition module, where their signal energy at each investigated scale

40
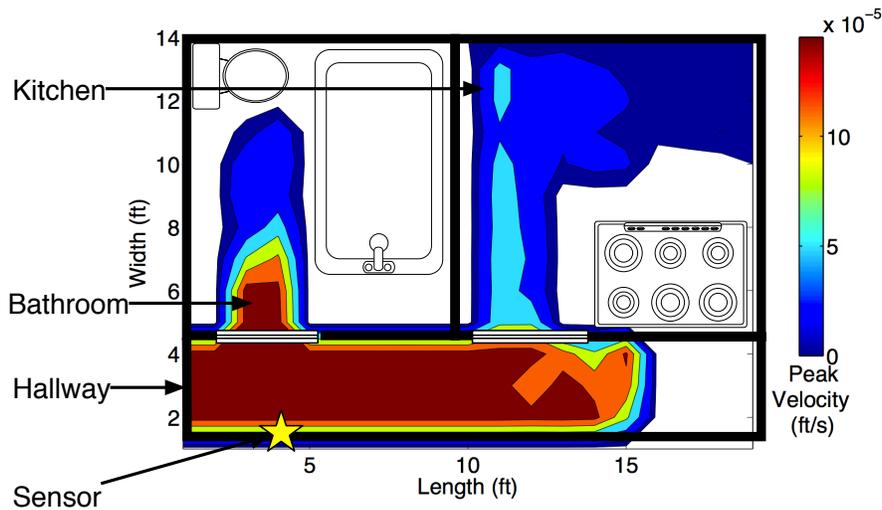
Figure 4.11: Impulse responses at a townhouse, showing structural factors effect on signal decay [97].

is calculated (Section 4.4.3.3). Once the wavelet scale energy profile of each sensor is collected, the algorithm compares these profile of sensors that are deployed in adjacent physical locations and based on their similarity, the algorithm determines if there is structural variation between two sensors that causes significant signal variations (Section 4.4.3.4). The algorithm selects the sensors that have no structural variation detected between them to be in the same group for collaborative sensing later.

4.4.3.1 Physical Insight: Structural Factor

When the excitation happens at different locations of the structure and travels through a different route to reach the sensor, the decay model may vary due to the structural factors on the path. To demonstrate this effect, Figure 4.11 shows the signal energy of uniform ball drop [1] impulse-induced floor vibration. Compared to the signal decay in the hallway (horizontal direction in the figure), the signal decay through the bathroom door and the kitchen door shows a significant drop instead of a gradual decrease in the hallway. This indicates the structural factor between

---

[1]Uniform ball drop experiment is to let a ball free fall from a particular height at varying locations to characterize structure responses for a particular sensor.

41

the hallway and the bathroom and kitchen separate the structure and increase the signal decay rate significantly when the wave travels by this structural factor.

When two sensors are placed on two sides of this structural factor, the propagation of the wave may be altered when traveling through the structural factor (e.g., loss of certain frequency components, higher decay rate, longer average propagation time). These sudden changes in Step Event signals can be used as landmarks to indicate locations [97]. However, when multiple sensors need to learn information collaboratively, these changes become an obstacle that leads to inaccurate estimation.

4.4.3.2 Detect Ambient Vibration Signals

Each sensing node mainly contains a geophone sensor, an op-amp, and an ADC module. The geophone picks up the ambient structural vibration of the surface it is placed on. The op-amp amplifies the signal and this amplified signal is then digitized by the ADC module. When placing a sensing node on the floor, the sensor picks up the ambient vibration of the structure, which characterizes the fundamental frequency band of the structure and is considered 'noise' in terms of human monitoring. In this work, we analyze this 'noise' to group the sensor so that the human information monitoring accuracy can be improved.

4.4.3.3 Wavelet Decomposition and Analysis

The wavelet decomposition is selected to analyze the 'noise' signal because it provides both fine and coarse details of the signal in both time and frequency domain [132]. The wavelet transform is represented as

$$T_x(b, a; \Psi) = w(a) \int_{-\infty}^{\infty} x(s) \Psi_{b,a}^*(s) ds \tag{4.7}$$

where (s) is the mother wavelet, which is translated by b units in time and scaled by the factor of a. We use the Mexican hat wavelet for (s). The wavelet decomposition is applied to the collected noise signal. Figure 4.12 shows an example of the decomposition. Figure 4.12 (a) shows an segment of ambient structural vibration signal. Figure 4.12 (b) shows the wavelet transform of the signal shown in Figure 4.12 (a). Then we sum up the total energy for each scale, which indicates a specific frequency component. This array of the scale-energy, which is shown in
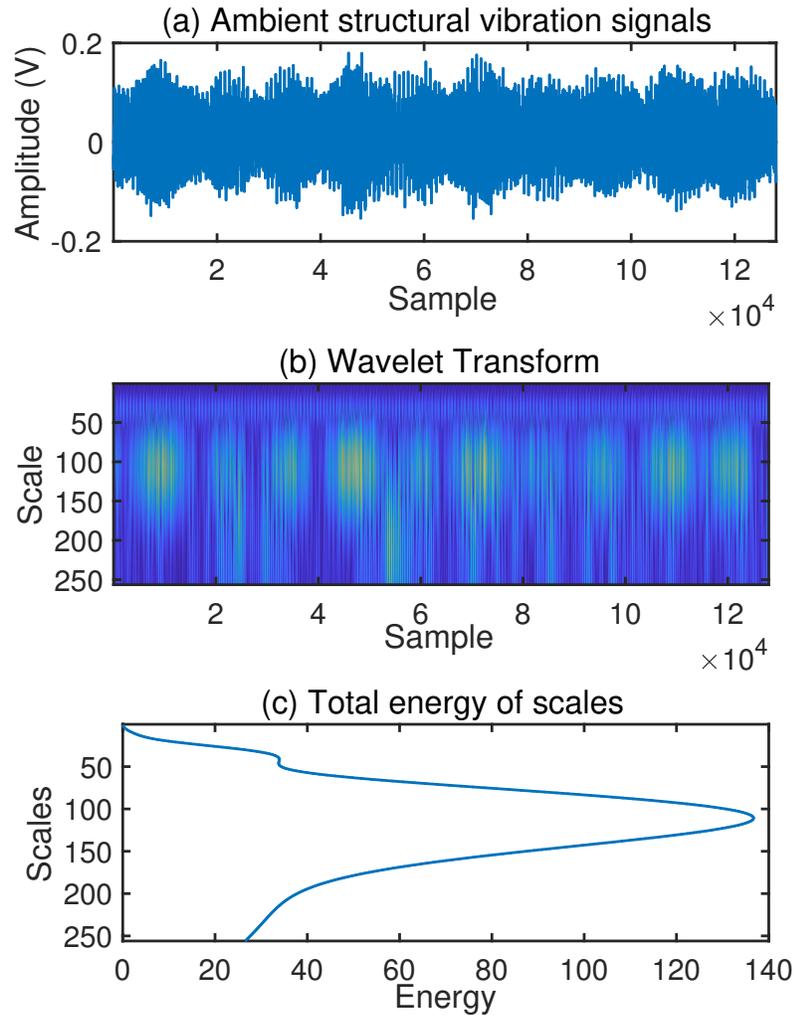
Figure 4.12: Wavelet decomposition and scale energy profiling.

Figure 4.12 (c), characterizes the structure where the sensor located and is then used as the structure-based sensor profile in Section 4.4.3.4.

4.4.3.4 Structure-Based Sensor Profiling Comparison

Once the wavelet decomposition scale energy profile of each sensor is collected, the grouping algorithm detects peaks of this scale energy profile. For example, for the sensor example in Figure 4.12, Figure 4.12 (c) shows two peaks at scale 38 and 109 respectively. The algorithm extracts the locations of the peaks to determine if there

are structural factors between a sensor pair. If two physically adjacent sensing nodes show different number/frequency component of peaks, the algorithm considers that there is a structural factor between this sensor pair.

4.5   Evaluation

To understand the importance of obtaining 'good signals' or 'high fidelity signal', as well as the efficiency of the physical measurement guided hardware configuration adaptation, various pedestrian monitoring experiments are conducted. First of all, the implementation of the system is introduced in Section 4.5.1. Next, metrics used to define the 'high fidelity signal', which is used to measure the performance of the system, is presented in Section 4.5.2. Then, three experiments targeting at different aspects of the system are explained and the results of experiments are presented in Section 4.5.3, 4.5.4, and 4.5.5.

4.5.1   Implementation

A few different version of sensing nodes are studied and introduced in this thesis. The basic version of sensing nodes introduced in Section 4.1 is used for the first two experiments in this Chapter and the enhanced version 2 sensing nodes are used for the third experiment, which localization is conducted to demonstrate the sensor grouping performance.

For the basic version of sensing nodes, the amplifier board in Figure 4.1 is implemented with $n = 3$ amplification settings. Three operational amplifiers (LMV385) with customized amplification gains are installed on the sensing configuration board. The processor board is connected to these amplifiers through three analog-to-digital converter (ADC) pins.

The sufficient resolution range defined here is $A_{sufficient} = 1/2 A_{max}$. Therefore, based on Eq. 4.6, the ratio of the optimal can be calculated from $A_{sufficient}/A_{max} = 1/2$, which is $(1/2)^2 : (1/2)^1 : (1/2)^0 = 1 : 2 : 4$. Through empirical measurements of other constants ($A_{sufficient} = 1.5, k = 3 \times 10^{-4}, R = 9$), the optimal gains are calculated as $2000\times$, $4000\times$, and $8000\times$.

The geophone SM-24 is used in all three experiments [54]. It has a sensitivity of 28.8V/m/s. The theoretical sensing range of the sensor is limited by its max

coil excursion, which is 2mm. However, in practical scenarios, the sensing range is limited by the amplifier input voltage, which is 3.3V for this implementation. Therefore, when an amplifier with $g_0 = 1$ is applied, the sensing range of the sensor is 0.1146m/s.

When a 10-bits analog-to-digital converter (ADC) is used, the resolution of the system is 1.12×10^{-4}m/s, which is not enough to observe signals with peak values fall in the range of 10^{-6}m/s and 10^{-4}m/s. Therefore, when an amplifier with a gain of $2000\times$ is applied, the sensing range of the sensor is 5.73×10^{-5}m/s, with a resolution of 5.6×10^{-8}m/s. Compare to the setting of $g_0 = 1$, this setting has less sensing range but higher resolution. Similarly, the gain of $4000\times$ and $8000\times$ enables even higher resolution (respectively 2.8×10^{-8}m/s and 1.4×10^{-8}m/s) with less sensing range (respectively 2.865×10^{-5}m/s and 1.43×10^{-5}m/s). Therefore, by combining multiple settings, the system achieves high resolution (1.4×10^{-8}m/s) as well as high sensing range (5.73×10^{-5}m/s) at the same time.

Implemented amplification gains, however, differ slightly due to practical constraints, such as resistor value availability and filtering effects of the amplifier. The LMV385 Sparkfun OpAmp Breakout board is used in the application, whose schematic is shown in Figure 4.13. LMV385 is a two-stage amplifier. Each stage has
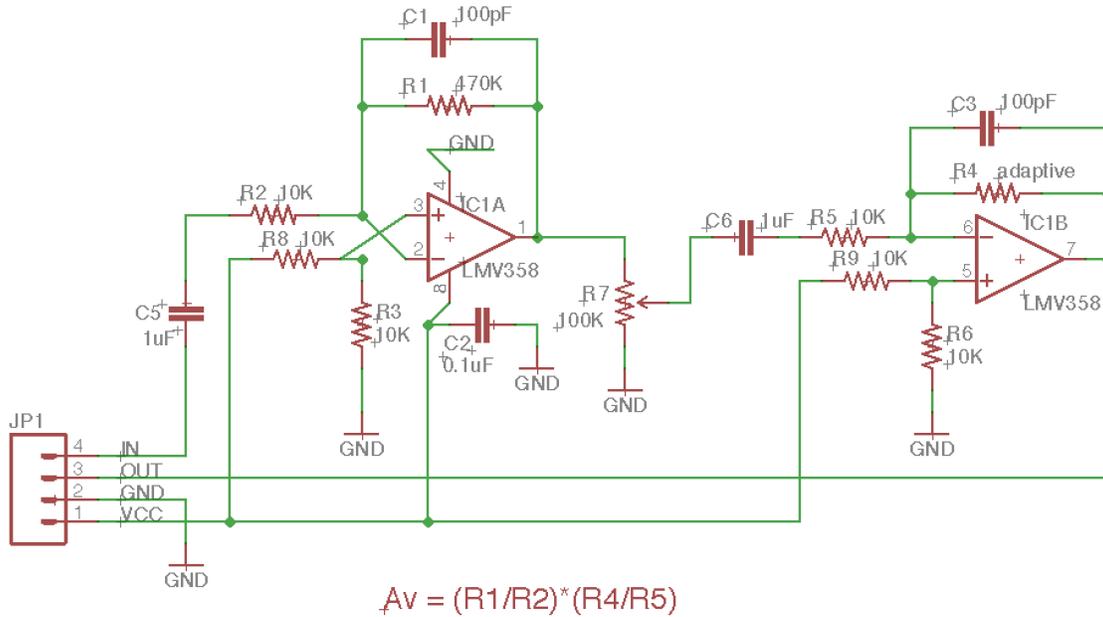
Figure 4.13: The schematic of the OpAmp LMV385 [135].

a differential amplifier serving as a bandpass filter. For the first-stage amplifier, a resistor with a value of $470K\Omega$ over $10K\Omega$ is used to achieve an amplification gain $R_1/R_2 = 470/10 = 47$. When selecting the first-stage gain, the corresponding resistor should be available and the gain should not cause clipping under most circumstances; otherwise, the clipped signal is smoothed by the second stage's filter. If that happens, the output signal of the second stage will not show evidence of clipping, even though it is distorted due to the clipping happening in the first stage.

For the second-stage amplifier, the resistor R_4 (values of $470K\Omega$, $1M\Omega$, and $2M\Omega$) and R_5 (values of $100K\Omega$) are used respectively to achieve the gain R_4/R_5. The corresponding gain for two stage amplifier is calculated as $Av = (R_1/R_2) \times (R_4/R_5)$, which are approximately 2000, 4000 and 8000. The calculated gains from this combination were $2200 \approx 47 \times 47$, $4700 = 47 \times 100$, and $9400 = 47 \times 200$. However, due to the limited open loop gain and filtering effects of the two-level op-amp circuit, the actual gains of the configuration were $g_1 = 2200$, $g_2 = 4400$, and $g_3 = 6400$ approximately [77]. With chosen configurations, over 90% of the impulses induced by detected footsteps are not clipped with g_1, and the background structural vibration noise after amplification is still less than $1/10$ of the entire resolution range with g_3.

4.5.2   Metrics and Parameters

Signal distortion and resolution are two major factors to evaluate the 'quality' or 'fidelity' of obtained sensing signals because they affect the information acquisition accuracy. Thus, the 'high fidelity' signals are defined as signals that have minimized distortion and noise and maximized signal resolution. Therefore, the metrics are signal resolution (Section 4.5.2.1), signal distortion (Section 4.5.2.2) and signal magnitude (Section 4.5.2.3).

On the other hand, two parameters/factors that affect the system performance are 1) the definition of the sufficient resolution rate and 2) number of amplification gains that are implemented (Section 4.5.2.5). In the rest of this section, I will use an example signal Step Event $[s_1, ...s_L]$ of L samples to explain these metrics, assuming

$$\frac{\sum_{i=1}^{L} s_i}{L} = 0 \tag{4.8}$$

.

4.5.2.1 Signal Resolution

Signal resolution usually refers to the number of bits used to represent a signal. In this section, for each Step Event $[s_1, ...s_L]$, the signal resolution is represented by a boolean value: sufficient or insufficient.

$$SE\_Resolution = \begin{cases} 0 & \text{argmax}_i\, |s_i| < threshold \\ 1 & \text{argmax}_i\, |s_i| \geq threshold \end{cases} \tag{4.9}$$

Therefore, for a series of detected Step Events $SE_1, ..., SE_N$, the $SE\_Resolution_j$ where $(j = 1, ..., N)$ is calculated.

The sufficient resolution rate over all detected SEs measures the general signal resolution level, which is further defined as Sufficient Resolution Rate (SRR) in this thesis:

$$SRR = \frac{\sum_{j=1}^{N} SE\_Resolution_j}{N} \tag{4.10}$$

The higher the SRR value, the more SEs of high resolution. In the analysis later in this section, the SRR is normalized by the maximum possible SRR value the given system hardware configurations can achieve.

4.5.2.2 Signal Distortion

Signal distortion refers to the degree a measured signal shape differs from the defined baseline. In this section, the distortion caused by signal clipping is targeted, which is calculated as:

$$SE\_Clipping = \begin{cases} 0 & |s_i + s_{i+1}| < 2 \times A_{max}, i = 1...L - 1 \\ 1 & |s_i + s_{i+1}| = 2 \times A_{max}, i = 1...L - 1 \end{cases} \tag{4.11}$$

The distortion rate is calculated as the proportion of SEs that are clipping, which is referred to as the clipping rate.

$$clippingrate = \frac{\sum_{j=1}^{N} SE\_Clipping_j}{N} \tag{4.12}$$

The lower the clipping rate, the less signal distortion the system experiences, hence a higher signal fidelity.

4.5.2.3 Signal Magnitude

Signal magnitude is defined as the maximum absolute value of a zero-mean step event signal:

$$SE\_Magnitude = \underset{i}{\operatorname{argmax}} |s_i|(i = 1...L) \qquad (4.13)$$

It indicates how many values are actually used to represent the signal. In the ideal scenario, the system should achieve maximum signal magnitude for each predicted step event signal. However, due to the variation and randomness in human activities as well as the monitored structure, the prediction result can vary, i.e., even a Step Event is count as of sufficient resolution, it might not have maximum magnitude. On the other hand, for different definitions of sufficient resolution, the same magnitude may be of sufficient or insufficient resolution. Therefore, the signal magnitude is adopted to reveal fine details about SE resolution.

4.5.2.4 Performance in Applications

To evaluate the importance of obtaining high fidelity sensing data, the comparison between the physical measurements guided signal acquisition and the traditional signal acquisition is used as metrics as well to demonstrate the system performance. The application used in this chapter is the occupants tracking. Different tracking algorithms that target at different resolution (e.g., area level, step level) are explored based on the signals obtained in each scenario.

4.5.2.5 Parameters

The system design is determined by two factors: 1) the definition of sufficient resolution and 2) the implemented number of amplification gains. When the sufficient resolution is of high magnitude, the adaptation would be more frequent to achieve the high SRR for the series of Step Events. Therefore, the noise of the prediction may reduce the SRR and increase the clipping rate. When the sufficient resolution is of low magnitude, the Step Events falls into the sufficient resolution range more often. Therefore the adaptation happens less often and the wrong prediction may happen less often as well. On the other hand, if we have more option of amplification gains, the dynamic range of the sensing system is higher, therefore the Step Events that falls into the sufficient resolution range is of higher probability. Therefore, in

the next section (Section 4.5.3), we first specifically evaluate the system behavior in these two factors under perfect amplification settings by generating an amplified 10-bit signal through a high-resolution oscilloscope signal of people walking by one sensor.

4.5.3 Experiment I: System Parameters

To explore how the the system parameters affects the system performance for the 10-bit ADC configurations, 15 traces are collected as the seeds for the 10-bit signal generation. Then each seed is used to generate N traces of different amplification settings. The amplification gains are selected so that, for a entire trace of the signals $[s_1,...s_L]$, the minimum amplification gain does not have any signal beyond the sufficient resolution:

$$\underset{i}{\text{argmax}}\,|s_i| < threshold(i = 1...L) \tag{4.14}$$

while the maximum amplification gain has maximum 0.5% clipped signal among the entire trace of signals:

$$SampleClipping(s_i) = \begin{cases} 0 & |s_i| < A_{max} \\ 1 & |s_i| = A_{max} \end{cases} \tag{4.15}$$

$$\frac{\sum_{i=1}^{L} SampleClipping(s_i)}{L} < 0.5\% \tag{4.16}$$

This discrepancy means the starting and ending steps are not clipped while most of the close-to-sensor step signals are clipped.

In the simulation, five sensors with different structural impulse response strength rates are generated for each collected trace. The pedestrian is simulated to walk from the sensing area covered by the first sensor to the fifth sensor. For the first sensor, the step strength for each trace is derived from the seed, and for the rest of the sensors, the step strength for each step is calculated with a ratio of $structural\_rate \times (1 + human\_noise)$ to simulate the human behavior noise as well as structural variation.

The system performance under five cases are compared: 1) only the LPP algorithm

is applied; 2) the baseline approach, which is defined as the median amplification level available; 3) the ground truth, which is the upper bound performance the system can achieve with the implemented hardware, i.e., the system rejects the settings that result in clipping signal and keeps the highest resolution signal that is not clipped; 4) only the GPP algorithm is applied; and 5) both the LPP and GPP are applied to conduct collaborative sensing. The acronyms used in the evaluation section are summarized in Table 4.2.

4.5.3.1 Sufficient Resolution Definition

To understand the effects of different sufficient resolution definition, the simulation experiments with parameters defined as $A_{max} = 1024$ and $A_{sufficient} = i/16 \; A_{max}$, with $i = 1...15$ are explored. For each i, N level of amplified traces are generated. The simulation runs the LPP algorithm through the N level amplification gains. In addition, since the GPP is focused on decreasing the clipping rate and hence increasing the sufficient resolution rate, we further explore a fourth metric, which is referred to as the 'critical signal SRR', which will be referred to as CSRR later. It calculates as the SRR of the five steps that is closest to the sensor, which have the highest signal-to-noise ratio in a trace.

Figure 4.14 demonstrates the SRR, clipping rate, signal magnitude, and CSRR values from the simulation: 1) the blue line with + markers demonstrates the LPP algorithm, 2) the red line demonstrates the baseline, 3) the yellow line demonstrates the ground truth result, 4) the purple line with circle markers shows the GPP algorithm, and 5) the green line with cross markers demonstrates results with both LPP and GPP algorithms.

When the value of $A_{sufficient}/A_{max}$ is low, the change between different amplification gain is large ($g_{i+1}/g_i = A_{max}/A_{sufficient}$). It is because that a large portion of the signal between -512 and 512 is considered as sufficient resolution. Therefore,

Table 4.2: Acronyms

Acronym	Meaning
LPP	Local Profile Prediction
GPP	Global Profile Prediction
SRR	Sufficient Resolution Rate
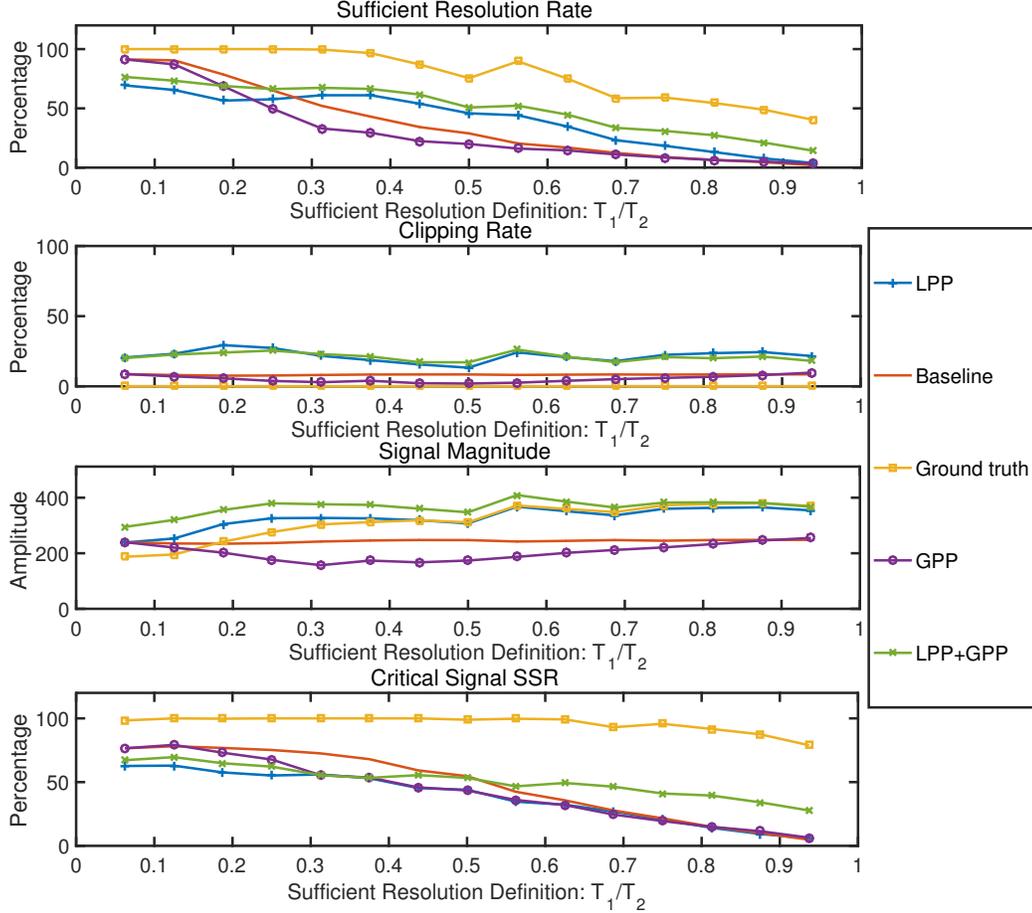
Figure 4.14: The relationship between defined sufficient resolution and the signal fidelity.

a lower number of amplifiers (N) is needed to cover the variation of the footstep signals. It also means that more low magnitude step signals are considered sufficient resolution, and have a high SRR value and low signal magnitude value.

Figure 4.14 shows that the clipping rate remains stable when the value of $A_{sufficient}/A_{max}$ increases. On the other hand, the signal magnitude increases when the value of $A_{sufficient}/A_{max}$ increases. This means that the signal quality increases, but due to the increment of the sufficient resolution definition, the SRR decreases.

LPP, in general, outperforms the baseline in terms of SRR and signal magnitude when the definition of the sufficient resolution is over $1/4$ of the entire resolution range. The average incremental values are 5% and 34% respectively.
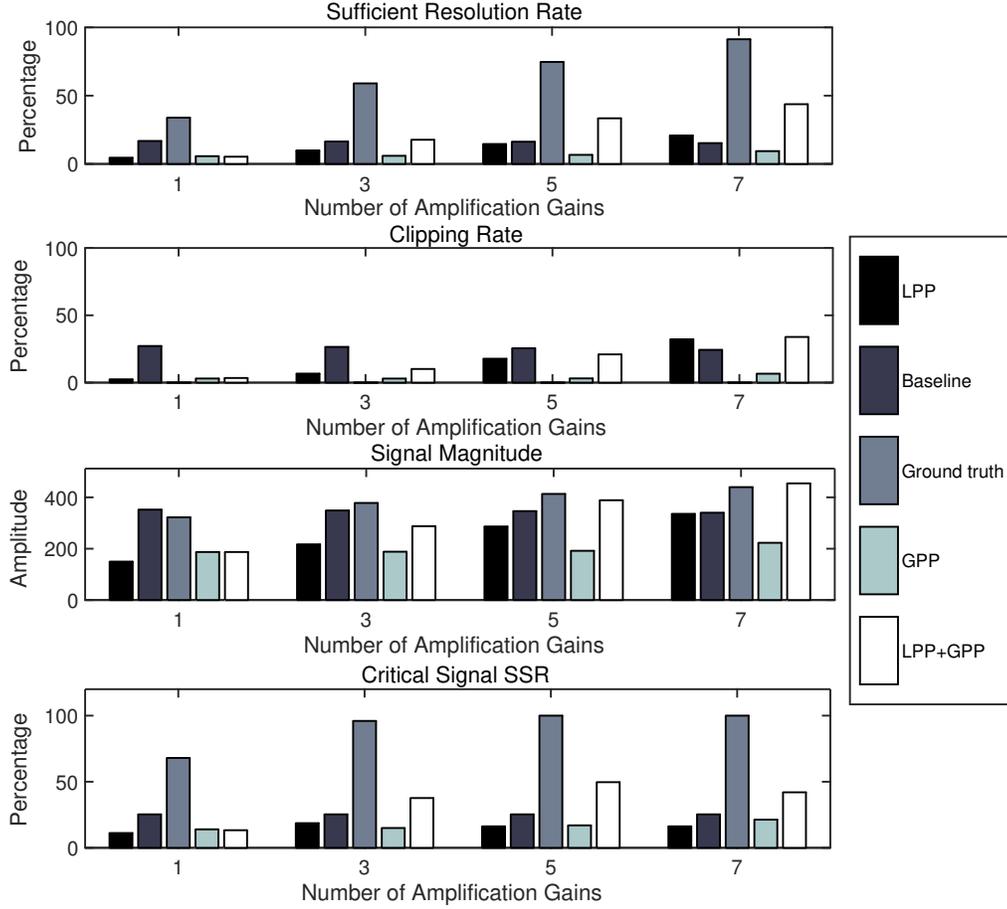
Figure 4.15: The relationship between the number of amplification gains and the signal fidelity.

On the other hand, GPP reduces the clipping rate compared to the baseline when the sufficient resolution is between 1/4 and 3/4 of the entire resolution range. It causes a clipping rate $1.6\times$ lower than the baseline. When LPP and GPP are combined, it achieves an SRR higher than either algorithm performing alone by 10% and raises the signal magnitude by 12% on average. The LPP and GPP combination follows the trend of LPP and outperforms LPP in the CSRR. The CSRR shows an average increase of 10% and up to $4\times$ increase for the highest $A_{sufficient}/A_{max}$ value when the definition of the sufficient resolution is of a high standard ($A_{sufficient}/A_{max}$ value high).
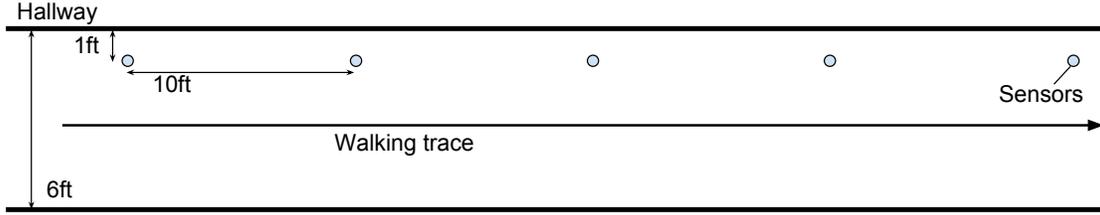
Figure 4.16: Collaborative adaptive sensing experimental setup.

4.5.3.2 Number of Amplifications

Since the number of amplification gains implemented is limited, this number affects the amplification range the system can achieve. Based on the analysis in Section 4.5.3.1, the sufficient resolution definition of $T_1/T_2 = 12/16$ is selected. With $T_1/T_2 = 12/16$, the scenario with seven levels of amplification gains is used to further conduct analysis on the number of amplification gains. The number is selected so that there are enough available amplification gains to cover the target signal range.

To understand effects of implemented amplification gain number, simulations with different number of gains are conducted. The median level gain is selected as the baseline with only one gain available. Then the number of levels is increased by adding one smaller and one larger amplification gain for each case. Figure 4.15 shows the SRR, clipping rate, the signal magnitude and CSRR under different numbers of amplification gains.

Each metric shows an increasing trend for all evaluated scenarios except the baseline. This is because that the baseline is a fixed amplification gain configuration, and it will only be affected by SRR. The more amplification gain levels are implemented, the more adaptable levels are used for selection. As a result, SRR and signal magnitude values increase.

On the other hand, due to the increasing options on the high amplification gain, the chance that the system selects a configuration that causes clipping increases too. This explains the increasing clipping rate observed in Figure 4.15.

4.5.4 Experiment II: Adaptive Amplification

To evaluate the system performance in the real-world scenario, a small-scale experiment in a school building is conducted with five sensing nodes. Geophones on sensing

nodes are mounted on the floor in a hallway (approximately $20m \times 2m$ area, tile floor) as shown in Figure 4.16. The experiment parameters are listed in Table 4.3.

Table 4.3: Hardware configurations

Parameters	Values
Sampling Rate	1000 Hz
Number of Gains	3
Distance between Sensors	10 ft
Number of Subjects	10

Subjects were asked to walk along the deployed hallway with no restrictions. Figure 4.17 demonstrates an example of floor vibration signals when each subjects walks along the hallway passing five sensors. Figure 4.17 (a,c,e,g,i) show the raw footstep-induced vibration signals perceived by each sensor. Figure 4.17 (b,d,f,h,j) show the windowed signal energy of raw signals shown in Figure 4.17 (a,c,e,g,i).

From the figure, both the variation caused by the distance between sensor and excitation and the structural changes (i.e., beams underneath the floor) can be observed. The red solid lines mark out the Step Events with the highest signal energy and it shows a trend of consecutively appearing from Sensor 1 to Sensor 5. This
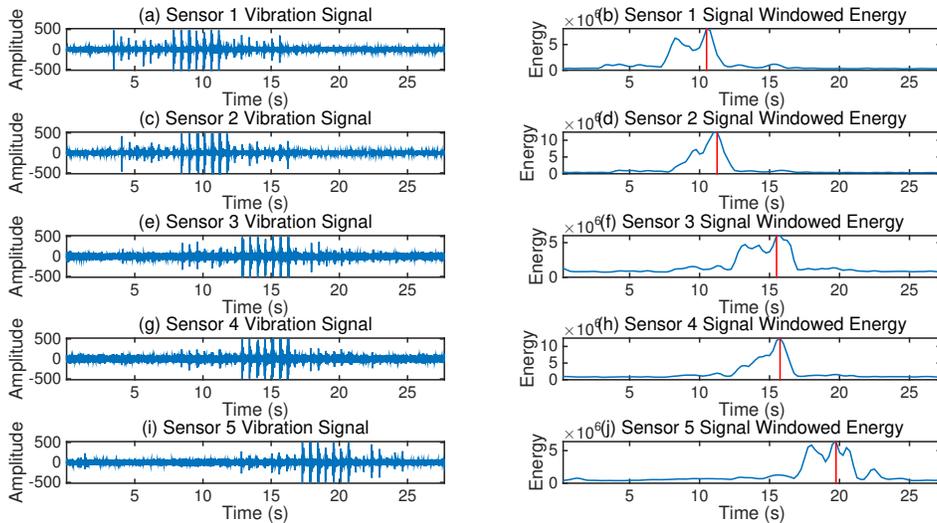
Figure 4.17: Structural vibration signal detected by sensors when a pedestrian walks by.

indicates that the person moved from Sensor 1 to Sensor 5.

The collected data is then replayed and the combined configuration adaptation algorithm is applied. SEs collected with the adaptive algorithm is compared to those collected with fixed configurations. Figure 4.18 (a) shows the normalized SRR when 1) fixed amplification configurations ($g_1 = 2200$, $g_2 = 4400$, and $g_3 = 6400$,), 2) the adaptive configuration using only LPP, and 3) the adaptive configuration when LPP and GPP combination algorithm is used. The normalized SRR value for these five cases are 32%, 36%, 61%, 67%, and 69% respectively. The system improvement comparing to g_1, g_2, and g_3 are between 1.7× and 2×. The performance of LPP, as well as LPP and GPP combination, has higher SRR values than the fixed configurations.

Note that the algorithm is designed for regular footsteps, i.e., footsteps from the same person, which are assumed to induce similar structural response. In addition, fixed padding values (P_1 and P_2 are used as described in Section 4.4.1.3). The randomness of human behavior introduced footstep induced signal amplitude prediction errors, leading to an approximately 30% lower SRR value compared to the maximum SRR value that the system can achieve. The LPP achieves higher SRR compared to that of g_1, g_2, and g_3. g_3 and g_2 amplify the near sensor signal beyond the ADC output limitations and cause a high clipping rate and a low SRR value.

To validate that, the clipping rate of these configurations are demonstrated in Figure 4.18 (b), of which values are 3%, 15%, 21%, 11%, and 11% respectively. g_1 obtains most of the near-sensor SEs without clipping. On the other hand, the SEs captured with g_1 configuration shows a low resolution value for footstep signals that are induced by steps far from the sensor due to an insufficient amplification gain. As a result, it shows low SRR values.

In order to understand the low-resolution effects, the average signal magnitude is presented in Figure 4.18 (c). g_3 amplifies the signal with the maximum gain, therefore it has the highest average signal magnitude. As a result, it also has a high clipping rate value.

The figure shows that fixed gains have an expected effect on magnitude while the combined adaptive method increases SRR values despite the reduction of average signal magnitude values. The GPP achieves slightly higher SRR comparing to LPP in this experiment due to the relatively uniform nature of the testing structure.

The system is further evaluated with a 1-D localization application. The system
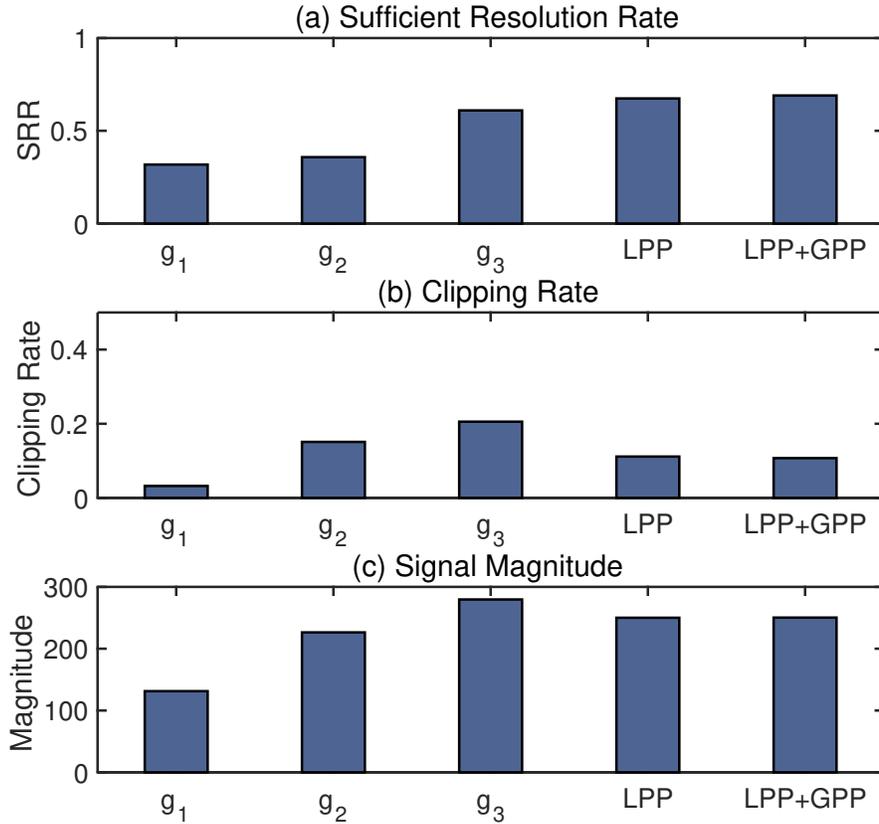
Figure 4.18: Collaborative adaptive sensing on pedestrian walking shows higher signal fidelity compared to fixed sensing configuration.

estimates the location of the pedestrian based on their footstep induced vibration amplitude decay model [147]. The system localizes the pedestrian by detecting SEs where the person-sensor distance is minimized. To evaluate that, the parameters we investigated in Section 4.5.3 are fixed to $A_{sufficient}/A_{max} = 12/16$ and the number of amplification levels as 7.

The step count error of the LPP and GPP combination algorithm is compared to that of the fixed amplification, in this case selecting the middle level (level 4). The average error for the combined algorithm in the aforementioned localization is $0.47m$, and the average error for that from a fixed amplification gain is $1.13m$. It means that the adaptive hardware configuration allows the system to achieve a $2\times$ less error when used to locate the pedestrian steps.

4.5.5   Experiment III: Sensor Grouping

The structure-based sensor grouping algorithm is evaluated through the step-level indoor localization application with data collected from two different buildings on campuses.

4.5.5.1   Application: Step-Level Indoor Localization

Step-level indoor localization provides useful human information for various smart building applications [85, 86, 115]. The step-level localization through footstep-induced structural vibration is achieved through Time Difference of Arrival (TDoA)-based multilateration. A key factor of accurate structural vibration TDoA estimation is a constant propagation velocity on the signal path. Therefore, structural factors between sensor pairs may cause inaccurate velocity estimation, hence increase the localization error.

The experimental setup on the two campus buildings is shown in Figure 4.19. Eight sensors are placed across three sensing areas, where Area 1 and Area 3 do not have any visible overhead beams, which is considered as potential structural factors, and Area 2 does. Experiments are conducted in hallways of two different buildings. Because these two buildings have different sizes, the distances between sensors are different.

The pedestrian is asked to walk through the sensing area stepping at the designated spots 10 times. The localization error and the detecting rate of the footstep-induced vibration signals using different sensor groups are compared later.
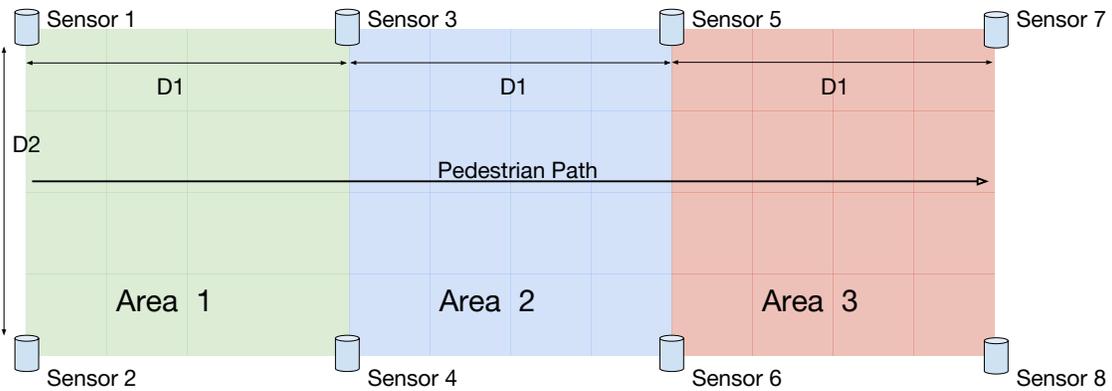
Figure 4.19: Sensor grouping experimental setup.

Table 4.4: Evaluation metrics

Metrics	Definition	Equation
Localization error	The distance between estimated and ground truth location	$\sqrt{\sum (estLoc - trueLoc)^2}$
Detecting rate	# of footsteps that has estimated locations within the sensing area ('detected') over total collected footsteps	$\frac{detectedSEs}{allSEs}$

4.5.5.2 Location I: Wood and Concrete Structure

The scale energy profile for deployment in Location I mainly falls in two clusters, indicating two slab regions. Figure 4.20 shows the two types of clusters, where the first one has one major peak at around scale 60 and the second one has two major peaks at around scale 30 and 65. The algorithm groups sensors into two groups: Sensor 1,2,3,4 and Sensor 5,6,7,8. To evaluate the performance of the grouping results, the sensor groups as shown in Table 4.5 are compared.

The wavelet analysis also indicates that the Area 1 (where Sensor 1,2,3,4 are placed on) and Area 3 (where Sensor 5,6,7,8 are placed on) in Figure 4.19 are separated floor slabs. Furthermore, the structural difference between Sensor 3,4 and Sensor 5,6 is significant.
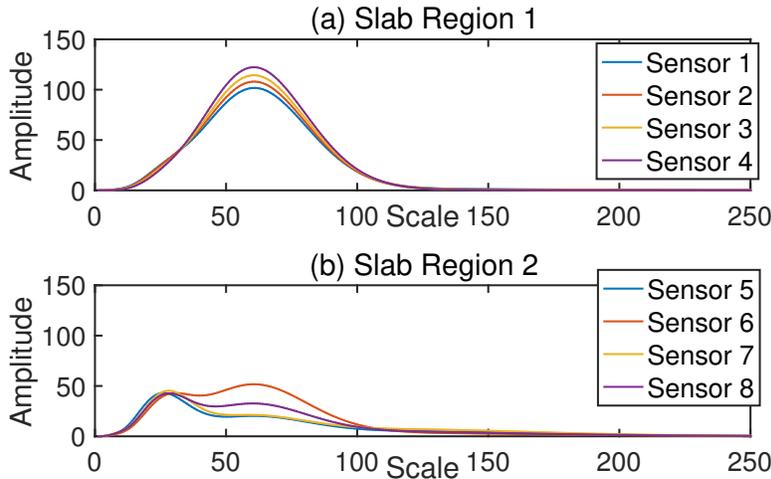
Figure 4.20: Wavelet scale energy profile at Location I.
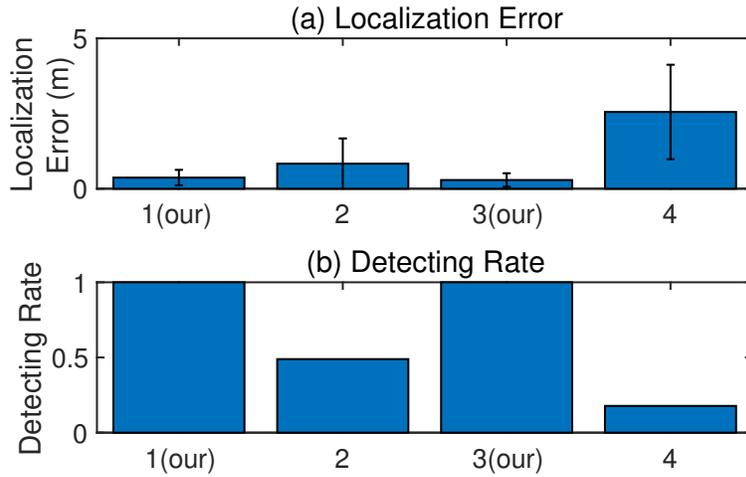
Figure 4.21: Sensor group comparison at Location I.

Table 4.5: Sensor group for Location I

Group ID	Sensors
1	1,2,3,4
2	3,4,5,6
3	5,6,7,8
4	1,2,3,4,5,6,7,8

The system performance demonstrated in Figure 4.21 based on the metrics listed in Table 4.4 to further understand our grouping algorithm. When localizing with sensor group 1 and 3, the algorithm shows significantly higher detecting rate (100% and 100% respectively) and lower localization error (0.37 m and 0.29 m respectively). On the other hand, if the sensor group 2 (sensors in Area 2) is used for localization, the detecting rate drops to 49% and the localization error increases to 0.83m (2.86× compare to group 3). When all available sensors are used, the average localization error is 2.56m and the detecting rate is 18%. This is because of the significant structural factor within Area 2, causing a different wave propagation velocity in Area 2.
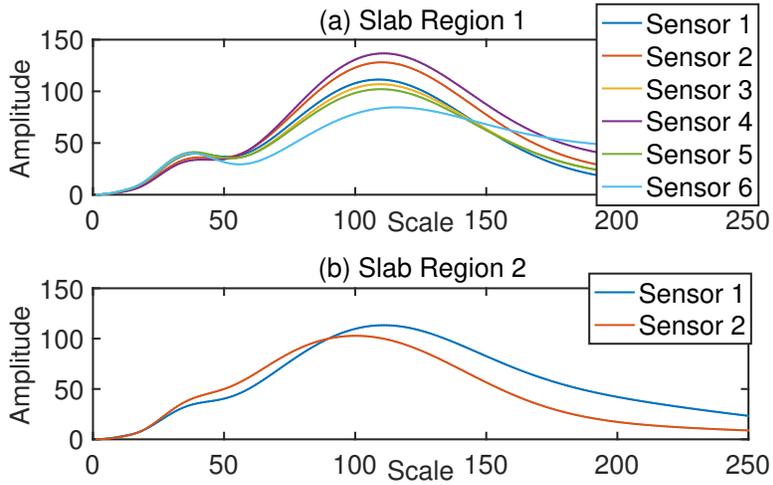
Figure 4.22: Wavelet scale energy profile at Location II.
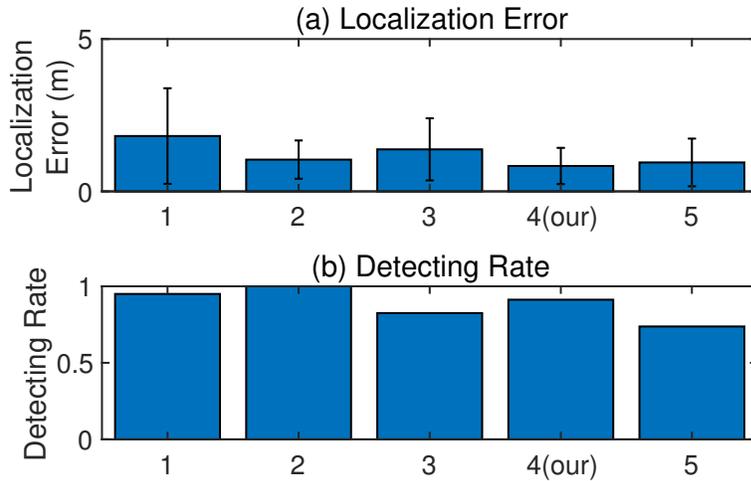
Figure 4.23: Sensor group comparison at Location II.

4.5.5.3 Location II: Concrete Structure

The scale energy profile for deployment in Location II mainly falls into two clusters as well. The first group consists of Sensor 1,2,3,4,5,6 while the second group consists of Sensor 7,8. Figure 4.22 shows scale energy profiles from the two slab regions, where the first one mainly has two peaks at around scale 40 and 110 and the second one has one major peak at around 105. Despite the fact human observation would consider a structural factor between Sensor 3,4 and Sensor 5,6, the scale energy

Table 4.6: Sensor group for Location II

Group ID	Sensors
1	1,2,3,4
2	3,4,5,6
3	5,6,7,8
4	1,2,3,4,5,6
5	1,2,3,4,5,6,7,8

profile reveals the underneath structural variation differently. It is also observed that the structural differences between the two regions are not as significant as the one at Location I.

The following sensor groups listed in Table 4.6 are compared and the results are shown in Figure 4.23. When conducting localization with sensor group 1, 2, and 3, the average error is 1.82m, 1.04m, 1.38m respectively, with standard deviations of 1.60m, 0.63m, 1.02m. The detecting rates vary over the three grouping scheme and are respectively 95%, 100%, and 83%. The sensor grouping based on our algorithm achieves the lowest average localization error of 0.83m with a standard deviation of 0.6m and a detecting rate of 91%. When all sensors are used for localization, the detecting rate drops to 73%, and an average localization error of 0.95m with a standard deviation of 0.78m. This indicates that the structural-based sensor grouping assists structural vibration based human sensing applications to achieve a higher accuracy in information acquisition.

4.6   Related Work

In this section, the two main aspects explored to improve the signal quality for more accurate information learning are 1) increase the sensing signal fidelity, and 2) group the sensors for collaboratively information acquisition based on the structural effects on the perceived signals. The related work on these two aspects will be introduced in this section, and the gaps filled by the work introduced in this thesis will be discussed.

4.6.1 Improve Signal Quality

Prior works that focus on improving sensing signal quality mainly fall into three categories: 1) utilizing expensive enhanced sensors [15], 2) post-processing to restore signal shape [55, 65, 89], and 3) adaptive hardware settings to obtain high fidelity signals [76, 165]. The cost of **enhancing** sensing device to achieve high dynamic sensing range as well as high resolution could make large-scale deployment unrealistic. Previous methods for obtaining high-fidelity sensing data mainly fall into two categories: post- and pre-processing.

Post-processing methods restore unknown or lost data after data collection [55, 65, 89]. These methods are usually used for audio data and evaluated by the signal-to-noise ratio (SNR). Janssen et al. proposed an adaptive interpolation method to restore lost data, with the restrictions that the positions of the unknown samples are known [55]. Miura and his group introduced their clipping removal method through recursive vector projection [89]. Kitic et al. approached the problem from another perspective with iterative hard thresholding and evaluated the results using both signal-to-noise ratio and human listening [65]. However, for those feature-oriented applications such as identification [104] or TDoA-based localization [31], restored data is not dependable enough since it introduces signal artifacts.

Pre-processing methods utilize signal processing techniques to predict signal clipping and limit distortion of an amplified signal [66]. In addition, Zhang et al. proposed the robust taking pressure control (RPC) algorithm to adjust the system sensing configuration for better signal collection [165]. For pedestrian induced excitation, the rapid change and variation make it difficult if not impossible to achieve high fidelity with those methods.

4.6.2 Sensor Grouping/Selection

Various sensor placement algorithms has been explored, including entropy-based method [149], uncertainty-based [46], etc. On the other hand, it is difficult if not impossible to measure and estimate the optimal sensor placement before deployment, because fingerprint-based structural characterization is extremely labor intensive. In addition, for structural vibration sensing systems, once they are deployed, it is difficult to retrofit [116]. Therefore, for the deployed system, different sensor selection

methods have been explored to achieve various sensing purposes.

Prior work on sensor selection scheme focus on various goals, including coverage [126], task assignment [126], detecting performance [63], tracking and localization [40], etc. However, none of these schemes target at the physical information media characteristics, which, in our case, the structural factors. Therefore, a specific structure-based sensor grouping or selection scheme is needed, and to the best of our knowledge, this is the first work to group sensor based on their placements' structural characteristics.

4.7 Chapter Summary

In this chapter, a methodology that utilizes physical insights/measurements on both human-induced structural vibration signal and ambient noise signals to enhance the sensing configuration for more accurate information acquisition is introduced. Two specific methods that follow this methodology are explained with the real world experiments on pedestrian tracking application to demonstrate the importance of obtaining 'good signals'. The physical insights that are used in this chapter are three-fold: 1) the footstep induced vibration signal decay model, 2) the continuity of the pedestrian walking trajectory, and 3) the detection of the structural changes/factors that alter signal decay models.

The high fidelity footstep induced structural vibration acquisition sensing system improves the signal quality by predicting the upcoming Step Event's amplification as well as the optimal amplifier configuration and adapts the hardware settings during the sensing process accordingly. The prediction is achieved through two key aspects: 1) each individual sensor predict the step strength change based on a pedestrian walking model (LPP), 2) networked devices collaboratively predict the step strength through a global profile on a structural variation model (GPP). In the pedestrian footstep monitoring application, the presented system demonstrated up to 2× increase on SRR in the evaluation experiments and up to 2× less error rate when used to locate the pedestrian walking along the hallway. This signal acquisition system can be applied to various future applications in smart buildings for human activity induced excitation vibration data acquisition.

The structural signal grouping algorithm is based on structural variations. The

algorithm utilizes wavelet decomposition to analyze the ambient vibration signal perceived at different locations of a structure. Then it groups the sensors based on the scale energy profile of each sensing location. Our grouping scheme demonstrates up to a 2.86× reduction in average localization errors in the real world experiments conducted in two campus buildings.

Chapter 5

Human-Induced Structural Vibration Signal Characterization

Signal characterization is to analyze the signal with various signal processing methods to obtain features that describes the characteristics of the signal. It is a process of abstracting the information from the signal in a designated form of features. These extracted features from the processed signal are often served as input in the information learning module. For example, if the learned information is the number of people walking together, the signal characterization extracts features that describe the person's footstep overlapping [99]. On the other hand, if the learned information is the pedestrian identity, the signal characterization extracts the gait variation instead of focus on the overlapping traits [104, 106]. If the learned information is step-level pedestrian location, the signal characterization focuses on the Time Difference of Arrival (TDoA) between sensor pairs and the estimation of the wave propagation velocity [86, 103].

The accuracy of characterizing the signal affects the learning results for the application. However, due to the complication of the physical world, accurate signal characterization for various applications becomes a challenge. The method discussed in this chapter aims to improve the signal characterization based on the physical properties of the wave and is evaluated through a real-world application of excitation tracking. Since the system this thesis focuses on is the structural vibration sensing system for indoor human information acquisition, I will first introduce the wave propagation properties and dispersion effects (Section 5.1) in this Chapter. Next,

I will present the methodology to utilize physical insights – wave properties and human behavior – to guide the signal characterization process (Section 5.2). The evaluation of two applications that utilizes the aforementioned methodology will then be discussed in Section 5.3 and Section 5.4.

5.1   Wave Properties

When waves propagate in solids, they disperse and attenuate. These effects distort the waveform of the vibration signal during their propagation and introduce challenges to the characterization of the signals. There are mainly two types of excitation, the impulse, and slip-pulse. The impulse and slip-pulse excitation on structures produce different mechanical waves. The former induces surface waves due to point impact force while the latter produces body waves due to friction. Different waves have different degrees of *attenuation* and *dispersion*. In addition, when multiple vibration signals are overlapping, they form a superposition signal.

5.1.1   Wave Generation: Impulse v.s. Slip-Pulse

When an object contacts a surface at a single point, such as a foot striking on a floor or a pen tapping on a table, this impact interaction induce vibrations [14]. The force applied to the surface of a structure causes it to deform. As the contact point is relieved of the force, the structure surface retracts due to its elasticity. This elasticity generates surface waves propagating outward from the point of contact, similar to ripples generated in water when a stone is dropped [14].

On the other hand, when two objects slide against each other, such as dragging a chair through a room, swiping a finger on a table, the friction between two objects contacting surface will induce vibration. Stick-slip is a general form of friction that induces vibrations [2, 71], such as a chair dragged on a floor or a pen swiping on a table. When an object slides on a structure, it will 'stick' because of static friction and the unevenness of the structural material, then the force applied to it causes it to overcome the static friction and slide or 'slip' [2, 112]. When these two states are alternating, the friction between the object and the structure changes between static friction and kinetic friction [112]. Since typically static friction is larger than kinetic friction, such alternating friction causes a sudden jump in the velocity of the

66

movement, resulting in slip pulse [2, 112] along the swipe. These slip pulses induce a wave that travels at an angle in the material as a combination of different types of waves dominated by body wave [2].

5.1.2 Attenuation

When the wave travels through a medium, it will gradually lose the flux intensity, which is referred to as the attenuation. For an impact-induced vibration (impulse) that is dominated by surface waves (Rayleigh-Lamb waves), its attenuation rate is $\propto r^{-1/2}$) [146, 147]. On the other hand, for a friction-induced vibration (slip-pulse) that is dominated by body waves (shear wave), its attenuation rate is $\propto r^{-1}$) [146, 159]. This is caused by different particle movement of these waves: when the Rayleigh-Lamb wave propagates, it induces the particles in the medium to move in circles; while when the shear wave propagates, it induces the particles in the medium to move in line (up and down) [147, 159].

These wave attenuation characteristics are the physical insights that guide the signal characterization that will be introduced later in this Chapter. For example, since the friction-induced vibrations decay faster than impulse-induced vibrations and thus have less reflection on edges, it is possible to use the entire windowed signal to estimate Time Difference of Arrival. While for the impulse-induced vibration that demonstrates higher reflection on edges, only the onset of the signal can be used to avoid the heavy reflection signals that overlap with the original signal. The detailed examples and the algorithm will be further introduced in Section 5.2 later.

5.1.3 Dispersion

The frequency dispersion of the wave indicates the phenomena where the different frequency components travel at different velocities [147]. As consequences, when an excitation occurs and induces the structure to vibrate, the sensors at different distances from the excitation will detect varying waveform due to the dispersion effect.

For Rayleigh-Lamb waves, different frequency components travel at different speeds in the range of 50 to 300 ms^{-1} [87, 109, 147]. Therefore, when the wave travels through the solid, the dispersion introduces distortion into the vibration

signal, making vibration-based localization a challenge [87]. On the other hand, for shear waves, the attenuation rates are higher and most of the frequency components decay before reaching the sensor. Therefore, less dispersion effect is observed in the shear waves compared to that in the Rayleigh-Lamb waves. This leads to different filter strategies for these two different waves: for Rayleigh-Lamb waves, a wavelet filter can be applied to reduce the dispersion effect before TDoA estimation; while for shear waves, the cross-correlation can be directly applied for TDoA estimation considering the requirement of continuous windowed signal calculation.

5.1.4 Superposition

When two impulsive waves overlap, they superposition with each other. Figure 5.1 shows the floor vibration induced by two tennis ball dropping on the floor detected by two sensors placed 5ft away from each other. The blue lines are the signal when two balls are dropped at the same time and their mixed signals are perceived by each sensor. The red and yellow lines are the signals when each ball are dropped alone at the same dropping location and perceived by the two sensors. The superposition of the signals detected by different sensors shows different delay and decay for each
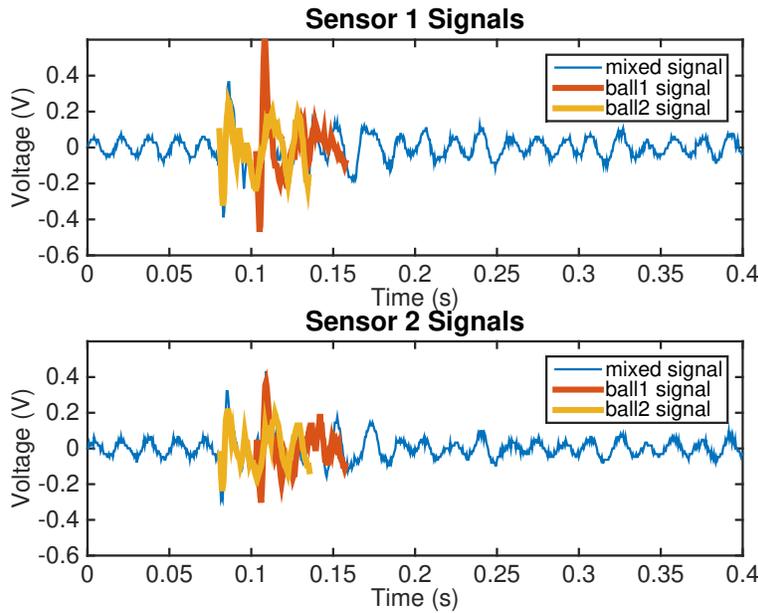
Figure 5.1: Ball-drop signals overlapping and decomposition

68

single ball drop signal. These observations will be used for feature design later (Section 5.2.2.2).

5.2 Physical Properties Guided Signal Characterization

The wave properties discussed in Section 5.1 can be applied to assist the signal characterization. Since different wave properties cause different perceived waveform properties, the information acquisition methods should vary taking these properties into consideration. In this section, I will introduce two types of characterization guided by physical properties of waves: 1) the pairwise Time Difference of Arrival (Section 5.2.1), and 2) signal features for pedestrian counting, as examples to demonstrate the advantage of utilizing physical properties.

5.2.1 Wave Property Guided TDoA Estimation

Human activities induce excitation on the ambient structure (e.g., floors, walls, tables) when performing activities. Tracking the source location of the excitation enables further inference of the human activity information. This information can be used in many smart building applications, such as kid/patient monitoring, customer behavior analysis, ubiquitous user input. The structural vibration-based sensing, compared to other methods including vision-, RF-, mobile-, acoustic-, and load-based sensing, enables sparse and non-intrusive sensing on people [51, 96, 119, 129, 143, 167]. The excitation mainly falls into two categories: impact induced impulse excitation (such as footstep, object dropping, door closing, tapping) and friction induced slip-pulse excitation (such as drag a chair, swipe on a surface) [14]. When there are multiple sensors in the sensing environments, these excitation caused structural vibrations travel to different sensors at different times, and the difference between the arrival time (TDoA) of pairwise sensors can then be used to estimate the excitation locations via multilateration. Therefore, the accurate TDoA estimation leads to accurate location estimation.

5.2.1.1 Time Difference of Arrival (TDoA) Estimation

Various methods have been explored to calculate TDoA values. The mainstream methods include time domain cross-correlation, frequency domain cross-correlation, and peak-based method, which is listed as follows.

- **Time Domain Cross-correlation.**

  When two time domain discrete signals are perceived by two sensors at different locations, e.g., $s_1[N]$ and $s_2[N]$, the cross-correlation of these two signals is defined as $(s_1 \star s_2)[n] = \sum_{m=-\infty}^{\infty} f^*[m]g[m+n]$ [121]. When applied on the same signal with different delay, the position of the cross correlation results with the highest value indicates the shift of the signals, which can be represented as $\tau = \operatorname*{argmax}_{n}(s_1 \star s_2)[n]$[121]. Therefore, for signals that have only delay effect and no attenuation, by applying the time domain cross-correlation, we can find the signal shift by locating the highest peak in the cross-correlation result.

- **GCC-PHAT.**

  When the target signal is collected under an echo-rich environment, the generalized cross-correlation with phase transform (GCC-PHAT) is often used [18, 140]. To calculate the generalized cross-correlation, the algorithm first calculates the Fourier Transform of the two signals, $S_1(f)$ and $S_2(f)$. Then the GCC-PHAT is defined as

  $$\frac{S_1(f)[S_2(f)]^*}{|S_1(f)[S_2(f)]^*|} \tag{5.1}$$

  and the TDoA is estimated as

  $$\tau_{PHAT} = \operatorname*{argmax}_{n} IFFT(\frac{S_1(f)[S_2(f)]^*}{|S_1(f)[S_2(f)]^*|}) \tag{5.2}$$

  Note that the weight of each signal is $\frac{1}{|S_1(f)|}$ and $\frac{1}{|S_2(f)**|}$ respectively, which indicates that the signal strength/decay effect is removed by the normalization of the signal magnitude.

- **Peak-Based Detection.**

  When the similarity between the two signals is low, due to various distortion effects on the vibration signals, the peak-based methods can be used to calculate
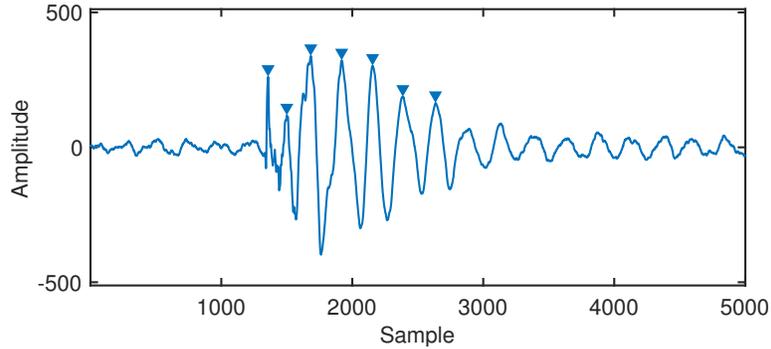
Figure 5.2: Example of footstep induced vibration and its peaks.

the shift of the signals τ_{peak} by selecting a particular peak within the extracted footstep signal. For each footstep signal, the peak detection algorithm is applied to the extracted footstep signals Figure 5.2 demonstrates an example of peaks detected from a footstep signal.

5.2.1.2 Physical Property: Wave Attenuation and Dispersion

The two main factors that distort the vibration signal through its propagation are discussed in Section 5.1: the attenuation and the dispersion. When the target signal travels through different distances and reaches different sensors, the attenuation rate of the signal varies. Furthermore, since the dispersion effect is also affected by the wave propagation distance, the signal arrives the different sensors will show different waveforms despite the amplitude variation.

When dispersion happens, it is difficult to estimate TDoA from either cross-correlation or peak detection based methods. This is because, for the cross-correlation based method, the signals do not have high similarity in terms of waveform, even they are normalized by energy (remove the variation of attenuation). On the other hand, since the waveform varies, the location and shape of the peaks vary. Therefore, the peaks for different signals may representing different frequency component, which is not comparable for signal shift estimation.

Different type of interaction between two objects causes different types of waves. For different types of waves, the dispersion effects, as well as time domain characteristics of the signals, vary. There are mainly two types of waves induced by human activities: impact induced impulsive signal and friction induced slip-pulse like

71

signal. These types of vibration signals perceived by sensors may be most suitable being calculated and processed by different TDoA estimation methods, which I will introduce respectively in Section 5.2.1.3 and 5.2.1.4.

5.2.1.3 Impacts Induce Impulse-like Vibration Signals

When an impulsive excitation occurs on a structure, the contact happens on a single point. The TDoA based localization of this contacting point is useful for various location-based applications. However, as discussed in Section 5.1, the dispersion effects make the TDoA estimation a challenge. Therefore, to obtain accurate TDoA for impulsive excitation that generates waves dominated by Rayleigh-Lamb waves, the system first decomposes the vibration signal and filters based on the observed structural character using wavelet analysis. Then the TDoA is calculated based on the detected first peaks of the signals from different sensors.

5.2.1.3.1 Impulsive Signal Decomposition

For impulsive signals, wavelet-based decomposition is an effective way to analyze the frequency components and to filter the signal. This is because that the mother wavelet has high similarity with the shape of the fast-decaying oscillating waveform [87]. Figure 5.3 (a) demonstrates an example of impulse excitation signals obtained by two different sensors. The original raw signal shows clear signal characteristics difference caused by dispersion effects, including the peak and valley location variation. The cross-correlation peak value for these two signals is 0.62, indicating a high variation of the compared signals.

5.2.1.3.2 Structure-Based Signal Filtering

The raw signals are first decomposed with wavelet transform and then the filtered signals are reconstructed on a specific scale. The selection of the reconstruction scale determines the filtering frequency band. We conduct this signal decomposition and show the filtered signal in Figure 5.3 (b). Compared to the raw signal in Figure 5.3 (a), the filtered signal demonstrate a higher similarity. The cross-correlation peak value for filtered signals is 0.84, showing a clear increase in signal similarity after the filtering is applied. Therefore, the wavelet filtering reduces the dispersion effects in the vibration signals significantly.
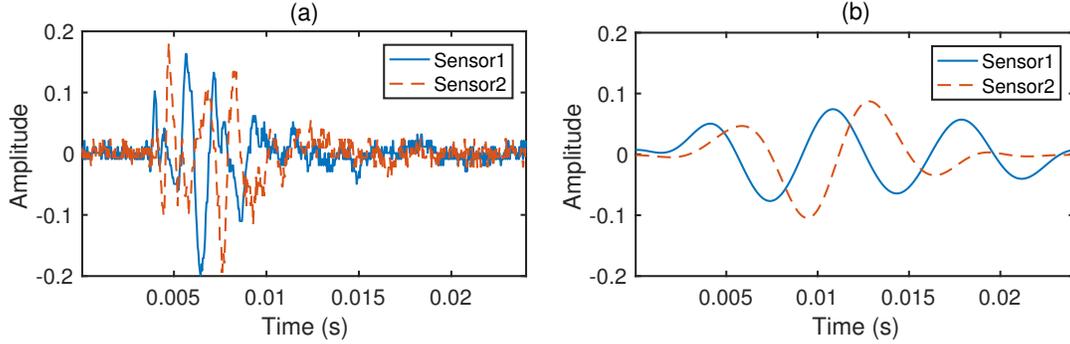
Figure 5.3: Examples of impulse-like excitation signals obtained by different sensors before and after wavelet filtering.

The filtering scale selection can be done in different ways, and two particular ones will be introduced here in this thesis: 1) The ambient noise signal is used to select the scale. For the noise signal decomposition coefficients, the signal energy values of each scale are calculated. The scale with the highest signal energy value is then selected as the filtering band. This gives us the fundamental frequency band for the floor structure, on which the signal travels farthest. 2) We conduct a low-cost calibration by generating impulse signals at known locations (e.g., between each pair of sensors) for each sensing area, and select the scale and velocity that cause the least error at these designated points. This gives us the optimal parameters for the targeting area despite the material and the size of the surface.

The former method is applied to the human footstep tracking application and the later one is applied to the interaction tracking application. The human footstep tracking application usually requires a larger scale deployment than that of the human interaction application. Therefore, the ambient noise analysis based labor-free scale selection is a better fit in that case. On the other hand, calibration on a small human interaction area is easy to conduct, and the improvement on accuracy is significant and important for interaction purpose.

5.2.1.3.3   Structure-Based TDoA Estimation

As discussed in aforementioned Section 5.2.1.1, various TDoA estimation methods have been explored for different deployment details [87, 103, 140]. The GCC-PHAT methods are used for acoustic signals under echo-rich environment [140]. For the

human footstep tracking application, since the sensing area is relatively large, the reflection signal dies down before overlapping with the original signals. Therefore, the entire impulse signal can be used as a reference to estimate TDoA via cross-correlation [86]. For the human interaction tracking application, since the sensing area is relatively small and the reflection signal may affect the original signal, the first peak is used since the first peak is less likely to be impacted by reflections [103, 109].

5.2.1.4   Frictions Induce Slip-pulse-like Vibration Signals

When a slip-pulse signal occurs on the structure surface, there is a sequence of contact points forming a trajectory. Based on the wave properties comparison in Section 5.1, we know slip-pulse signals have higher attenuation rate and the frequency components are more concentrated. It means that the dispersion effect is lower than that of the impulsive signals. On the other hand, this sequence of contact points need to be localized consecutively, therefore filtering may cause additional computational load that makes the real-time processing difficult. With that in mind, the TDoA estimation is done with the unfiltered signals.

Figure 5.4 (a) shows an example of slip-pulse signals detected by two synchronized sensors. Figure 5.4 (b, c, d) show three segments selected at the fore-end, middle, and tail-end part of the signal marked out in Figure 5.4 (a) with black lines. Figure 5.4 (e) shows the frequency components of the initial tap signal and (f) shows the frequency components of the signal segment in (d). Compared in Figure 5.4 (e, f), slip-pulses show more concentrated frequency than that of the impulses. This is caused by the higher attenuation rate of the body wave, so that when the wave arrives at the sensor, most of the frequencies already die down, leaving fewer frequency components, hinting at lower dispersion.

In addition, the reflection may also be lower than that of impulses due to the higher attenuation rate. With less dispersion and reflection, correlation is visible between segments from different sensors without the need for wavelet analysis as shown in Figure 5.4 (b, c, d). Unlike impulse excitation, there is no 'first peak' for a segment of a slip-pulse signal. Therefore, we extract TDoA using cross-correlation from multiple consecutive slip-pulses and localize each of them. Their locations make the trajectory of the slip-pulse excitation.
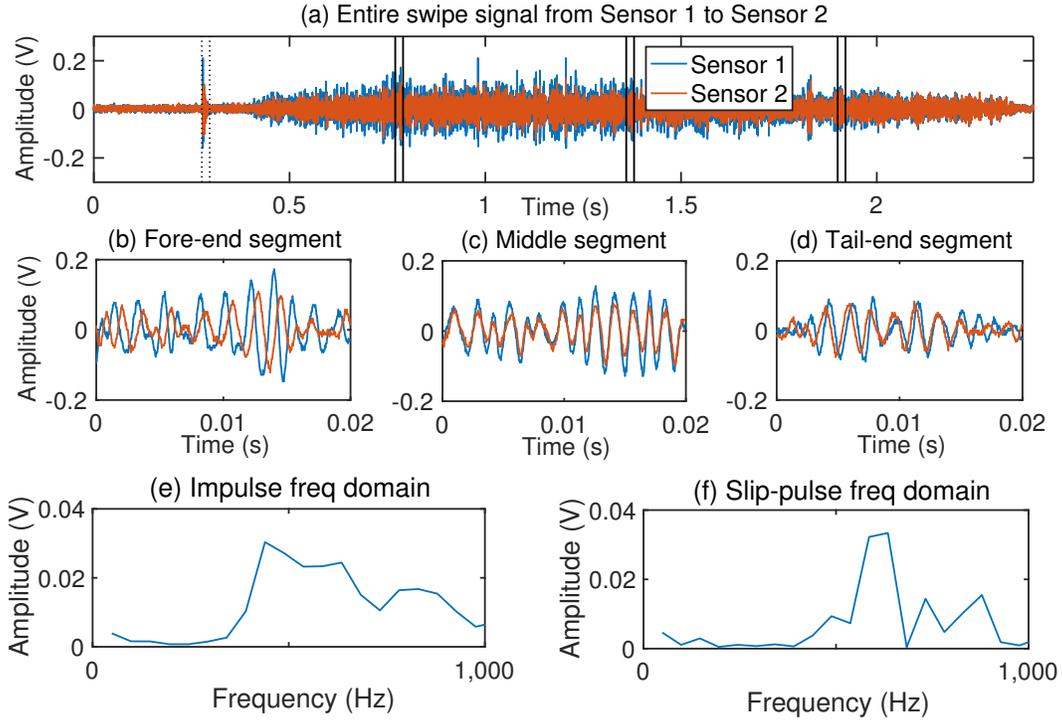
Figure 5.4: Impulse and slip-pulse signal TDoA progression.

5.2.2 Wave Property Guided Feature Extraction

Other than obtaining TDoA values from sensor pairs, the wave properties, can also be used to extract features for classification problems. In this section, an example of obtaining the number of people passing through different indoor areas is used to explain the importance of the wave property guided feature extraction.

This information can be used in various smart structure applications, including occupancy-based building energy/space management, marketing research, security, etc. To obtain this information through structural vibration induced by footsteps is challenging. The main challenge lies in distinguishing multiple simultaneous walkers by developing features that can effectively represent the overlapping signals. The intuition that these different traffic conditions can show distinguishable features are two-fold: 1) human behavior has its randomness, therefore when the pedestrian traffic contains more people, this randomness – chaos – increases (Section 5.2.2.1); 2) when there are multiple excitation sources and multiple sensors, the mixture of the signals detected at each sensor are the superposition of signals of each excitation

after particular decay (Section 5.1.4).

The signal characterization is conducted on the signal extracted that contains human footstep induced floor vibration, which is referred to as the Signals of Interest (SoI) (Section 4.2.2). Then the selected features for the occupant traffic estimation are extracted from these SoIs for further occupant number estimation (Section 5.2.2.2). These features reflect the signal characteristics under different occupant traffic conditions.

5.2.2.1 Human Behavior: Group & Randomness

Studies have been done on interpersonal distance in group walking. When multiple people walk as a group, they tend to fragment in smaller units of one to three members [26]. Therefore, fine-grained occupant traffic estimation for one to three people in a walking group is critical for a larger group of occupant traffic estimation. Different occupant traffic conditions cause different feature characteristics. Due to the randomness in human motions, these features are distinguishable under different traffic conditions.

5.2.2.2 Feature Selection for Occupant Counting

Four features are extracted from each SoI extracted (Section 4.2.2) for further estimation of the pedestrian count: 1) cross-correlation between signals collected by sensors at different locations induced by the same footstep, 2) cross-correlation between signals induced by consecutive footsteps collected by the same sensor, 3) SoI duration, and 4) SoI entropy. These features are selected based on 1) the understanding of the structural vibration signal attenuation [97], 2) the assumption of linear addition mixture model for structural vibrations [70], and 3) the understanding of the human gait consistency[104]. These features and their underlying principles will be explained in the rest of the section.

- **Cross-correlation between SoIs from Different Sensors for the Same Footsteps.** The intuition of this feature is that when different number of signal overlap, their signal superposition at sensor of different locations vary. The normalized cross-correlation between SoIs from different sensors for the same footsteps, which later we refer as **spatio-different SoIs**, reflects the divergence of the signal attenuation observed by those sensors.

76

Therefore, the attenuation of each footstep signal at different sensors is different. Due to the linear addition assumption, when these signals add up, the similarity between mixed signals observed by different sensors will be low. Therefore, the cross-correlation between the spatially different SoIs can be used for inferring the occupant count.

Assume that there are n impulses, $S_1...S_n$ happening at the approximately same time, and the observations at two sensors are

$$
\begin{cases}
O_1 = a_{1,1}S_1 + a_{1,2}S_2 + ... + a_{1,n}S_n \\
O_2 = a_{2,1}S_1 + a_{2,2}S_2 + ... + a_{2,n}S_n
\end{cases}
$$

where $O_i, i = 1, 2$ are signals observed by sensors, $S_j, j = 1, 2, ..., n$ are signal impulse sources (i.e., footsteps), and $a_{i,j}, i = 1, 2; j = 1, 2, ..n$ is the signal decay coefficient. When the **number of impulses** n increases, there is a higher chance that $a_{1,i} \neq a_{2,i}$, where $i = 1, 2, ...n$. This causes signals observed by different sensors to exhibit less similarity, lowering normalized cross-correlation values. Therefore, the relation between $a_{1,i}$ and $a_{2,i}$ is also determined by the **impulse location** relative to different sensors.

- **Cross-correlation between SoIs for Consecutive Footsteps from the Same Sensor.** The intuition of this figure is that when multiple people walk at the same time, the superposition of the signals overlapping has a lower consistency than there is no overlapping. The normalized cross-correlation between different SoIs of the same sensor, which later we refer to as **temporal-different SoIs** reflects the consistency of waveforms, discarding the attenuation factor, induced by footsteps/impulses from a continuous footstep sequence (within a trace) and obtained by one sensing unit.

  When one person walks by a sensor, their footstep induced vibration signals are consistent (high cross-correlation value) due to human gait consistency [104]. However, when multiple people pass by, the inconsistent temporal offset between mixed signals and the inconsistent signal attenuation rate can break such consistency between the footstep signals.

  Therefore, the cross-correlation between the temporal-different SoIs can be used to infer the number of occupants. For each trace, the detected SoIs are

compared pairwise (i.e., left foot signals are compared to left foot signals, same for the right foot signals). Since the SoIs can be of different length, for this feature, we find the highest peak within each SoI and then redefine the SoI as the signal with the specific length before and after the peak.

- **SoI Duration.**

  SoIs are the extracted signal segments that contain multiple consecutive sliding windows of signals whose signal energy is an anomaly compared to sliding windows of noise signals. Different paces or number of people can cause overlapping signals, resulting in elongated event duration and increase in SoI duration variation, as indicated by the average and standard deviation. With the SoI detection algorithm introduced in Section 4.2.2 which extracts a SoI through multiple sliding windows, we define SoI duration as the length of these sliding windows added up, which means the sliding window size determines the SoI duration resolution.

- **SoI Signal Entropy.**

  Signal randomness is another feature we explore. The randomness can be quantified as the degree of order/disorder associated with a multi-frequency signal response [125], which is carried by the wavelet entropy

  $$E(s) = -\sum_i s_i^2 \log(s_i^2)$$

  where s is the signal and s_i is the relative wavelet energy of the signal at resolution level i using orthogonal discrete wavelet transform of s.

  The noise signal is expected to be most random (i.e., high entropy). When there is only one impulse source, the randomness of detected signal will be low. However, when the number of mixing footstep induced signals is large enough, their distribution addition will converge to the Gaussian distribution by central limit theorem (CLT), which is used to model the noise. Therefore, we expect to see higher entropy when the number of mixing impulse increases.

5.3 Evaluation I: TDoA Based Tracking

To evaluate the wave property guided signal TDoA estimation, the localization/-tracking application for human activity induced excitation is used as the application example. In the rest of the section, I will first introduce the implementation of the tracking system (Section 5.3.1), then I will explain the two experiments on impulse signal localization (Section 5.3.3) and the slip-pulse signal tracking (Section 5.3.3.2).

5.3.1 Implementation

The overview of the TDoA estimation (signal characterization) guided by physical properties implementation is shown in Figure 5.5. From the human aspect, the physical properties used is the type of waves generated by different interactions, and it determines the filtering as well as the TDoA estimation algorithm. From the structure aspect, the physical properties used is the material variation caused dispersion/velocity difference, and it determines parameters of the filtering method in the signal characterization module. The wave properties of different types of excitation vary [100, 103]. The characterization presented in this section takes the excitation type into account and selects different characterization processes based on the wave types. Therefore, the system first needs to identify the type of an excitation.

Intuitively, the impulse excitation decays fast within a short duration due to the dissipation of the energy, while the slip-pulse excitation often lasts over seconds with continuous high signal energy due to the overlapping of multiple stick-slip iterations during the movement [103]. The system identifies an event to be a slip-pulse if the segments above the threshold last over an empirical threshold (one second). Most of human-induced impulse signals on various testing surfaces do not contain segments over a second.

This identification of excitation signal types enables the selection of signal processing methods to achieve highly accurate signal source tracking. Once the system obtains the excitation signal and their types, it tracks the excitation accordingly. This physical property guided signal characterization is evaluated through two applications: human activity tracking and human interaction tracking. The experimental setting implemented is as shown in Figure 5.6.

79
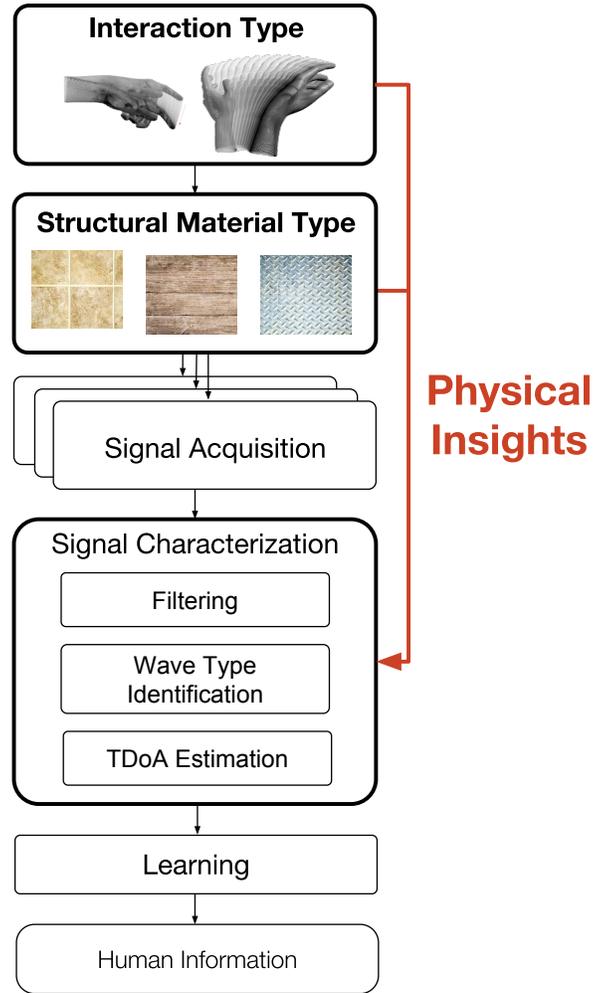
Figure 5.5: Human excitation induced vibration signal characterization system overview.

5.3.1.1   Application I: Human Activity Tracking

The excitation investigated in this application include footsteps falling (impulse) and object dragging (slip-pulse).

The experiment is conducted in a room of wood floor with sensors placed as shown in Figure 5.6 (a). Four sensors are deployed along a 4m × 3m sensing area and illustrated as circles in Figure 5.6 (a). They are placed to cover as large sensing area as possible. The crosses in Figure 5.6 (a) indicate the footstep locations. The number of footstep excitation collected at each location is 10.

For object dragging excitations, metal bars of different sizes and weights are

used. The experiments are conducted on a 60cm × 60cm wood platform as shown in Figure 5.6 (b). The wood platform is selected to avoid potential scratch damage to the floor surface. Metal bars are used to represent different types of objects that may be dragged through a floor, such as tables and chairs. Different sizes of metal bars represent different weights of slip-pulse signals.

5.3.1.2 Application II: Human Interaction Tracking

The experiments are conducted on different sample materials placed on a large enough floor. The excitation investigated here include tap (impulse) and swipe (slip-pulse). For each scenario, the evaluated surface is placed at the same location of the room for consistency. Since the sample board is not coupled with the floor like these materials will appear in people's everyday life (e.g., as part of a table/wall), sand-filled weights (20cm × 5cm × 2cm) are used. They are placed around the sensing area in some experiments to enhance the coupling and damp the wave reflection at the loose edges to represent real-world table/wall conditions. A pen is used to tap and swipe for consistency.

Compared to the human activity tracking experiment, the investigated sensing area in this case is in general smaller (< 1m × 1m area) and the force applied on the structure is lower. On the other hand, a short distance between sensor and excitation does cause a higher signal energy. Due to the contacting point difference, the accuracy of human interaction tracking application is higher than that of the human activity tracking.
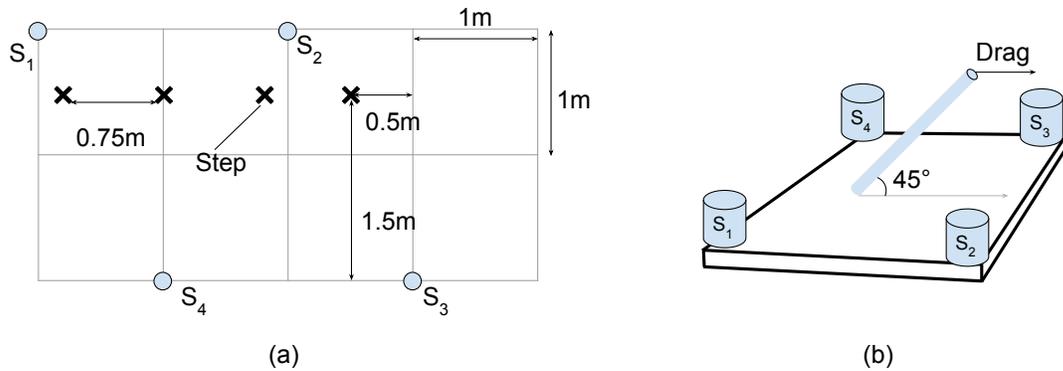
Figure 5.6: Human activity induced excitation experiment setups [100].
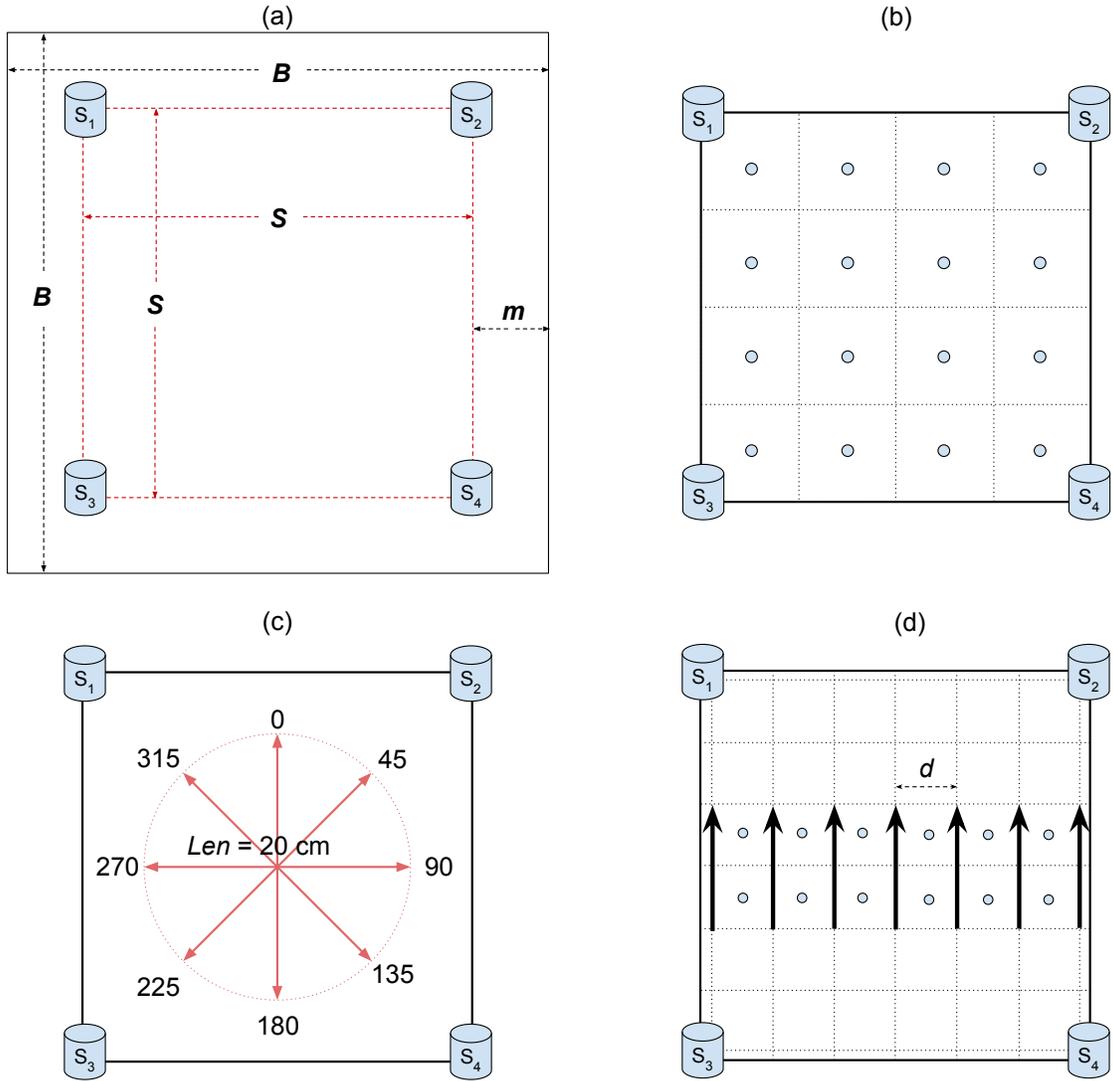
Figure 5.7: Human interaction tracking experiment setup [100].

5.3.1.3 General System Settings

For both experiments, the data is collected at a sampling rate of 25 kHz to provide 1cm resolution for up to 250 ms^{-1} waves when used to estimate TDoA. To generating consistent slip-pulse signals, the object moving speed is controlled by a metronome at a speed of 10cms^{-1}. The sliding window size used in these experiments is 0.02s to cover a 0.2cm distance. This selected size is small enough to realize the resolution limitation of the system. The number of incidences for each scenario investigated is 10.

5.3.2   Metrics and Parameters

Two metrics are selected to evaluate the impulse localization and slip-pulse tracking estimation. For each metric, both the accuracy (error) and precision are evaluated. The error and precision of impulse and slip-pulse excitation are different as follows.

Impulse Error. For impulses, the distance between estimated and ground truth locations is measured as the localization error. The precision is measured by the scattering radius of the estimated location with the corresponding ground truth location.

Slip-pulse Error. For slip-pulses, length and angle errors of the estimated trajectory are measured and compared to that of the ground truth. In this thesis, linear trajectories are investigated, therefore the length and angle errors are sufficient to describe the signal.

The length error is positive when the estimation length is longer than that of the ground truth, and negative when shorter. As for the angle error, the clockwise errors are considered as positive values, and the counter clock wise errors as negative values. Based on this definition, the standard deviation of the error would be a more effective metric to describe the accuracy here.

The precision indicates the accuracy of the tracking in following the trajectory. It is measured by the distance between estimated location of each segment and the trajectory, which will be referred to as the 'trajectory error'.

Various of parameters may affect the system performance, which mainly fall into three categories: 1) the implementation factors, 2) the excitation factors, and 3) the structural factors. These factors include structural material, structure size, sensing area size (distance between sensor pairs), excitation location on the sensing area, excitation strength, excitation type, etc. The experiments taking these factors into account are summarized in Table 5.1.

Application I targets at building or room scale detection. Application II targets at small scaled flat solid surfaces (table or tablet scale). Therefore, for Application II, various sampling boards are experimented on due to the feasibility. This allows us to understand system robustness through evaluations on 1) different materials, 2) surface sizes, 3) sensor deployment distances (sensing area margin size), and 4) different areas on the board.

Five surface materials are investigated, including wood, iron, cement, stone, and

App	Structure	Type	Sec.	Excitation	Variable
I	floor	impulse	5.3.3.1	footstep	impulse location
II	sample	impulse	5.3.3.2	tap	structure material
II	sample	impulse	5.3.3.2	tap	structure size
II	sample	impulse	5.3.3.2	tap	distance between sensors
II	sample	impulse	5.3.3.2	tap	sensing area
I	sample	slip-pulse	5.3.4.1	drag	object weight (strength)
II	sample	slip-pulse	5.3.4.2.1	swipe	window size pointer type
II	sample	slip-pulse	5.3.4.2.2	swipe	swipe length swipe speed swipe direction
II	sample	slip-pulse	5.3.4.2.3	swipe	structure material structure size distance between sensors sensing area

Table 5.1: Human excitation tracking experiments summary [100].

ceramic. These surfaces were chosen as a representation of materials and surfaces found in a person's everyday life, such as tables, cabinets, and walls. For the first four materials, we find samples of size 61 cm × 61 cm, therefore we set $B = 61$cm, $S = 40$cm shown in Figure 5.7 (a) as the baseline. The ceramic surface sample used here is 41cm × 41cm as it is the only size available.

To evaluate the sensor deployment distance, board margin size m is first set as $m = 1$cm as shown in Figure 5.7 (a). Then different m values are studied, from 1cm to 30 cm, with intermediate values of 10 cm and 20 cm. The experiment is done on wood surfaces only, since it is easier to cut particle wood board into these sizes.

For the largest surface ($B = 101$cm), the sensing area value S is investigated from 40cm to 80cm with incremental 10cm. For each case, the tap and swipes are at the same location as shown in Figure 5.7 (b, c) within the central 40cm × 40cm area of the board.

Furthermore, to evaluate effects on different parts of sensing area, taps and swipes are evaluated at locations shown in Figure 5.7 (d). To ensure the consistency of other parameters, this experiment is only done on the largest surface, where $B = 101$cm and $S = 80$cm.

5.3.3 Experiment I: Impulse Excitation Localization

The challenge faced by impulse excitation localization through surface vibrations lies in the dispersion of wave propagation. We address that by studying properties of the impulse-induced surface wave and utilizing a wavelet filter to extract the designated scale of signals and reduce the dispersion effects. To evaluate our localization algorithm, we compare localization error rates when different methods are applied.

The footstep localization experiments evaluate the system performance in a specific structural condition (the floor in a room) and of a larger area. The tap localization experiments evaluate the 'micro' area (a sample surface) where we can further evaluate on different materials and surface sizes. The goal is to evaluate the performance of our system handling dispersion and attenuation effects. Therefore, in the example sections (Section 5.3.3.1, 5.3.3.2), we compare the localization accuracy with v.s. without applying the wavelet filter.

5.3.3.1 Impulse Example I: Footstep

To understand the performance of the signal characterization algorithms with or without taking wave properties into account, footstep localization results under these two algorithms. The filter scale is selected based on ambient noise vibrations. The system conducts wavelet decomposition on the ambient noise signal and selects the scale with highest signal energy. Because this scale corresponds to building fundamental frequencies, on which the vibration signal travels the farthest [86].

Figure 5.8 (a) shows the accuracy of estimated locations of four footstep demonstrated in Figure 5.6 with boxplots [1]. The light box plot shows localization error without applying wavelet filter and the dark box plot shows that with wavelet filter applied. The average location estimation error without filtering is 2.18m, with the standard deviation of 1.88m and median of 0.91m. This value drops to 0.49m, with the standard deviation of 0.17m and median of 0.46m, when the structure-based decomposition is conducted and wavelet filter is applied. Figure 5.8 (b) shows the location estimation precision. The average precision values are 0.17m and 0.07m, with the median values of 0.13m and 0.07m for the cases without and with wavelet filtering respectively. In summary, the average localization error rate decreased 4×,

---

[1]A box plot shows the mean (circle), median (middle line), quartiles (rectangle box), fences (outside box line), and outliers (cross markers) of the data points.
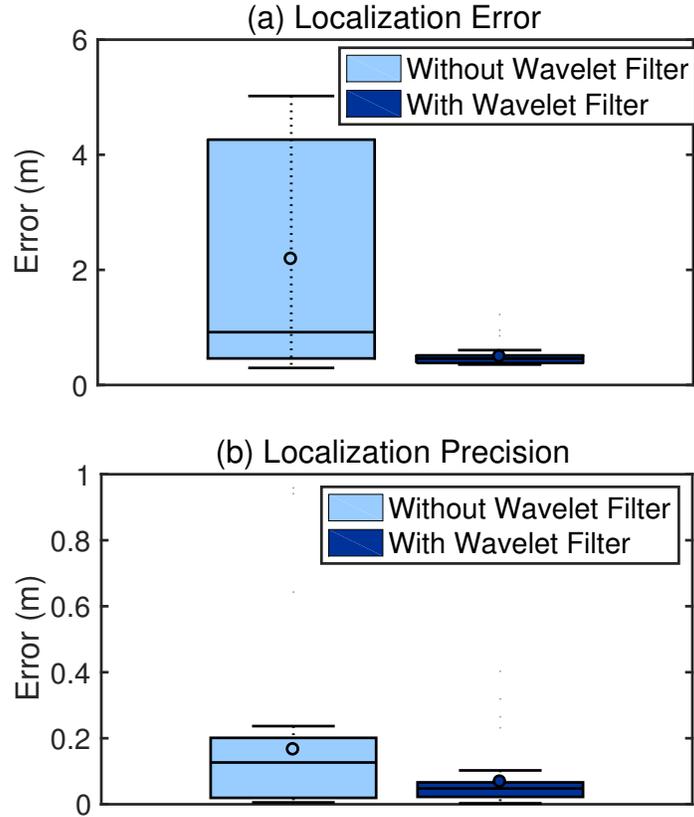
Figure 5.8: Footstep localization error and precision.

and the average location precision decreased $2\times$.

5.3.3.2 Impulse Example II: Surface Tap

To understand the system robustness, the experiment conditions discussed in Section 5.3.1 is conducted. As an example, Figure 5.9 shows estimated tap locations when the dispersion is handled through different methods. Figure 5.9 (a) shows the estimation using raw signal and cross-correlation to obtain TDoA estimation. Taps at four locations near the center show higher accuracy and precision than the rest. This is because the noise and dispersion make the TDoA estimation unstable when tap points are far from the board center, where dispersion effects for all sensors are similar.

For small scale deployment aforementioned, the calibration by generating impulse

signal at designated locations is suitable. The filtering scale is selected to minimize the estimation error at designated locations. Figure 5.9 (b) shows the results when the system filters the signal on a global band instead of a calibrated band. The estimation shows a higher accuracy and demonstrates a high precision compared to Figure 5.9 (a). Furthermore, when filtering bands are calibrated but only the lower bound of the velocity is used, the results shows a high accuracy of the relative locations and a consistent offset to their absolute locations in Figure 5.9 (c). Finally, the estimation when both filtering band and wave velocity are selected through calibration is shown in Figure 5.9 (d). The estimated locations shows less than 3cm error and less than 1cm precision.
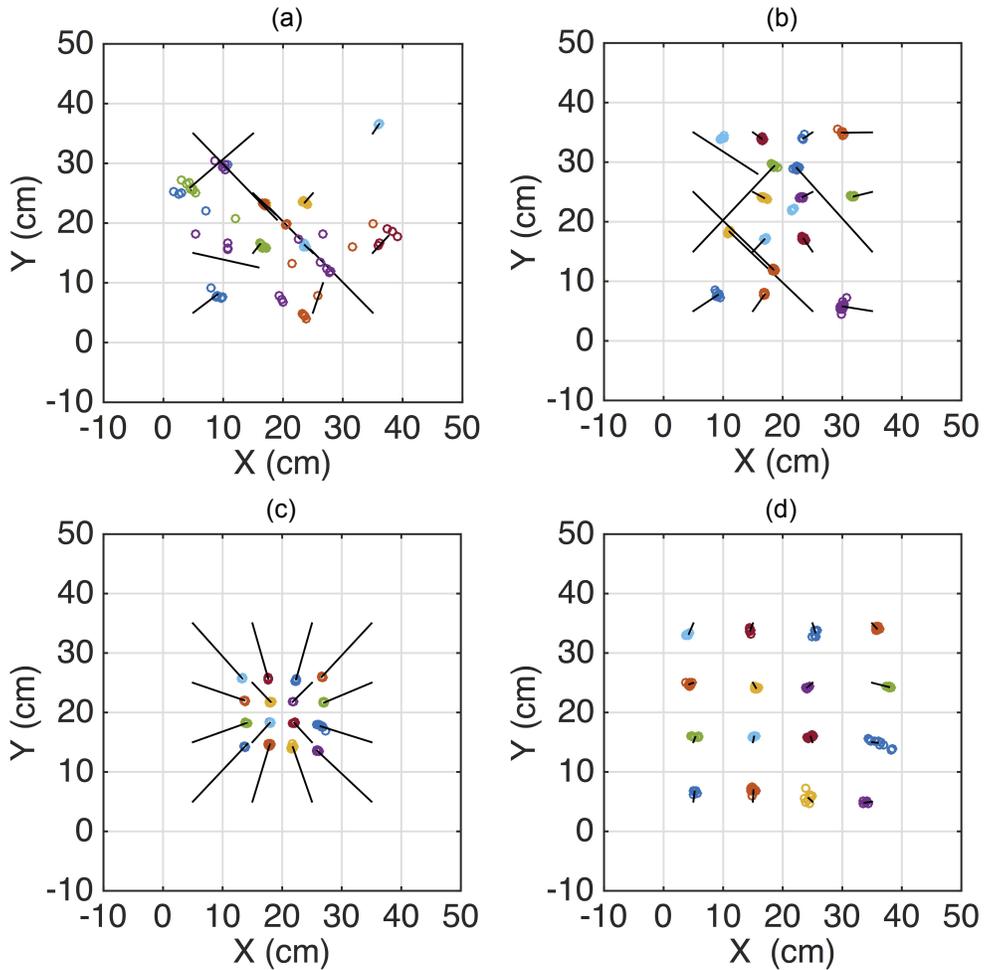
Figure 5.9: Tap localization example under different methods.
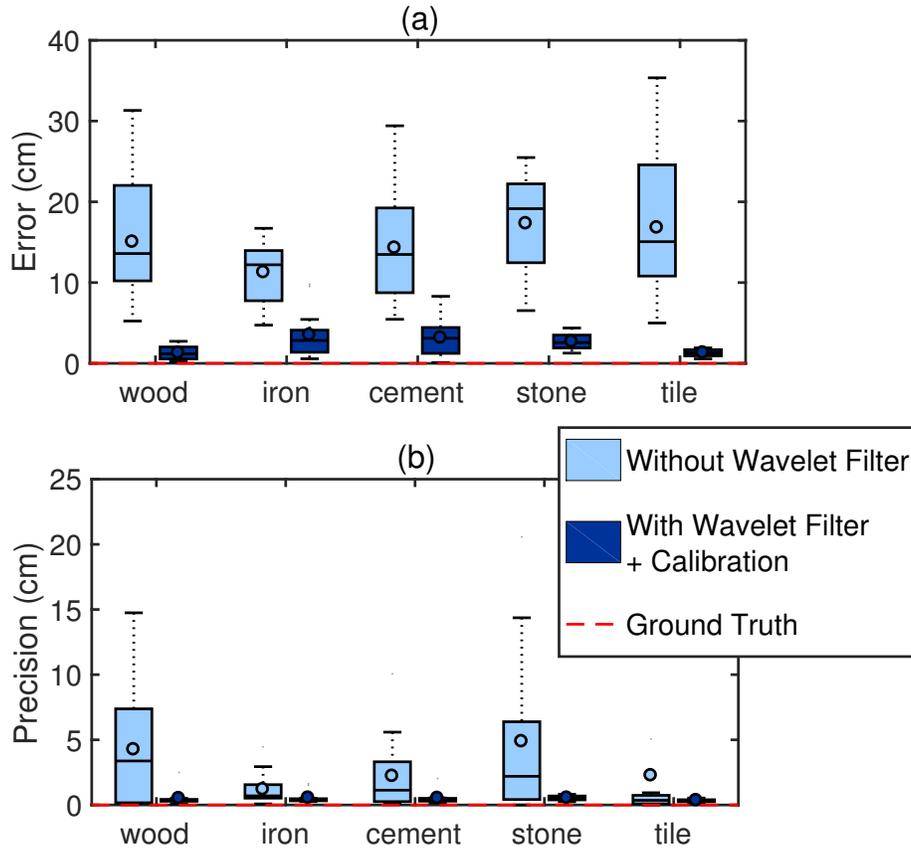
Figure 5.10: Tap localization accuracy on different materials.

Surface Tap: Surface Material Figure 5.10 shows box plots of tap localization results on different materials. Estimations utilizing calibration on both wavelet filter band and calibrated velocity have consistently lower errors than that without wavelet filter. The average localization error reduction is 14.8cm, which means the estimation that takes wave properties into account achieves $6\times$ lower average error compare to the baseline. In addition, when filtered on the selected band, the precision error is lower than 0.5cm, as compared to the 2.9cm when no filter is applied. The results show that with the properly selected filtering band and wave velocity, the system achieves up to $6\times$ lower localization error.

Surface Tap: Margin Size For the same size of sensing area ($S = 40$cm), different margin sizes including $m = 10$cm, 20cm, 30cm shows similar level of localization error (respectively 1.3cm, 1.1cm, and 1.2cm) as shown in Figure 5.11
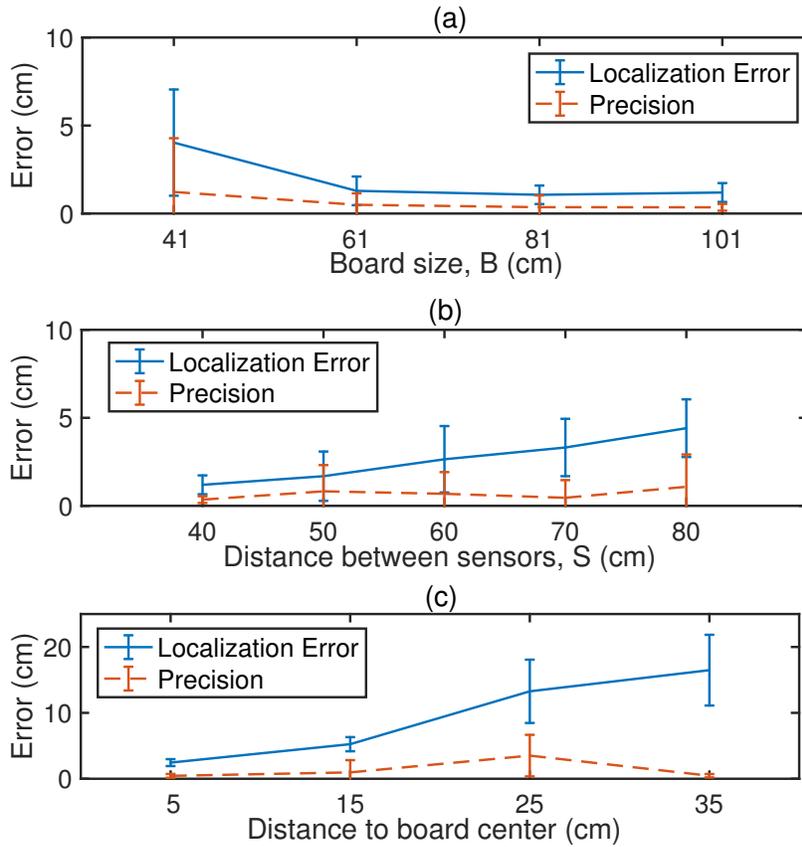
Figure 5.11: Tap localization performance over (a) different board size, (b) different between sensor distances, (c) different distance to board center.

(a). For size $m = 1$cm, the localization error is up to 4cm. This could be caused by the fact that the size of the sensing area in this case is almost the same size as the board. Therefore, the sensors are placed right at the corners of the board, where the boundary conditions are more complicated. Another possible reason for this higher localization error is that the smaller board resides less stably when taps are applied, resulting higher noise and error.

Surface Tap: Distance between Sensors For the same size of interaction area and same size of surface, the difference in sizes of sensing areas also affects the sensing ability of the system. To evaluate this factor, surface settings of $B = 101$ cm, $S = 40$ cm, 50 cm, 60 cm, 70 cm, 80 cm are investigated. Figure 5.11 (b) shows localization accuracy and precision respectively. When the distance between sensors (size of the sensing area) increases, the average localization errors goes up from

1.2cm to 4.4cm. This is because of the fact that as the interaction area increases the calibrated velocity (average velocity for the entire board) is no longer well suited due to heterogeneity of the plywood material. That is also why the localization error increases while the precision error stays at a similar level.

Surface Tap: Effective Sensing Area For the largest sensing area setting $B = 101$cm, different parts of the sensing area are investigated when the sensors are farthest from each other ($S = 80$cm). Figure 5.11 (c) shows that the further the excitation is away from the center, the higher the localization error is. When the distance of the excitation is within 20cm range from the center, the average localization error is 5.1cm, while the excitation outside this range has an average localization error up to 18.4cm. The precision error values are lower than 5cm for all testing points. The increase of the localization error is caused by attenuation and dispersion, where the first peak decays to the noise level and cannot be detected.

5.3.4 Experiment II: Slip-pulse Excitation Tracking

The slip-pulse excitation are tested through surfaces of different materials and settings. Different surfaces are compared based on parameters listed in Section 5.3.2. The system demonstrates robust interaction tracking through different parameters.

5.3.4.1 Slip-Pulse Example I: Metal Bar Dragging

For the application of human activity tracking, object dragging is evaluated on a sample wood surface with metal bars of different weights. To control the interaction strength consistently, metal bars with different weights are used. The metal bars are dragged through the designated trajectory without further pressing as shown in Figure 5.6 (b). Four metal bars with different weight (in a ratio of $1 : 1.5 : 2 : 2.5$) are dragged along testing trajectories.

Figure 5.12 shows the tracking accuracy of the four bars dragging. The tracking accuracy values are averaged through eight investigating directions as shown in Figure 5.7. The angle error standard deviation values are 36°, 31°, 42°, and 10° for the investigated four bars respectively. The length error average values are -4cm, -3.3cm, -4.5cm, and -2.6cm respectively. The results are shown a similar level of error for first three bars and the error rate is lower for the heaviest bar dragging.
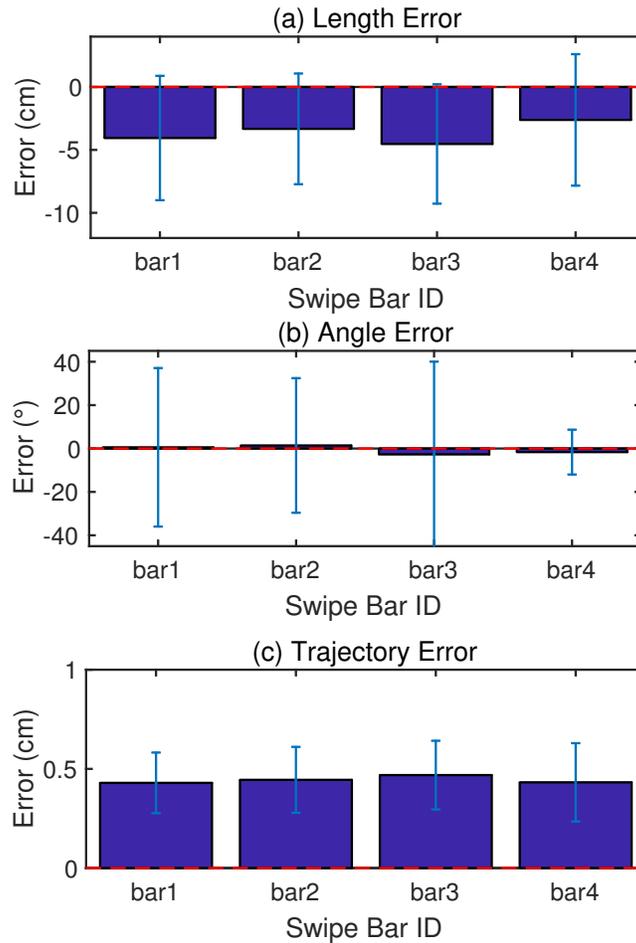
Figure 5.12: Object dragging experiments tracking accuracy v.s. object weights.

5.3.4.2 Slip-Pulse Example II: Interactive Swipe

Three different types of parameters are investigated here: 1) system implementation parameters, including the sliding window size, the mounting condition, and the type of pointers used to perform scratch, 2) swipe parameters, including swipe direction and length, and 3) surface parameters.

5.3.4.2.1 Swipe: Implementation Factors

There are three implementation factors that may affect the system performance: 1) test surface edge damping/mounting condition, 2) sliding window size used in the implementation, and 3) swipe instruments with the baseline algorithm. Experiments

91

varying these factors are conducted on a wood surface of size $B = 61$ cm.

Window Size. The sliding window size applied on swipe signals determines the range of slip pulse covered in each window. Therefore, when the window size is too small, the signal might be too short and lack of characteristics to achieve an accurate match. On the other hand, when the window size is too large, it covers the signal segments with different TDoA values. The estimation of the TDoA of the entire windowed signal will be inaccurate. The effects of sliding window sizes are then further evaluated. Five different window sizes are experimented with, including 100, 200, 500, 1000, and 2000 sample per window. With the sampling rate of 25kHz, these window sizes are corresponding to window sizes in terms of time of 0.004, 0.008, 0.02, 0.04 and 0.08 seconds.

Mounting Condition. Swipe signals are sequences of slip pulses, therefore if the reflection of preceding stick-slip signal does not die down after bouncing back at the edge, it will interfere with the later signal, and when a board is well mounted this is less likely. To prevent the reflection, the edges of the investigated surfaces are damped with sand-filled weights placed between edges and the sensing area. The wave energy dissipates to the weights before reaching to the edge.

Figure 5.13 shows the length error, angle error, and trajectory error when varying the window size and the mounting condition. The x-axis shows the varying window size, and the y-axis shows the error rate of the corresponding metric. Light blue box plots show the accuracy under undamped experiment conditions, and dark blue box plots show that under damped conditions. When the window size is less than 500 samples, the angle error is relatively high because the order of the windows appearing may differently due to the mismatch of the TDoA estimation.

For the undamped condition, average angle error values for 100-sample and 200-sample window are 68° and 52° respectively. Once the window size increased to over 500-sample, the error rate reduces to lower than 45° (26°, 1°, 3° respectively for 500-sample, 1000-sample, and 2000-sample). Similarly, the trajectory error is high when the window size is small, due to the TDoA estimation noise. Figure 5.13 shows a clear decreasing trend (0.57cm, 0.44cm, 0.29cm, 0.23cm, and 0.18cm) for the trajectory error when the window size increases.

On the other hand, the length estimation error increases when the window size increase (−6.7cm, −7cm, −8.25cm, −8.9cm, and −9.8cm). Therefore, by taking multiple evaluation metrics into account, 500 sample (0.02 s) window size is chosen
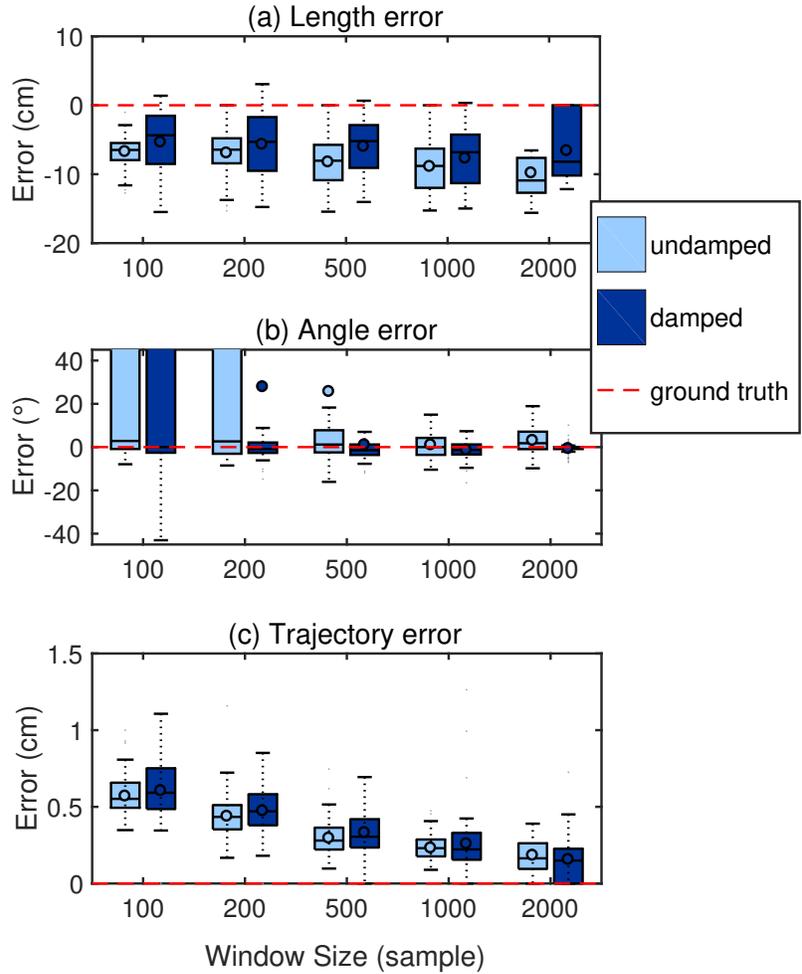
Figure 5.13: Swipe interaction tracking accuracy v.s. sliding window size.

in the surface parameter evaluation later in Section 5.3.4.2.3. Because it can achieve relatively accurate length estimation while maintains a low angle and trajectory error rate. For the damped condition, the estimation shows a lower error rate for the window size of 500 in both length error and angle error.

Pointer Material. Three different materials, including pen, fingernail, and metal bar, are studied. The tracking accuracy results are shown in Figure 5.14. These three pointers are selected to represent three different levels of signal strength: the light swipe of the fingernail, the medium strength of the pen butt, and the heavy strength of the metal bar scratching. Each of the pointer is tested through eight
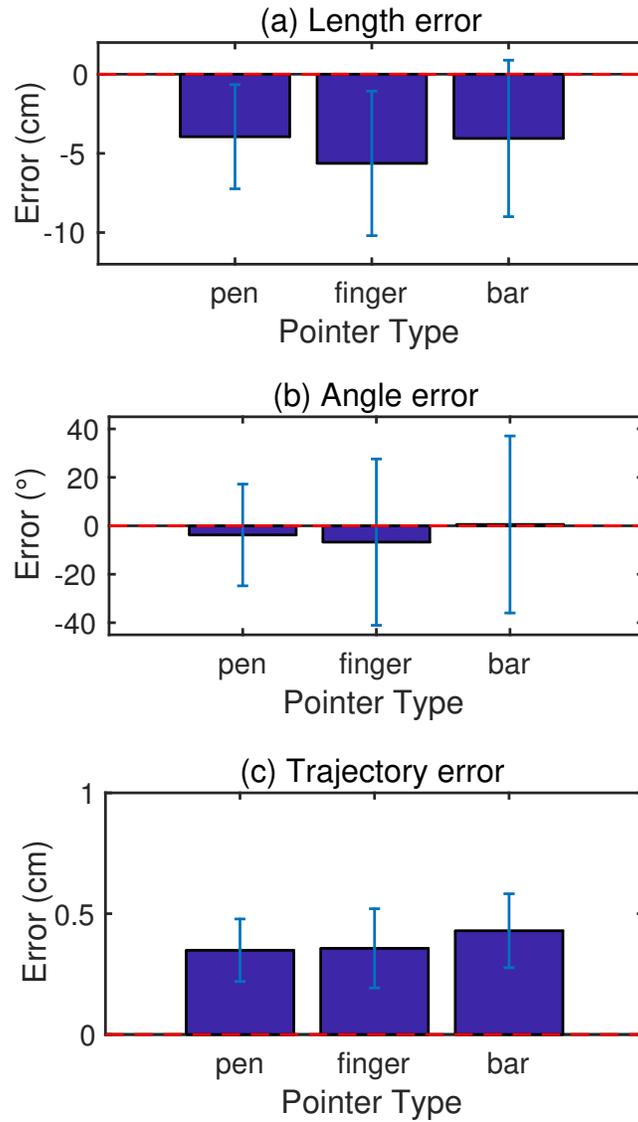
Figure 5.14: Swipe interaction tracking accuracy v.s. pointer materials.

directions, each contains 10 swipes.

Figure 5.14 shows that the pen and the bar achieve a similar level of length error value of -4cm, while the weak scratch from fingernails shows slightly higher error value of -5.6cm. These three pointers demonstrated a similar level of error on angle and trajectory errors. This indicates that the system is robust to various pointers interaction in the real-world scenario, such as pen butt on a meeting table, fingernail
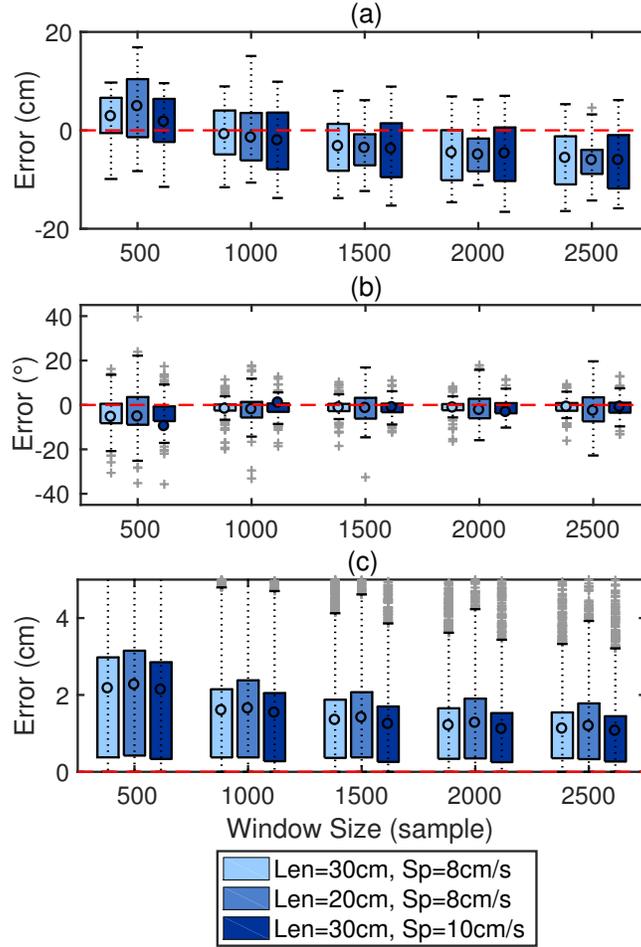
Figure 5.15: Swipe interaction tracking accuracy for different swipe length and speed.

on a wall, or a chopstick on a counter-top. In the rest of the swipe evaluation, a pen is used to perform all the swipes due to its consistency.

5.3.4.2.2 Swipe: Swipe Variables

The parameters that describe a swipe include the length, the speed, and the direction, which will be investigated in the rest of this section. The experiments are done on a ceramic tile, which demonstrates high signal to noise ratio in the aforementioned material analysis.

Swipe Length. For different sensing area, the system performs may vary due to different dispersion effects. When the length of swipes vary, the coverage of swipes

are different, therefore the dispersion effects for these lengths varies. Therefore, other than the 20 cm swipe, a longer swipe of 30 cm is further investigated to verify this assumption.

Swipe Speed. When swipes occur at different speeds, a specific window size may cover slip pulses with locations of different ranges. Therefore, when speeds of swipes varies, the optimal window size to characterize the excitation may vary. On the other hand, since the speeds of the excitation induced by human activities are limited, the potential error rate change is also within a limited range. To evaluate the speed difference caused performance change, the swipe with 8 cms^{-1}, which is slower than the 10 cms^{-1} tested in the rest of the evaluation section, is evaluated.

Figure 5.15 shows the system performance of three combinations of the swipe lengths (Len) and swipe speeds (Sp): 1) $Len = 30$ cm, $Sp = 8$ cms^{-1}, 2) $Len = 20$ cm, $Sp = 8$ cms^{-1}, and 3) $Len = 30$ cm, $Sp = 10$ cms^{-1}. Light blue, medium blue, and dark blue box plots in Figure 5.15 (a, b, c) show these combinations respectively. Figure 5.15 (a, b, c) show the length error, angle error, and trajectory error of swipes respectively. Figure 5.15 shows that when the swipe length increases, the length error standard deviation increases, and the angle error decreases. This is because that with more windowed signals, the more reference points can be used to estimate the direction of the swipe.

Swipe Direction. The swipe signal can be considered as a sequence of slip pulses, therefore when swipe to different directions, the reflection on edges may vary. The linear swipes are evaluated in this thesis to quantify the performance.

Figure 5.16 (a-h) shows an example set of swipes detected by the system towards eight designated directions plotted in Figure 5.7 (c). Figure 5.17 shows errors of the investigated swipes by their directions. Different directions show different error rate, especially the trajectory errors shown in Figure 5.17 (c). This indicates the heterogeneity of the sample material causing different reflection effects.

5.3.4.2.3   Swipe: Surface Parameters

With the understanding of the system and swipe parameters, the effects of the surface parameters is further studied in this section, including 1) surface material, 2) surface size (margin size), 3) size of the sensing area, and 4) interaction area.

Swipe: Surface Material Different surface materials have different decay rates
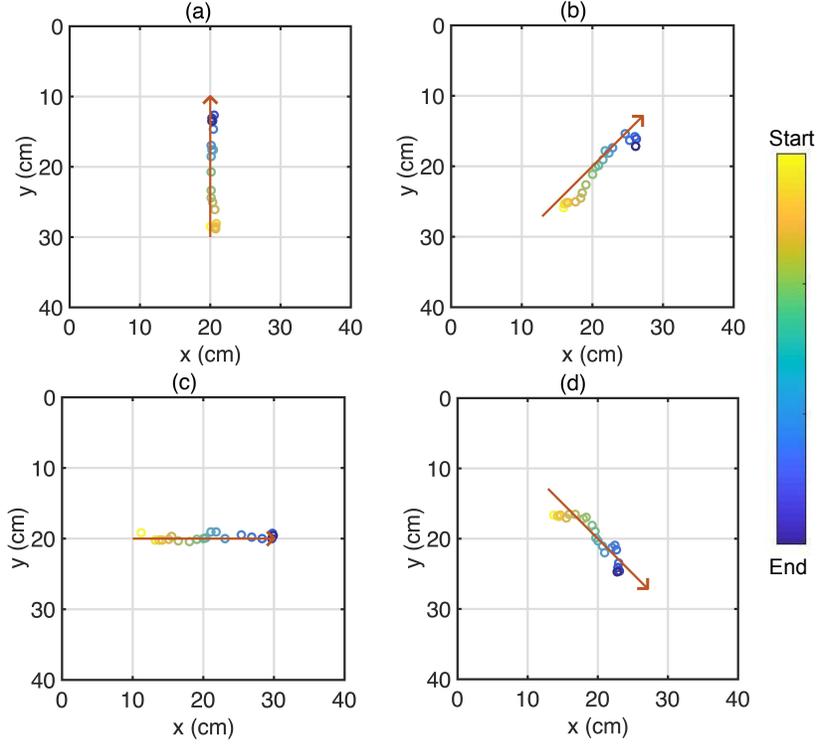
Figure 5.16: Swipe interaction tracking examples.

and reflection effects. Figure 5.18 shows tracking accuracy for five investigated surface materials under four scenarios: 1) undamped surface with baseline algorithm, 2) damped surface with baseline algorithm, 3) undamped surface with initial point correction, and 4) damped surface with initial point correction.

Figure 5.18 shows that for wood, iron, cement, and stone, the damped surface achieves lower length and angle errors. The average length estimation error over five materials are -8.7cm for damped cases, and -10.3cm for undamped cases. Similarly, the average angle error over five materials are $-7.16°$ with std of $59.3°$ for damped cases, and $-12.8°$ with std of $91.6°$ for undamped cases.

Ceramic tile shows an opposite trend on length estimation when the damped surface is used. This could be caused by the different size of the ceramic sample due to the manufacturing limitation. Hence, considering the first four materials, damping reduced the average length error from -11.1cm to -6.9cm, and the average angle error from $-15.5°$ with std of $96°$ to $-2°$ with std of $39°$ ($7\times$ reduction, $2\times$ reduction).

97
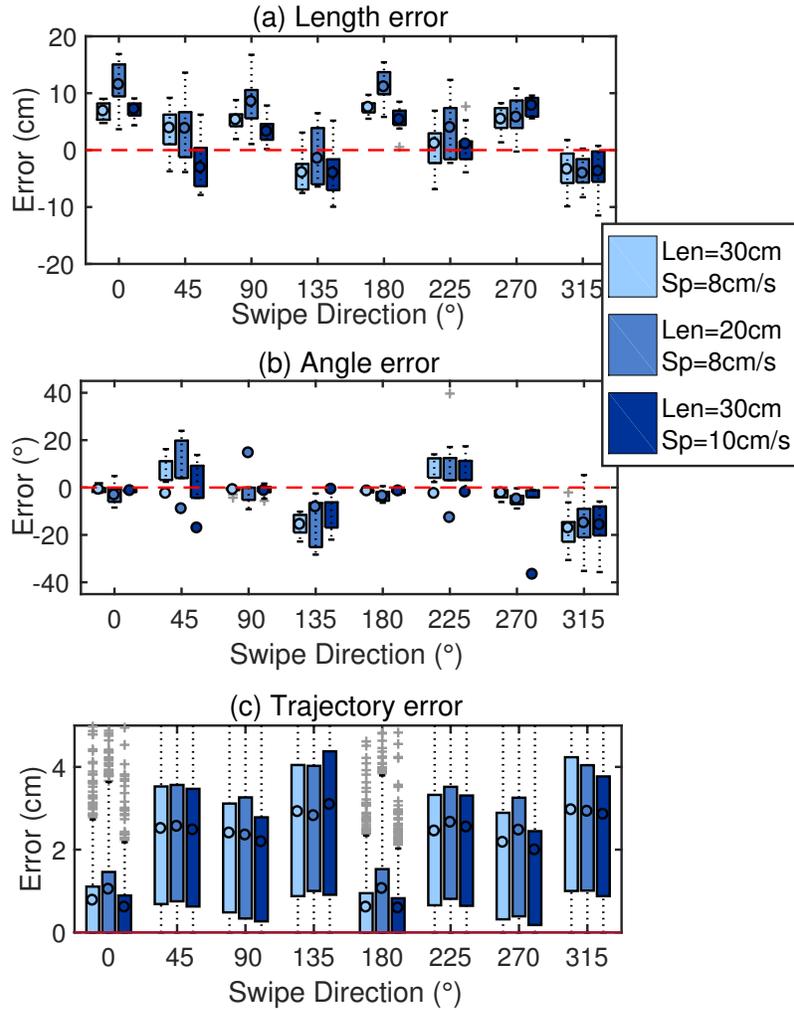
Figure 5.17: Swipe interaction tracking accuracy v.s. swipe direction.

When compared to the scenario where the initial tap location is not used, using the initial tap location allows the least length error in all scenarios. In addition, the average length error is reduced from -10.5cm to -8.1cm, and the average angle error is reduced from -11.2cm to -8.5cm. While the trajectory errors have an average of less than 0.5cm in all scenarios.

When the damp and initial tap trajectory correction were both applied, the average length error over five materials reduced from -11.3cm to -7.2cm (by a factor of 1.5). In addition, the average angle error reduced from $-14.6°$ with std $91°$ to $-6.39°$ with std $61°$ (by a factor of 2). When comparing only the first four
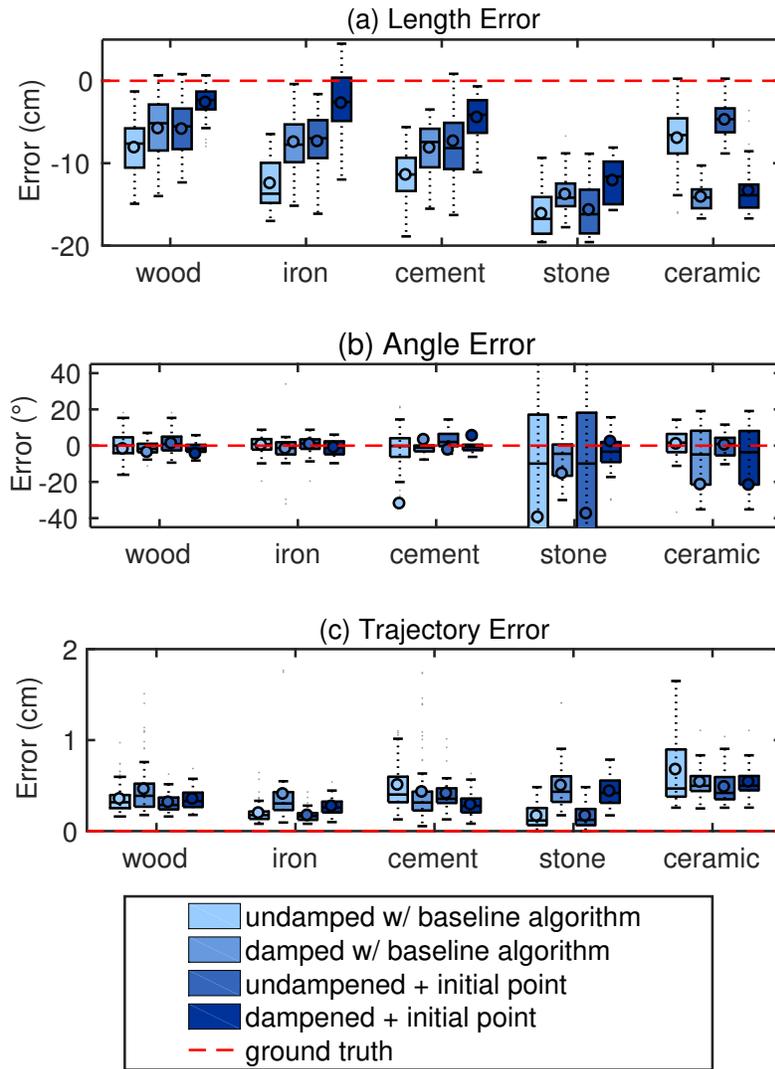
Figure 5.18: Swipe interaction tracking accuracy v.s. surface materials.

materials, this reduction is from -12.2cm to -4.6cm (by a factor of 3) and 97° to 37° (by a factor of 2.5) for average length error and standard deviation of angle errors, respectively. Therefore, both damping and initial impact based trajectory correction contribute to reducing the length and angle errors. When both are applied, the system achieves the lowest error rate.

Swipe: Margin Size Figure 5.19 demonstrates tracking accuracy of four different sizes of wood boards with damped surfaces. When the board size is similar to the sensing area, the sensors are right at the corner of the board; hence experiencing
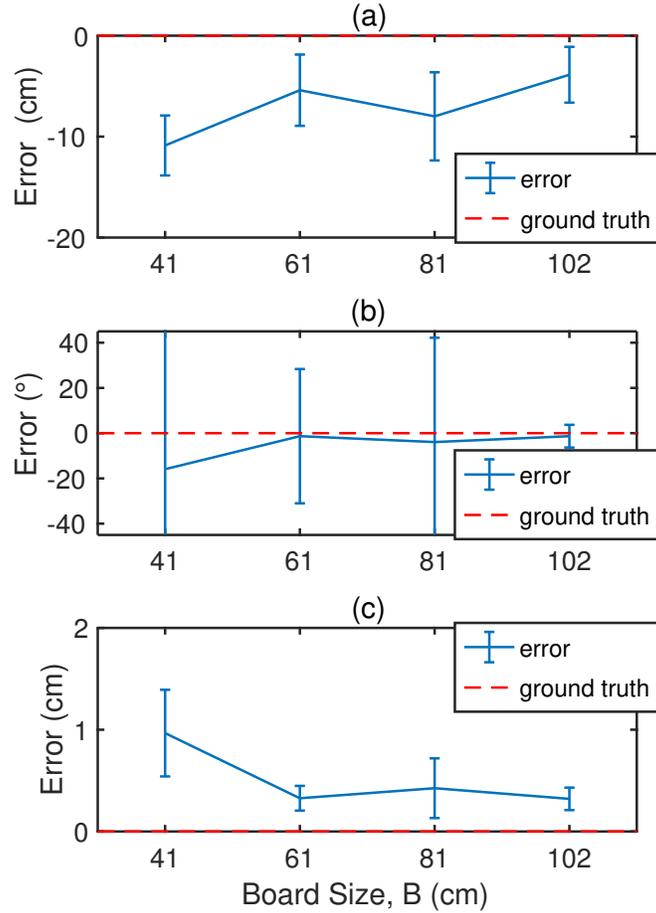
Figure 5.19: Swipe interaction tracking accuracy v.s. surface size.

the highest reflection effect. The average length error for the surface with settings of $m = 1$cm is -11cm. When m increases to 10cm, 20cm, and 30cm, the error is reduced to -5.4cm, -8cm, and -3.9cm respectively. Similarly, the average angle error for the surface of $m = 1$cm is 16° with std 101°. When m increases, the error std decreases to 30°, 46°, and 5°.

Figure 5.19 shows a trend that when the margin between the board and the sensing area increases, the swipe error decreases. This is due to the reduction of the reflection waves at board edges. When the surface size increases with the sensing area size remaining the same, it means that the distance between the board edge and the sensing area is increasing. Therefore, this additional dist acne allows the
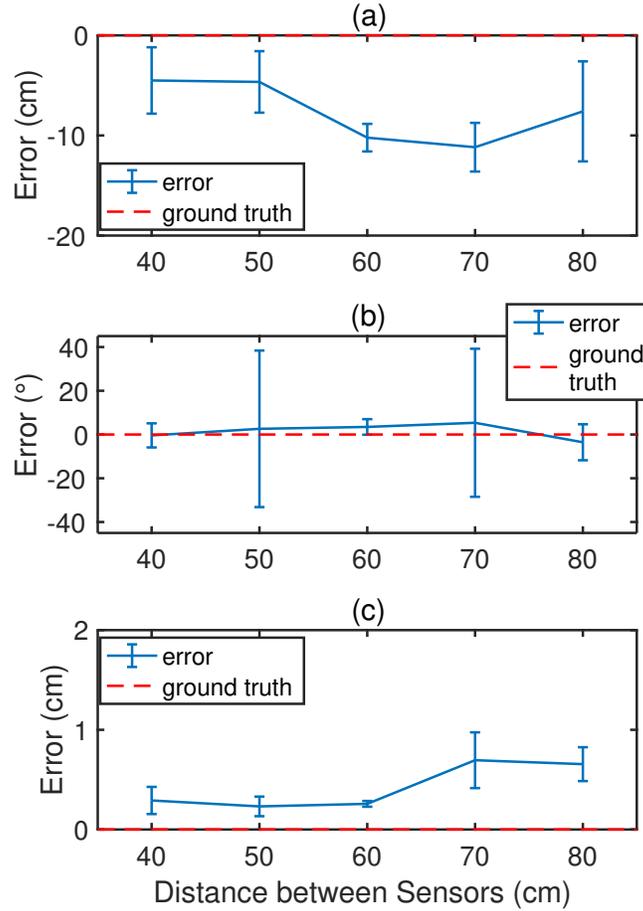
Figure 5.20: Swipe interaction tracking accuracy v.s. distance between sensors (sensing range).

reflection to die down before overlapping with the succeeding waves at sensors.

Swipe: Distance between Sensors The distance between sensors is the indication of the deployment density. The distance between sensors on large scaled surfaces will affect the attenuation rate of the interaction signal, as well as dispersion since the filtered signal is not a single band sine wave. Figure 5.20 shows the five distances between sensors that we investigated. When the distance between sensors is increased, the average length error increases (respectively -5cm, -4.7cm, -10cm, -11cm, and -7.6cm). This trend is similar to the one in the aforementioned tap experiments, which was caused by the emerging of the dispersion effect when the travel distance between the excitation and the sensor increases. The angle error std
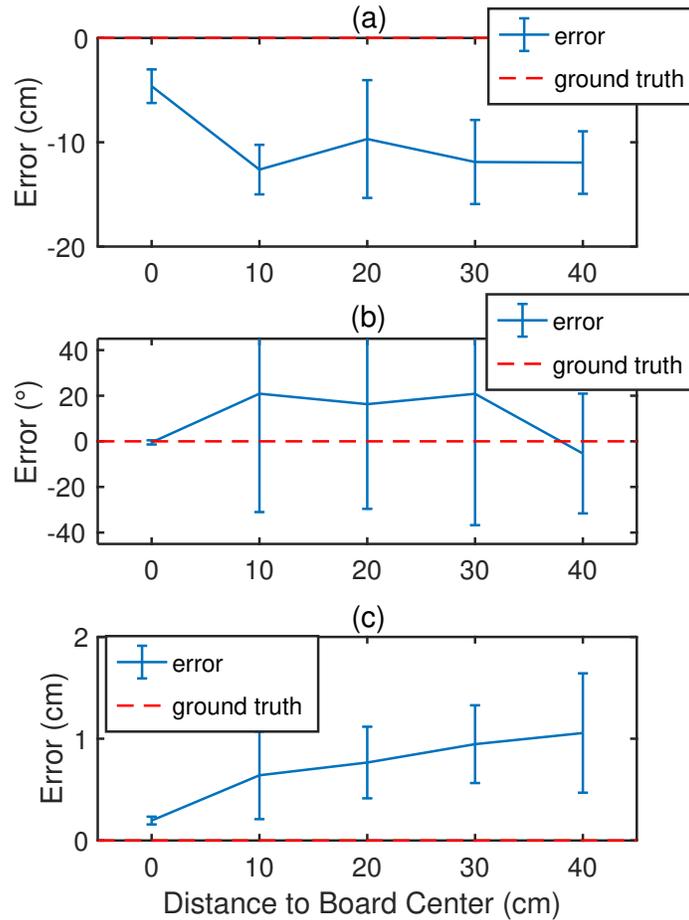
Figure 5.21: Swipe interaction tracking accuracy v.s. distance from the excitation to the board center.

raises up to 35° when the distance between sensors are 70cm. As discussed in the aforementioned tap evaluation, this change in angle error is due to the heterogeneity of the plywood surface.

Swipe: Effective Sensing Area This experiment was conducted on the largest surface as described in Figure 5.7 (d). Swipes that are at different distances from the center were conducted and the swipe tracking results are shown in Figure 5.21.

The length error and the trajectory error increased when the distance between the swipe and center was increased. The average length error for the center swipe (distance to board center is 0cm) is -5cm, while the rest vary between -10cm to -12cm. Similarly, the angle error std increased tremendously from less than 1° to

102

over 25° when swipes are moved away from the center and demonstrate more severe dispersion effects as discussed in the aforementioned tap evaluations.

5.4 Evaluation II: Occupant Traffic Estimation

To evaluate the signal characterization in terms of feature extraction and selection, two types of experiments are conducted: 1) feature validation experiments implemented by the controlled impulse load test (Section 5.4.2), and 2) scenario analyses with footstep induced vibration signals collected when occupants were asked to walk through a hallway with designated number of occupant within the group (Section 5.4.3). The research in group walking behavior that shows when multiple people walk together, they tend to fragment in smaller units of one, two, or three members [26]. Therefore, numbers of impulses/occupants from one to three are investigated in this section.

5.4.1 Implementation

The physical properties used in this implementation are the understanding of human behavior and the wave superposition properties respectively from the human side and the structure side. The non-fixed signal length SoI extraction is determined by the randomness of human behavior. The feature selection is determined by the wave propagation and superposition properties. Once the features are extracted, the system estimates the number of occupant passing by the sensing area using k-nearest neighbours [27]. The k-nearest neighbours is selected because it reflects the distribution of selected features directly. The classification procedure is equivalent to the following steps: 1) find the k neighbour points in the training set that are nearest to testing set; 2) find the labels of these neighbour points; and 3) assign the classification label based on the majority vote of its neighbors [27].

The classification model for each case is learned in the training phase with the extracted feature set $F = [f_1, f_2, f_3, f_4]$, where f_i is the ith feature value for each SoI detected from the vibration signals. Then we classify the testing cases by predicting the classification label using the k-nearest neighbor classifier model.
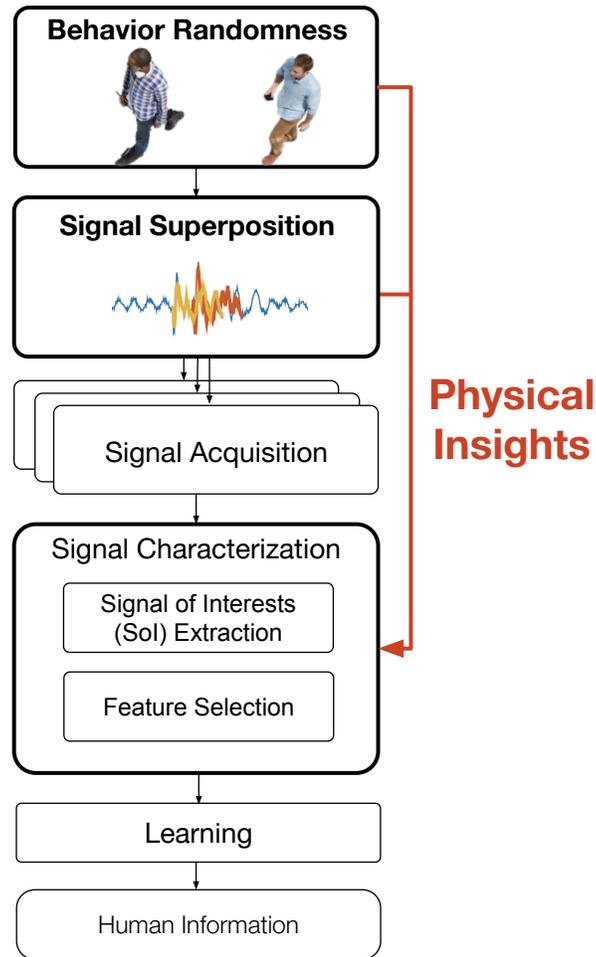
Figure 5.22: Pedestrian traffic estimation system overview.

5.4.2 Experiment I: Feature Validation through Load Test

Before conducting experiments on human, we conducted impulse load test with ball drops to understand the signal mixture conditions and their effects on the selected features. Impulse load tests with ball-drops often generate more stable vibration signals comparing to signals generated by human footsteps. In this section, we first introduce the load test experimental setup in Section 5.4.2.1, and then analyze the feature changes through different test conditions in Section 5.4.2.2.

Figure 5.23: Ball drop experiment setting. The diagram indicates the relative locations of geophones (circles) and ball drops (triangles). The grid dimensions are $1' \times 1'$.

5.4.2.1 Experimental Setup

Experiments were conducted with a high resolution oscilloscope, which samples at 10000 Hz. The experiment setting is shown in Figure 5.23 with two sensors put four feet from each other synchronized to sample. Three different conditions where the number of simultaneous ball-drops are respectively one (impulse 1), two (impulse 1 & 3), and three (impulse 1 & 2 & 3) are investigated to demonstrate the feature changing trend when the number of mixed impulse signals increases. For each investigating case, five ball-drops are conducted. The ball-drops are controlled by metronome at the speed of 45 impulses/min to achieve the impulses synchronization as much as possible. The selected features are studied through these conditions to understand the features' effectiveness under different signal mixture conditions.

5.4.2.2 Feature Analysis

To understand each feature's effectiveness, features from three traffic conditions are compared in Figure 5.24. The blue bars in Figure 5.24 display the average values of the features and the error bars indicate one standard deviations. The sub-figures (a) to (d) display results for cross correlation between spatio-difference SoIs, cross correlation between temporal-difference SoIs, SoI duration, and SoI signal entropy,

respectively.

Normalized Cross Correlation between Spatio-different SoIs The impulses at different locations introduce variation in the signal mixture. This variation in signal mixture can be observed by different sensors. Figure 5.24 (a) shows the bar graph of the mean (respectively 0.82, 0.71, and 0.64) and the standard deviation (respectively 0.02, 0.03, and 0.02) of the results. The figure displays the decreasing trend of the normalized cross correlation value with increasing number of impulses. Cross correlation is therefore a feasible method for distinguishing between different numbers of impulses when their vibration signals mix. This suggests that the feature is useful when the observed footsteps have different locations relative to different sensors.

Normalized Cross Correlation between Temporal-different SoIs When the occupant number is smaller, the chance that the SE signals within a TE are similar to each other is higher. When multiple footsteps applied on the floor without perfect synchronization, their additivity at different phases cause the temporal difference SoI to decrease. Even when multiple impulses are applied to the floor at the same pace, it is difficult if not impossible to have the step perfectly synchronized (i.e., $10^{-3}s$ for 1000 Hz sampling rate). Figure 5.24 (b) demonstrated the comparison of the results from the impulse load test, with values 0.91, 0.47, and 0.48 respectively. The result shows a clear reduction when there is more than one impulse sources. This suggests that the feature is effective for distinguishing traffic conditions between one-impulse test v.s. multiple-impulse test.

SoI Duration The more impulses mix, the higher the maximum impact time offset two signals may have, leading to larger value of the SoI duration. Figure 5.24 (c) plots the average (0.19, 0.27 and 0.28 seconds) and standard deviation (0, 0.1, and 0.04 seconds) of SoI duration corresponding to different number of impulses generated following the metronome. The mean value of the SoI duration increases from one-impulse test to multiple-impulse test. This suggests that the SoI duration is useful for distinguishing one person walking v.s. multiple people walking. It can also be observed that the two-impulse test has the higher variation in duration than the three-impulse test. This could be because that the three-impulse test is performed
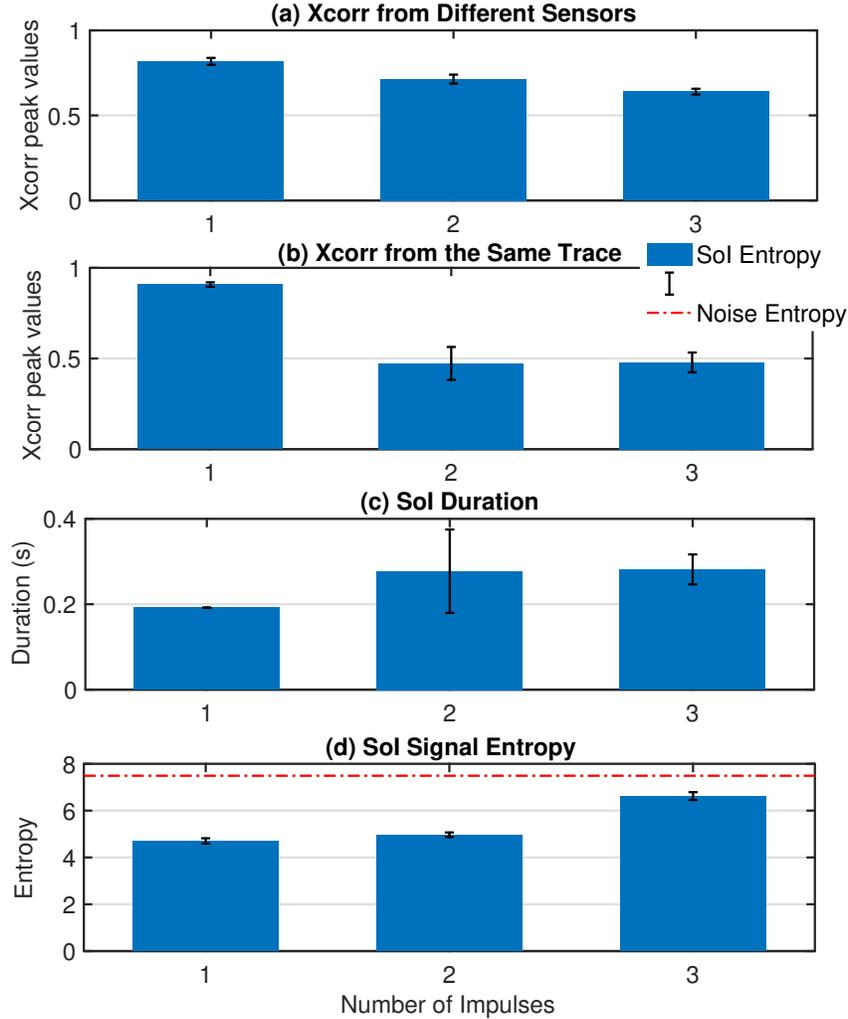
Figure 5.24: Feature analysis with impulse signals generated with different number of impulses happening at the same time in the load test.

after the two-impulse test. Therefore, people synchronize better after practicing from performing the two-impulse test. This suggests that SoI duration is sensitive to the synchronization of impulses. When the SoI duration is high, there is high probability that there are multiple people walking together, but when the SoI duration is low, other features need to be referred to determine the number of people.

SoI Entropy   The more impulses mix, the more 'complicated' the mixed signal will be. Figure 5.24 (d) shows that the entropy of a segment of noise signal is 7.49, higher than that of the SoIs. The average SoI entropy values for tests with respectively one,

two, and three impulses at the same time are 4.7, 5.0, and 6.62, with the standard deviations of 0.11, 0.1, and 0.17. The results show a clear increasing trend in SoI entropy when the number of the impulses increase. This could be caused by the blending of multiple footstep impulse signals, which makes the mixture impulse more similar to the noise (the maximum randomness). The more footstep impulses mix together, the more random the signal will be. The results indicate that increasing number of impulses will yield higher randomness, meaning entropy can be used to infer the number of impulses in occupant traffic conditions. This feature performance may degrade for SoIs with low SNR because the noise in the signal may dominate the entropy value.

5.4.3 Experiment II: Scenario Analysis through Human Test

To evaluate the selected features in the real-world application, human walking test is conducted. In this section, I will first introduce the experimental setting with people walking through the sensing area in Section 5.4.3.1. Then I will analyze and compare the characteristics of the features demonstrated in human walking tests in Section 5.4.3.2. At last, I will display traffic estimation results with different traffic conditions in Section 5.4.3.3, and explain the observations with the impulse load tests results from Section 5.4.2. Considering the randomness in human walking test is more than that in impulse load test, we extended the number of occupant to four to verify the change trend of the features observed in impulse load tests.

5.4.3.1 Experimental Setup

In the experimental setup, sensing node 1 and node 2 (described in Figure 5.25 (c)) with a distance of 10 ft in a hallway as shown in Figure 5.25 (a) and (b). We choose such distance so that 1) in our testing location, the two sensing nodes will be close enough to be on the same floor plan, hence observing the same signal without other complicated structural effect such as beams; and 2) when an occupant walks by the sensors, the sensing nodes are far enough so that their footstep energy attenuation at each sensor varies. The amplification values are set empirically so that when multiple people pass by each sensor at a distance of two feet, their footstep induced structural vibrations do not clip. The sampling rate is set to 1000 Hz. When occupants walk by the sensor group, the footstep induced vibration signal containing sequences of
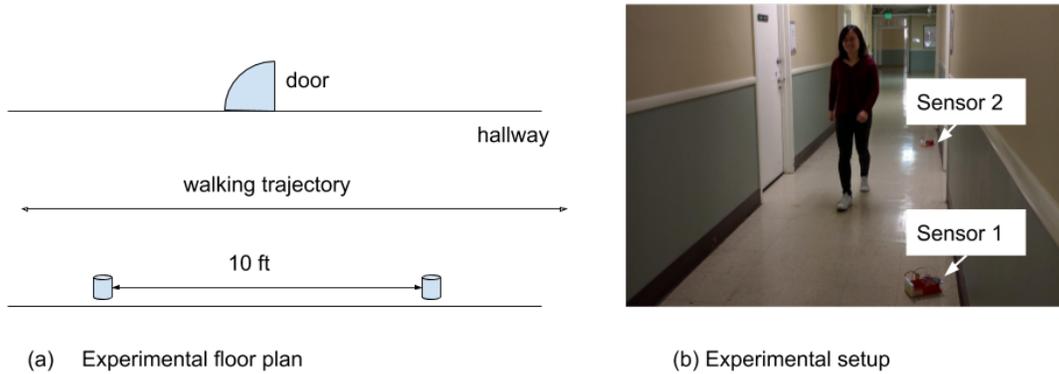
Figure 5.25: Occupant traffic estimation experimental floor plan and setup.

SoIs is called a *trace*. The selected features are extracted for each SoI within a trace, and values from the same trace are averaged for classification. For each scenario, we ask people to walk pass through the sensing area eight times. When there are multiple people walking as a group, they are asked to walk in a manner that is most comfortable for them.

5.4.3.2   Feature Analysis

To evaluate the selected features in terms of occupant counting ability, we conducted experiments in the hallway with an assigned number of occupants walking in the same direction as a group with their natural walking patterns. When they walk as a group, the distance between each individual is less than three feet. When two people walk together, they are walking side-by-side with a similar pace. When three people walk together, one of the person walked slightly behind the other two due to the limited width of the hallway. When four people walk together, two of them walked behind the other two approximately two feet away.

The cross-correlation between spatio-different SoIs reflects the signal mixture variation caused by different footstep locations. Figure 5.26 (a) displays the values of normalized cross correlation between spatio-different signals in different scenarios, whose averages over 8 traces are 0.96, 0.96, 0.89, and 0.91. The results show a general decrease trend when compare three or four occupant traffic condition to the one or two occupant traffic condition. This verifies the effectiveness of this feature when the footstep locations relative to different sensors vary. When two people walk together side-by-side (instead of one after another like the ball drop experiment setup in

Section 5.4.2), their relative location to different sensors are similar, therefore the cross-correlation between spatio-different SoIs may not be able to distinguish the case clearly.

The cross correlation between temporal-different SoIs reflects the synchronization between different footsteps. The values of normalized cross correlation between temporal-different SoIs of different occupant traffic conditions are shown in Figure 5.26 (b) with average values of 0.89, 0.79, 0.76, and 0.66, demonstrating a decreasing trend. As discussed in Section 5.4.2, this feature is effective for distinguishing one-impulse v.s. multiple impulses, which is demonstrated here as a clear drop from one-footstep to multiple-footsteps conditions.

The SoI duration also reflects the synchronization between different footsteps. The SoI durations from human walking tests are shown in Figure 5.26 (c) with average values (over 8 traces for each traffic condition) of 0.29, 0.4, 0.43, and 0.44, and standard deviation values of 0.18, 0.44, 0.49, and 0.66. Note: the plotted bars in this sub-figure shows the average of the standard deviation of SoIs within a trace. When the footsteps are more synchronized, the standard deviation is low, as observed in Section 5.4.2 (i.e., the duration is consistent). Thus, the standard deviation of the SoI duration for each trace increases as the number of persons increase. This is due to less synchronous footsteps as the number of persons increase. Similarly, the mean value of the SoIs increases when the number of people increase. Therefore, when the steps are synchronized, this feature behaves similar to Section 5.4.2. However, because there are less synchronization in human behavior, we observe that standard deviation within a trace can be used for occupant traffic estimation as well.

The SoI signal entropy indicates the complexity of the signal. Figure 5.26 (d) shows the SoI entropy value of different occupant traffic conditions, where the average over 8 traces are 4.84, 4.98, 5, and 5.04. The increasing trend, as we modeled earlier in Section 5.4.2, is shown with corresponding increasing number of occupants in a walking group. The difference between these values are not as obvious as those in Section 5.4.2. This is possibly caused by the low signal-to-noise ratio of the footstep induced signal comparing to impulse load induced signal.
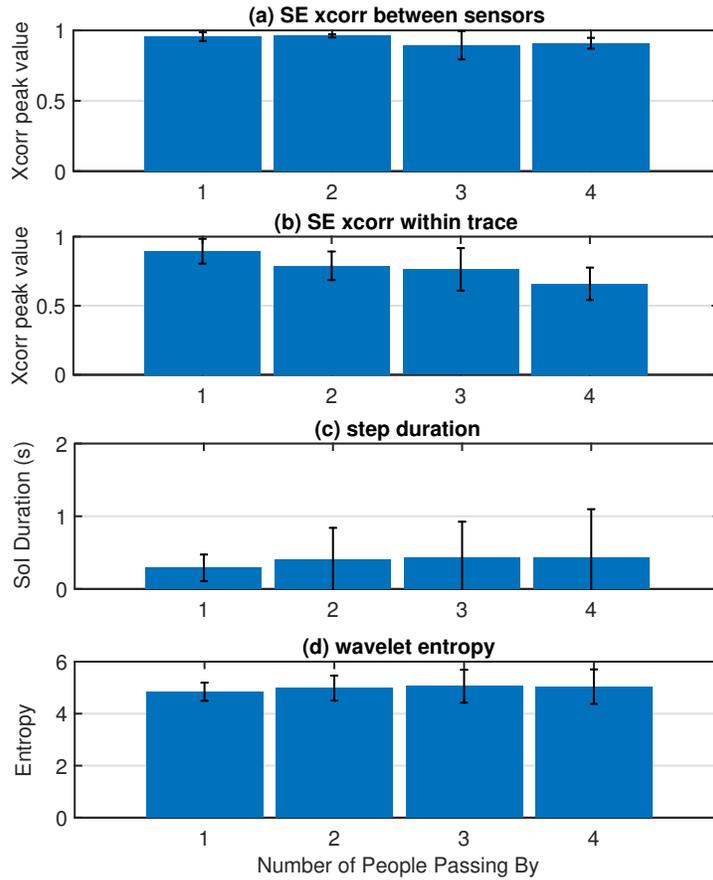
Figure 5.26: Human walking experiments with 1 to 4 pedestrians walking through the sensing area.

5.4.3.3 Classification Results

Efficiency of the selected features are studied through number of people classification using the k-nearest neighbor, which is described in Section 5.4.1. Five traces were used for training and three traces for testing through cross validation. This results in 12 data points per scenario. To understand the performance for each occupant traffic condition, the confusion matrix of this classification is shown in Table 5.2. The confusion matrix's (i, j) grid demonstrates the percentage of the scenario where number of i people walking scenario is classified as number of j people walking scenario.

Table 5.2 shows that 16.67% of the one occupant traffic condition is confused as two occupants traffic condition, but is not confused as three or four occupant

Accuracy	1P	2P	3P	4P
1P	83.33%	16.67%	0%	0%
2P	25%	66.67%	8.33%	0%
3P	8.33%	25%	33.33%	33.33%
4P	0%	8.33%	0%	91.67%

Table 5.2: Confusion matrix for people counting (1-4 people walking) [99].

traffic conditions. The two occupants traffic condition is mostly confused with one occupant traffic condition, which could be caused by the synchronized side-by-side walking that confuses feature cross correlation between spatio-different SoIs and the SoI duration. The three occupants traffic condition has the lowest accuracy and are confused with two and four occupants traffic conditions with up to 33% chance. As discussed in Section 5.4.3.2, the cross correlation between spatio-different SoIs can confuse the three v.s. four occupant traffic condition. The cross correlation between temporal-different SoIs can confuse the two v.s. three occupant traffic condition. This results in the three occupant traffic condition being confused with two and four occupant traffic condition. Finally, the four occupants traffic condition estimation is fairly robust, only confused with the two occupants traffic condition in one instance. The mean estimation of the number of occupant passing by for the one to four occupants traffic conditions are 1.16, 1.83, 2.91, and 3.83 respectively. The result demonstrates less than 0.2 people mean estimation error for each scenario.

5.5 Related Work

In this section, we mainly discuss prior works based on two applications introduced in this chapter: 1) the human activity excitation tracking and characterization through vibration and 2) indoor occupancy traffic estimation.

5.5.1 Vibration-based Tracking and Characterization

In this chapter, physics models (wave properties, wave dispersion) are used to enhance the accuracy of robust human activity induced excitation (vibration source) tracking through different materials. Works have been done on robust vibration source tracking and characterization using physical principles, despite limited sensing

data/labeling and human involvement [47, 56, 88, 103, 158]. Jia et. al. utilize bed vibration to estimate human heart rate by analyzing the wave propagation model [56]. Han et. al. measure the vibration of the vehicles to estimate their relative location (whether on the same lane) utilizing the physics insight of the vehicle vibration and road conditions [47]. Xu et. al. use structural vibration to infer traffic condition in the street [158]. Mirshekari et. al. present the footstep localization in dispersive structures specifically and how to utilize physical information about the structure to assist sensing [88]. Pan et. al. estimates the human-surface interaction locations by taking different excitation types (tap and swipe) into account in the tracking algorithm [103]. These prior works are all focus on a specific application, and in this chapter, a general approach that can apply to many different applications is presented.

5.5.2 Occupancy Traffic Estimation

The occupancy traffic information has been proven useful for various energy management in smart building environments [4, 33]. Many passive sensing methods and apparatuses have been proposed, including using cameras [22, 25, 157, 163], IR sensors [62, 69, 134], RF sensors [155, 156], ultrasonic range finders [52], etc. Optical-based methods (camera and IR sensor), usually require installation in the designated area/position (e.g., on the ceiling pointing down, or ankle girth height in both side of the doorway), which is costly in terms of installation and maintenance, esp. in large-scale deployments. Similarly, ultrasonic range finder Doorjamb [52] based methods require designated sensor installation at each monitored doorway. RF-based methods [155, 156] also require extra radio transceivers in high density. Compared to these existing methods and apparatuses, the structural vibration based system utilizes easy-to-install structural vibration sensing units to obtain occupant traffic estimations and faces less sensing constraints such as the requirement of line-of-sight.

5.6 Chapter Summary

The signal characterization is an essential module in the cyber-physical systems and can directly determine the information learning accuracy later. In this chapter, the

methodology of utilizing physical properties of the vibration signal waves to guide the signal characterization is introduced. Two information acquisition examples are presented to demonstrate the importance of the signal characterization and how the study of the wave physical properties can improve the information acquisition accurately.

The first example is pairwise Time Difference of Arrival (TDoA) estimation for human excitation induced structural vibration source tracking. Different excitation causes different types of waves, which vary in propagation properties and are characterized differently based on these wave properties. The impulse-like interaction-induced (such as footstep and tapping,) vibration signal is filtered based on the wavelet decomposition and reconstruction, and the TDoA is extracted with the peak detection and matching to avoid wave reflection. On the other hand, the slip-pulse-like interaction induces shear wave dominated waves that show less dispersion effect and can be used directly for TDoA estimation. The human excitation induced vibration source localization is then used as an evaluation example to demonstrate that the physics property guided TDoA estimation allows the tracking system achieves up to $6\times$ improvement on impulse localization and up to $3\times$ improvement on slip-pulse trajectory length estimation.

The second example is pedestrian traffic estimation, which the mixture of waves from separate excitation sources is the superposition of the two separate signals. This property determines that the features over multiple sensors at different locations capturing a mixture of signals from excitation at different locations can represent the traffic condition in terms of a number of excitation sources within the sensing area. The combination of the features that consider spatio-temporal variations between different sensors capturing the pedestrian traffic is evaluated in the real-world traffic estimation application and achieved less than 0.2 people mean estimation error for traffic condition from 1 to 4 people.

Chapter 6

Human Information Learning through Structural Vibration Signals

Various human information can be inferred from the structural vibration their activities induced, including pedestrian identity [104, 106], location [86], number of people [99], physical conditions of people (e.g., heart rate [56], foot balance [34]). Traditional data-driven machine learning methods that target on these learning problems rely on the labeled training data to build the model.

This assumption makes the accurate learning difficult for two reasons: 1) The physical world is complicated. To accurately model it, a large amount of labeled training data is needed to cover various sensing conditions, 2) It is often difficult to obtain sensing data from the physical world, let alone labeled training data. Therefore, in this chapter, I will first introduce the physical attributes that change data distributions, and why it is a problem for traditional data-driven learning methods (Section 6.1). Then I will discuss the key observation as well as how to utilize this observation to achieve accurate learning with limited initial labeled data.

6.1   Physical Attributes: Examples and Characterization

Various physical attributes or conditions would affect sensing signal data distribution. In this section, the pedestrian identification is used as an example to discuss the potential factors that may cause the variation in footstep-induced vibration signals.

Prior work demonstrates the difference between Step Events (SEs) from different people due to their gait variation [104]. Here, the sensing condition variations that cause the prior work identification accuracy decrease will be quantified based on 1) an individual's Step Events (Section 6.1.1) and their physical causes (Section 6.1.2).

6.1.1   Step Event Variation and Clustering

The SEs from an individual's dataset show variations that can be summarized by a few representative waveforms. In order to investigate the collected SEs in a systematic manner, we conduct hierarchical clustering [59] using pairwise SEs' similarity as distance, which is calculated from the peak values from the cross-correlation between time-domain signals of two SEs. The distance between each pair of SEs SE_i and SE_j is calculated as $distance = 1 - xcorr(SE_i, SE_j)$, where $xcorr$ calculates the peak cross correlation value of the signal pair. Then the pairs with the shortest distance are grouped into clusters made up of two SEs. This process is repeated for the resulting clusters until all the SEs in the original dataset are linked together in a hierarchical tree.

To decide how many clusters to form based on a specific tree generated from a person's data, a threshold value is empirically determined based on the experiments in the Load Test with ball-drops [106] to incorporate the structural variation within a designated area. For each node in the hierarchical tree, if the similarity between SEs within the internal node of the tree is higher than the threshold, they are consider to be in the same cluster. These clusters are then used to explain the cause of the SE variation in the following section.

6.1.2   Causes of Step Event Variation

The causes of the SE variation are two-fold: 1) a person may have multiple gait patterns that generate different waveforms when striking the floor at different step frequencies, and 2) when a person steps on different locations in the sensing area, the structural variation may cause the structural response difference. However, when under the same sensing condition in terms of these two factors, the SEs from a pedestrian are repeatable (hence can be clustered) despite the variations. To further investigate how these factors affect the data distribution, Person #1's SEs and their

(a) One person's SE cluster in different areas

(b) One person's SE cluster through different step frequencies
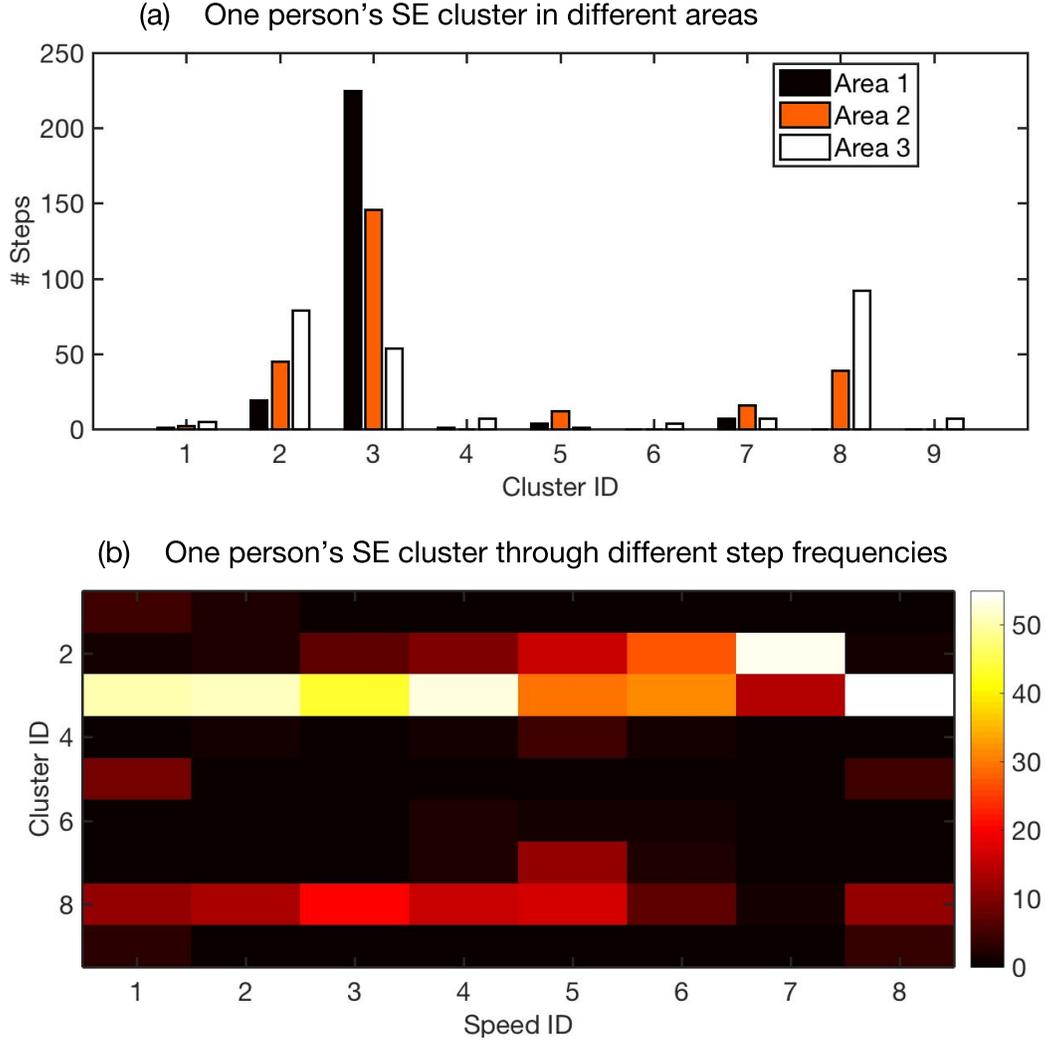
Figure 6.1: Footstep induced vibration signal cluster example of Person # 1 shows that the footstep signal of a person change gradually with their footstep frequency.

clusters are further analyzed as an example in Figure 6.1.

6.1.2.1 Structural Variation

A person's SEs in different locations within the sensing range may look different due to the structural variation. Figure 6.1 (a) categorizes the clusters based on the estimated step location. Area 2 is closest to the sensor location, while Area 1 and Area 3 are on either side of Area 2. SEs from cluster 2 appear in Area 3 more often than in Area 2 and Area 1, and SEs from cluster 3 shows an opposite trend
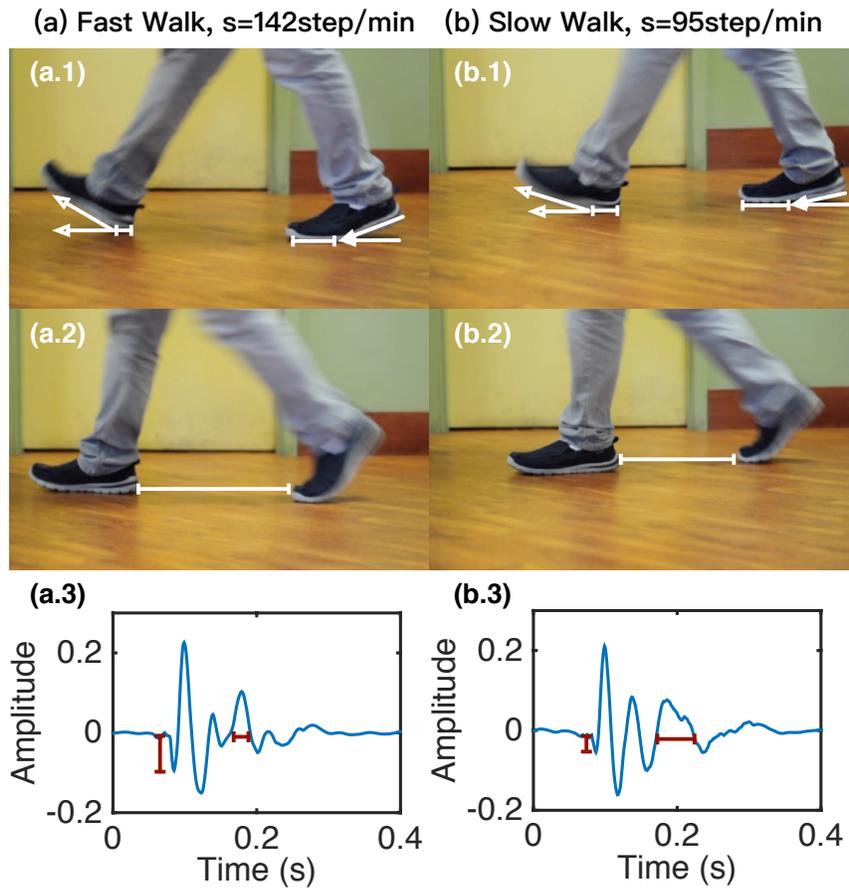
Figure 6.2: Step Events extracted under fast and slow step frequencies.

through the three areas. This indicates that the waveforms of SEs are affected by step locations. This could be caused by 1) the structural variation and the response signal alters or 2) the anisotropic characteristics of the floor and the wave changes differently during the propagation [133]. In addition, such structural variation is gradual, which explains the repetitive appearance of the cluster at the same area in Section 6.3.1.5.1.

6.1.2.2 Step Frequency

A person's gait changes when the person walks at different step frequencies [106]. Figures 6.2 (a.1-a.2) and (b.1-b.2) show the example photos of the same person walks at different step frequencies. Figures 6.2 (a.1) shows the initial strike of the person walking at a step frequency of 142 steps per minute, while (b.1) shows that of 95

steps per minute. The angle between the foot and the floor of at the initial contact is larger when the person walks faster. In addition, compared to Figures 6.2 (a.2), (b.2) shows a smaller stride length when the walking step frequency is lower.

When the gait changes, the corresponding SEs change too. Figures 6.2 (a.3) and (b.3) show the example step signal when a person passes by the sensor with a different step frequency. As we can see in the time domain signal in Figures 6.2 (b.3), the part of the signal around 0.2 seconds decays more smoothly than that in Figures 6.2 (a.3). This is caused by the slow stance phase (the foot remains in contact with the ground) in the gait cycle when people walk slowly. In addition, the first dip of the signal has a greater value in the fast walking SE than in the slow walking SE. This could be caused by the heavier heel strike when the walking speed is high. These SEs are repetitive in signals from the same person walking at the same designated step frequency. Therefore, a person's footstep-induced vibration when walking through a designated area with different step frequency can be summarized into a few representative waveforms, i.e., different clusters.

Figure 6.1 (b) categorizes the clusters of Person $\# 1$ by controlled step frequencies. The participant walks at a designated step frequency (speed ID 1 to 7) to generate SEs. Seven levels of designated step frequencies are assigned, including the average step frequency μ_{sf}, low frequencies from $\mu_{sf} - 3\sigma_{sf}$ to $\mu_{sf} - \sigma_{sf}$, and high frequencies from $\mu_{sf} + \sigma_{sf}$ to $\mu_{sf} + 3\sigma_{sf}$, where σ_{sf} is the step frequency standard deviation [1]. The experiment details will be introduced in Section 6.3.1.5.2.

In this figure, nine clusters are generated from the SEs of seven different step frequencies, and Clusters 2, 3, and 8 appear most frequently. Clusters 2, 3, and 8 appear more often in the high, low, and medium step frequency SEs respectively. From the low step frequency SEs to the high step frequency SEs, the portion of Cluster 3 decreased and the portion of Cluster 2 increased. This observation further verifies our conclusions on the relationship between step frequencies and gait changes.

6.2 Physical Attributes Guided Iterative Learning

Measurable physical attributes referring to data distribution change degree can be used to solve inaccuracy caused by the data distribution changes even for significant

---

[1]Note that in the rest of the paper, we refer to μ_{sf} and σ_{sf} as μ and σ for simplicity.

changes [162]. The algorithm Iterative Learning using Physical Constraints (ILPC) is presented in this section. ILPC trains a classification model iteratively (Section 6.2.1) and controls the order of unlabeled data used in each iteration according to measurable physical constraints or physical attributes (Section 6.2.2).

6.2.1 Using Multiple Domain Adaptation Models to Cover Gradually Changing Distributions

When data distribution changes significantly, ILPC iteratively constructs multiple domain adaptation models in order to handle the changing distribution. A single domain adaptation model can handle a limited range of distribution changes. In each iteration, ILPC labels the unlabeled data predicted in the current iteration with a prediction confidence score higher than an empirical threshold. In this way, our ILPC method extends the distribution of labeled data. In the next iteration, ILPC constructs another domain adaptation model using also the newly labeled samples, which then can predict accurately with a broader data distribution than the previous iteration. Multiple iterations of this extension process will eventually cover all changing distributions. The initial domain adaptation model is trained with the initial labeled data and a portion of selected unlabeled data. For example, in the pedestrian identification example, the initial labeled data are SEs with medium walking speeds. When a person walks faster or slower (unlabeled), the training model is extended gradually using both labeled and unlabeled data.

6.2.2 Guiding the Model Distribution Order with Physical Constraints

To use the iterative approach discussed in Section 6.2.1, it is assumed that, within each iteration, the unlabeled and labeled data distributions should not have a significant difference. To select the unlabelded data that has similar data distributions with the labeled data, ILPC utilizes the relationship between the data distribution changes and the physical constraints and selects the unlabeled data with the particular physical constraint values.

First, the physical constraints that cause data distribution changes is identified. For example, in the pedestrian identification application, we know from previous

research that walking speed and walking pattern (gait) are strongly correlated (Section 6.1). Once these constraints are defined, the sensing system extracts the data (feature) along with its corresponding physical constraints. Based on the physical constraint values, the system decides if the distribution of the test data 1) is within the labeled data distribution, 2) has a similar enough distribution to the labeled data to be trained in the next iteration, or 3) has a very different distribution from the labeled data and therefore needs to wait for the model to be extended.

6.2.3   ILPC Algorithm

ILPC algorithm is based on the two aforementioned key ideas (Section 6.2.1 and 6.2.2). The pseudo code is shown in Algorithm 1. $DomainAdaptation(D_{labeled}, D_{unlabeled})$ is a function that conducts traditional domain adaptation learning [107, 130]. The system is initialized by determining which physical measurements affect the data distribution. Then it discretizes the physical constraints into k levels. The range of the physical constraint $[x_{min}, x_{max}]$ is uniformly discretized into $\{x_1, x_2, \cdots, x_k\}$, where $x_1 = x_{min}$, $x_k = x_{max}$. It means that there are $k-1$ intervals $[x_n, x_{n+1}]$, where $1 \leq n \leq k-1$.

6.2.3.1   Iterative learning.

To address the distribution change in each interval, a domain adaptation model is constructed for each interval (two adjacent data distribution levels). Therefore, for k levels of data distribution, $k-1$ domain adaptation models are constructed. Then the n-th model is assigned to handle the data distributions subject to the identified physical attribute with values in the range $[x_n, x_{n+1}]$, where $n \in \{1, 2, \cdots, k-1\}$. The value k can be determined by observing the empirical histogram of the sample count. An k is considered to be appropriate when there are sufficiently many data samples in each interval $[x_n, x_{n+1}]$.

In the n-th iteration, ILPC constructs the n-th model by domain adaptation methods based on the selected unlabeled data $Data_{selected}$ with physical constraint values in $[x_n, x_{n+1}]$ and the labeled data $Data_{labeled}$. Selecting $Data_{selected}$ with physical constraint values in $[x_n, x_{n+1}]$ guarantees limited distribution change between $Data_{selected}$ and $Data_{labeled}$, which can be handled with high accuracy by existing domain adaptation methods.

define the physical constraint x;
uniformly discretize x into $\{x_1, x_2, \cdots, x_k\}$;
for $n \leftarrow 1$ **to** k **do**
    $Data_{selected} = []$;
    for *each sample in* $Data_{unlabeled}$ **do**
        if $x_n < x_{sample} \leq x_{n+1}$ **then**
            $Data_{selected} = Data_{selected} \cup sample$;
        end
    end
    $Prediction, Confidence = DomainAdaptation(Data_{labeled}, Data_{selected})$;
    $New\_Labeled\_Data = []$;
    for *each result in Prediction* **do**
        if $Confidence_{result} > threshold$ **then**
            $New\_Labeled\_Data =$
             $New\_Labeled\_Data \cup result$ *and the corresponding sample*;
        end
    end
    $Data_{labeled} = Data_{labeled} \cup New\_Labeled\_Data$;
end

Algorithm 1: The ILPC algorithm [162].

6.2.3.2 Model expansion.

In the n-th iteration, the n-th model is used to label the unlabeled data with high prediction confidence $Data_{selected}$. The final confidence score of a prediction can be calculated from multiple sample points measured for the prediction, e.g., summing based thresholding.

6.3 Evaluation

To evaluate the physical attribute guided iterative learning algorithm, an pedestrian identification system – FootprintID – is implemented (Section 6.3.1). The metric and parameters to measure the performance of the algorithm is then discussed (Section 6.3.2). Finally, the algorithm performance is discussed (Section 6.3.3 and 6.3.4).

6.3.1 Implementation: FootprintID

The FootprintID algorithm is implemented leveraging the ILPC algorithm to obtain a pedestrian's identity from their step-induced floor vibration signals. The algorithm consists of three modules: the information extraction (Section 6.3.1.1), the structural variation handler (Section 6.3.1.2), and the step frequency variation handler (Section 6.3.1.3).

6.3.1.1 Information Extraction

The information extracted from the raw ambient floor vibration are mainly two-folds: 1) the inter-footstep information (e.g., step frequency) and 2) the intra-footstep information (e.g., features of each SE). The system first needs to separate the vibration induced by the footsteps (SEs) from the ambient noise (Section 4.2.1). Based on these SEs, the system further extracts the step frequency and relative location, which is referred to as the inter-footstep information (Section 6.3.1.1.1). Finally, for each SE, the system extracts its signal frequency components, which is referred to as the intra-footstep information (Section 6.3.1.1.2).

6.3.1.1.1 Inter-Step Information Extraction

The inter-step information describes the sensing condition of the intra-step information. The step relative location is estimated based on the wave attenuation model tracking algorithm that utilizes SE signal energy change trend to determine the location [97]: the closer the step is to the sensor, the higher the signal energy of the SE. The step frequency is estimated based on the average time interval between consecutively detected SEs. To avoid noise affecting the step frequency estimation by causing the system to miss an SE, the system takes the mean excluding the highest and lowest K values of the time interval array within each trace. In this section, $K = 2$ is used.

6.3.1.1.2 Intra-Step Information Extraction

The intra-step information describes the interaction between the footstep and the floor. For each SE, first of all, it is normalized by the signal energy to remove the footstep-sensor distance difference on the signal energy. The frequency band is selected based
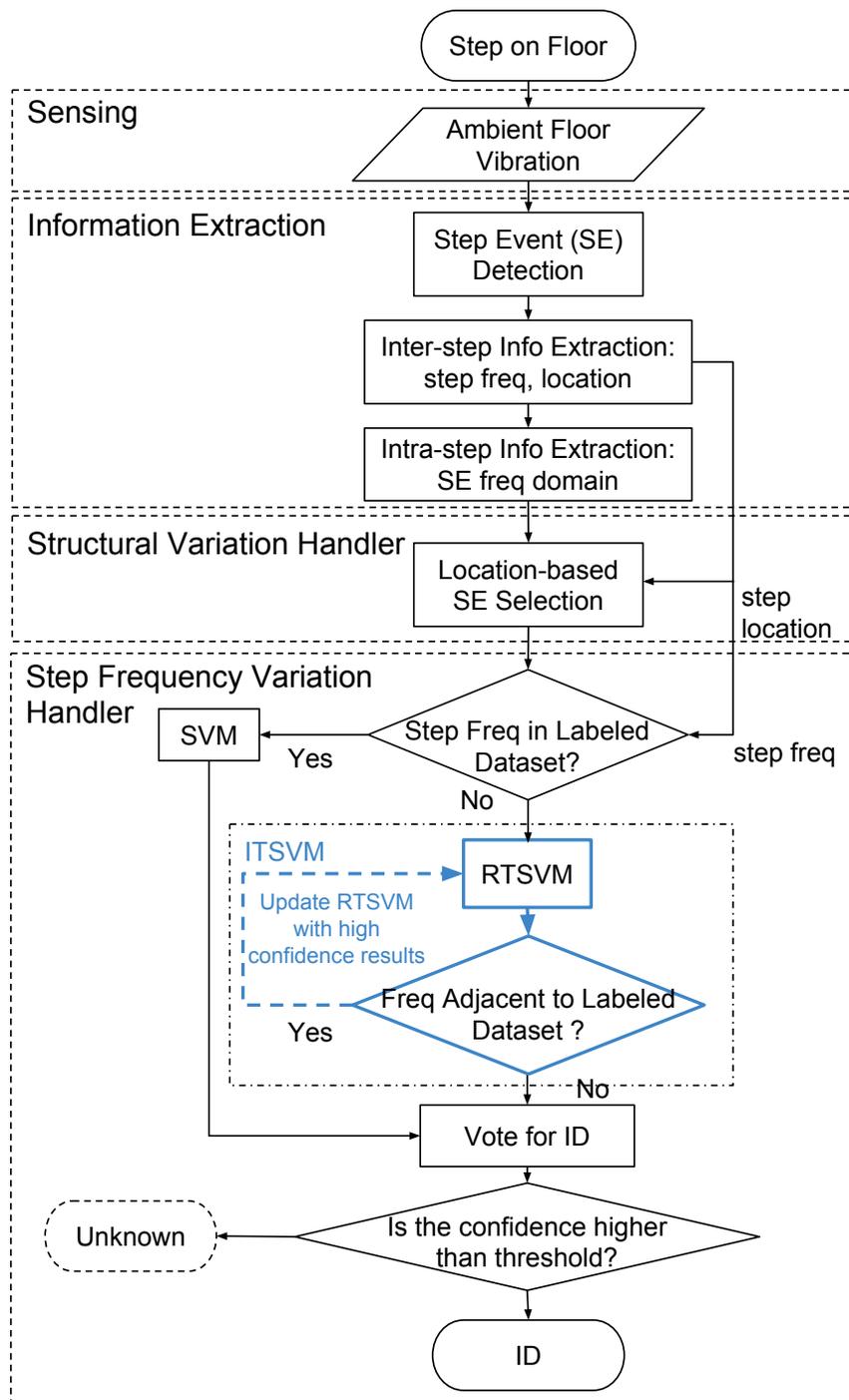
Figure 6.3: FootprintID system overview.

on the sensor properties: the sensor's response frequency is $f_{response} \leq 240$ Hz. Since the floor velocity response to footsteps has the characteristic frequency band between 20 and 90 Hz [32], the selected band is sufficient to represent the signal characteristics. Therefore, the cut-off frequency is selected to be at least twice as much as 90Hz and below 240Hz. In the implementation, for each SE, the 0 to 200 Hz power spectrum is selected as a feature to describe the characteristics of the signal.

The system then combines the inter-step and intra-step information as features in the form of an array:

$$
\begin{aligned}
Feature(SE) &= [f_1, f_2, ..., f_N] \\
&= [SE\_freq(1Hz), ..., SE\_freq(200Hz), step\_freq]
\end{aligned}
\tag{6.1}
$$

where N = 104. The first 103 features represent frequency domain signals between 0 to 200 Hz discretized evenly. The last feature is the step frequency of the walk discussed in Section 6.3.1.1.1. For each f_{num} $(num = 1...N)$, the system conducts feature normalization using the corresponding feature values from the labeled training data (f_{num_train}) to achieve uniform weight through all features, i.e.,

$$
f_{num\_norm} = (f_{num} - \min(f_{num\_train})) / (\max(f_{num\_train}) - \min(f_{num\_train})) \tag{6.2}
$$

Then the system generates the SE features $\boldsymbol{x}$ by normalizing each feature in the array:

$$
\boldsymbol{x} = [f_{1\_norm}, f_{2\_norm}, ..., f_{N\_norm}] \tag{6.3}
$$

which is used for SE classification in Section 6.3.1.3.

6.3.1.2 Structural Variation Handler: Location-based SE Selection

When a pedestrian walks through the sensing area, the structural variation through the sensing area causes the same pedestrian to have SEs with different waveforms despite a consistent gait. When the consistent gait is applied at the same location, the SEs share a similar waveform (Section 6.1). This means that even similar foot strikes are only directly comparable when they are from the close proximity area.

Therefore, to ensure the SEs are comparable, SEs that are from a close proximity from each trace is selected. The trend of the SE energy change can be used as an indicator of their relative locations to the sensor [97]. To use that that, a sliding

window is applied on the SE energy values to smooth the trend change on the *trace*, which is defined as one individual's consecutive SE sequence. The value of the sliding windows are then averaged, which smooths the SE energy trend changes. Finally, the peak of the smoothed SE energy sequence is selected as the point indicating the closest proximity of the person to the sensor. We will further discuss the structural variation and effects of location-based SE selection in the load test Section 6.3.1.5.1, and step location evaluation Section 6.3.2.2.

6.3.1.3 Step Frequency Variation Handler

When a pedestrian walks through the sensing area with different step frequency, their footstep-induced structural vibration signal varies. This is the one of the key challenges for gait-based person identification through structural vibration sensing. For a pure data-driven approach, to handle this data distribution variation, a large amount of labeled training data collected from diverse walking speeds is required to achieve robust identification. If the labeled training data misses SEs of a particular walking speed range, the testing classification accuracy for SEs in that range decreases.

FootprintID handles this problem by choosing between supervised learning and an iterative transductive learning method (`TSVM`) based on the detected footstep step frequency. As shown in Figure 6.3, for the testing data with step frequencies in the labeled training data physical attribute's range, FootprintID applies the supervised learning model directly. Otherwise, FootprintID chooses the transductive learning model [38] and extend the model iteratively [106].

For the implementation of the FootprintID, a Support Vector Machine based transductive learning model is adopted for three reasons: 1) SVM-based methods have proved an efficient classification method for decades [17], 2) compared to a neural network, SVM-based methods require less labeled data in general to achieve accurate modeling [43], which fits the application's profile, where the data labeling has high manual cost, and 3) the prior work on person identification through footstep induced floor vibration adopts SVM [104], which is the baseline to compare the learning accuracy with.

`ITSVM` (marked in blue in Figure 6.3), an iterative transductive support vector machine algorithm, resolves the SE variance caused by different step frequencies. In this section, I will first introduce the traditional support vector machine (SVM,

126

Section 6.3.1.3.1) and how it handles step frequency variation between training and testing datasets. Then I will introduce the transductive SVM (`TSVM`, Section 6.3.1.3.2) and how we tailor `TSVM` (Section 6.3.1.3.3, 6.3.1.3.4) to incorporate the physical properties of our data.

6.3.1.3.1 Support Vector Machine (SVM)

The Support Vector Machine (SVM) has been widely used to solve classification problems. Given two-class training data $(y_1, \boldsymbol{x}_1), \dots (y_l, \boldsymbol{x}_l)$, it aims to find the maximum-margin hyperplane $\boldsymbol{w}$ by minimizing the following loss function [17]:

$$\min_{\boldsymbol{w}, b} \quad \frac{1}{2}||\boldsymbol{w}||^2 + C \sum_{q=1}^{l} \max(1 - y_q(\boldsymbol{w}^T \phi(\boldsymbol{x}_q) + b), 0), \tag{6.4}$$

where $\boldsymbol{x}_q \in R^n, \forall q$ are training samples, $y_q = \pm 1, \forall q$ are corresponding labels. C is a penalty parameter balancing regularization term ($\frac{1}{2}||\boldsymbol{w}||^2$) and training losses ($\sum_{q=1}^{l} \max(1 - y_q(\boldsymbol{w}^T \phi(\boldsymbol{x}_q) + b), 0)$), which control the generalization of the model. The kernel function $\phi(\cdot)$ in nonlinear SVM, which projects the feature space into a higher dimension [19]. The kernel enables the models with high class-separability and generalization ability, even with a small amount of training data [19, 104]. After obtaining $\boldsymbol{w}$ with some optimization techniques, the sign of function value $f(\boldsymbol{x}_q) = \boldsymbol{w}^T \phi(\boldsymbol{x}_q) + b$ decides the class of the test data $\boldsymbol{x}_q$.

For multi-class problems, most existing multi-class SVM methods decompose the problem into several two-class classification problems [53]. The two commonly used methods are 1) the one-against-one strategy, which trains a model for pairwise classes of training data, and 2) the one-against-rest strategy, which trains a model for every class against the rest of the classes.

The traditional nonlinear multi-class SVM achieved up to 96.5% accuracy in identifying the participants when they walk at a specific speed during a short amount of time [104]. However, when participants walk with different step frequency, we observe significant variations of SE signals across different step frequencies (Section 6.1). When this variation appears between training and testing dataset, the traditional supervised learning may fail (Section 6.3.1.5.2).

127

6.3.1.3.2 Transductive SVM (TSVM)

Transductive learning utilizes both labeled and unlabeled data in the form on the loss function to take data distribution difference between labeled and unlabeled data into account. The gradually change between the labeled and available unlabeled data (Section 6.1) can be leveraged to improve prediction accuracy when the transductive learning is applied.

Instead of finding a decision boundary with maximal margins over labeled data by SVM, TSVM aims to train a boundary with maximal margins over both labeled data and unlabeled data. Therefore, compared to the loss function in Equation 6.4, TSVM has an additional loss function term representing the margin over all the unlabeled data. Given two-class training data $(y_1, \boldsymbol{x}_1), \ldots (y_l, \boldsymbol{x}_l)$ and unlabeled data $\boldsymbol{x}_{l+1}, \ldots \boldsymbol{x}_{l+u}$, TSVM aims to find the maximum-margin hyperplane $\boldsymbol{w}$ and bias term b by minimizing the following loss function:

$$
\begin{aligned}
\min_{\boldsymbol{w}, b} \quad & \frac{1}{2}||\boldsymbol{w}||^2 + C_1 \sum_{q=1}^{l} \max(1 - y_q(\boldsymbol{w}^T \phi(\boldsymbol{x}_q) + b), 0) \\
& + C_2 \sum_{q=l+1}^{l+u} \max(1 - |\boldsymbol{w}^T \phi(\boldsymbol{x}_q) + b|, 0),
\end{aligned}
\tag{6.5}
$$

where C_1 and C_2 are the penalty parameters balancing regularization term and training losses over labeled data and unlabeled data, respectively. Intuitively, TSVM tends to find boundaries in regions where there is less labeled and unlabeled data. Thus, it is a low-density separation method studied in transductive learning problems [20].

The binary TSVM can be extended to handle a multi-class problem using a one-against-one or a one-against-rest strategy (Section 6.3.1.3.1). For a k-class SVM, the one-against-rest strategy solves k binary SVM problems, each of which treats one class as positive and all the rest as negative, and its accuracy often yields to the one-against-one strategy [42, 53]. Therefore, in this work, the one-against-one TSVM is selected to handle the multi-class problem.

6.3.1.3.3   Refined `TSVM` (`RTSVM`)

One potential problem in the one-against-one setting for multi-class `TSVM` is the irrelevant unlabeled data. For example, in a k-class problem, when the binary model for class 1 and 2 is constructed, the unlabeled data is from all k classes, while only the unlabeled data from 1 and 2 is relevant to the binary classification problem between class 1 and 2.

If all the unlabeled data is used, the drawbacks are two-folds. 1) the model accuracy may be low due to the large amount of irrelevant data. When $(k-2)/k$ percentage of the unlabeled data is irrelevant, the binary `TSVM` may not effectively capture the information hidden behind the unlabeled data of class 1 and 2. 2) the training of `TSVM` can be slow due to the large amount of unlabeled data. Note that solving `TSVM` is a combinatorial optimization problem, where the size of search space grows exponentially when the number of unlabeled data increases. Therefore, the increasing amount of unlabeled data may significantly reduce the training speed. In this case, a refined (selected) unlabeled dataset, which contains less unlabeled test data for modeling, leads to a faster training speed.

A mechanism to refine the relevant unlabeled data for the training of each binary `TSVM` in the one-against-one setting is designed, which is called `RTSVM`. In `RTSVM`, every time the system trains a binary `TSVM` between class i and j, it utilizes supervised SVM to pre-select unlabeled SEs, which are 'most likely' to be class i or j. The multi-class SVM modeled from the labeled data is used to predict the identity of all unlabeled SEs. Then the `RTSVM` calculates the most frequently appearing class in each trace as the class of the trace and uses SEs from traces of class i or j for the binary `TSVM` modeling. `RTSVM` is an intermediate step in our algorithm.

6.3.1.3.4   Iterative `TSVM` (`ITSVM`)

One assumption in the `RTSVM` method is that for each one-against-one comparison, the unlabeled data and the labeled data are from the same two classes selected based on the SVM results. However, this assumption may not be true all the time, especially when the step frequency variation is large (e.g., step frequencies that are in the range of $\mu \pm 2\sigma$, $\mu \pm 3\sigma$).

On the other hand, for the unlabeled data with step frequencies that have a relatively small difference compared to that of the labeled training data ($\mu \pm \sigma$), the
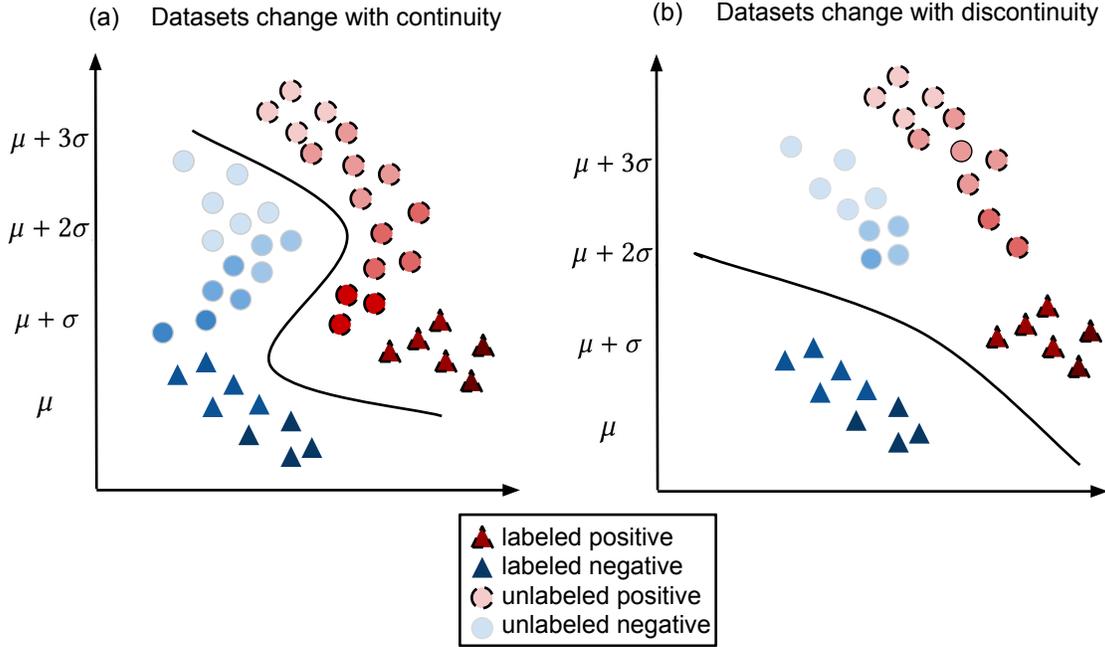
Figure 6.4: The importance of continuously changing frequency in unlabeled datasets for `TSVM` and the potential modeling error caused by the discontinuity in the unlabeled data.

SVM classification results a higher accuracy and the `RTSVM`, which is built upon that, also results a higher accuracy as well.

From the signal characterization in Section 6.1, it is observed that the dataset between rare and average step frequencies ($\mu \pm \sigma$) are a combination of Step Events with waveforms from those datasets (μ, $\mu \pm 2\sigma$, $\mu \pm 3\sigma$). Therefore, based on those observations, the multi-class `TSVM` model can be trained in an iterative way, which is referred to as `ITSVM`:

1. Step 1, the multi-class `RTSVM` is constructed for the test data in the frequency of $\mu \pm \sigma$. This model can be used to label some unlabeled data in the frequency of $\mu \pm \sigma$. After this labeling procedure, the size of the 'labeled' dataset increases. Note that the 'labeled' dataset grows in a conservative way. When the `RTSVM` result confidence is higher than a threshold, the system labels the unlabeled data. The system calculates this confidence as the percentage of unlabeled SE's in a single trace predicted to be the same class by `RTSVM`. The updated labeled dataset is across the frequencies of μ and $\mu \pm \sigma$, which allows the SVM with rare step frequency data to achieve higher accuracy compared to the SVM

with the original labeled dataset (μ).

2. Step 2, based on the updated labeled dataset, the multi-class RTSVM is constructed with the test data with step frequencies of $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$.

The transductive learning model is constructed with both the data of rare step frequencies ($\mu \pm 2\sigma$ and $\mu \pm 3\sigma$) and the data used in the Step 1 (μ and $\mu \pm \sigma$). The data from Step 1 is used because of the low-density separation intuition behind TSVM [20]. When there is a continuous changing in the unlabeled data from different frequencies, the TSVM is more likely to locate the true decision boundary.

The low-density separation cases are illustrated in Figure 6.4. Figure 6.4 (a) shows the intuition data distribution of all unlabeled data (between μ and $\mu + 3\sigma$), and (b) demonstrates the intuition data distribution of rare step frequencies (between $\mu + 2\sigma$ and $\mu + 3\sigma$). If the data with frequency $\mu + \sigma$ (similarly, $\mu - \sigma$) is excluded in the unlabeled set when training TSVM, the sparsity of data in $\mu + \sigma$ region may place the TSVM's hyperplane in that region.

6.3.1.4 Trace Identity Calculation

Once FootprintID conducts identity estimation on each SE using the classification algorithm introduced above, it further combines the results of each SE in a trace to estimate the identity of each trace. Since each trace can be assigned to one person, utilizing multiple data points (SEs) would improve estimation accuracy [104]. FootprintID conducts a vote and selects the most frequent ID as the representative of the trace ID.

Furthermore, for each trace, the system calculates the confidence of the identification decision. Then, based on the identification accuracy requirement for different applications, the system assigns the traces with confidence values lower than the application-based threshold as 'unknown'. The trace level decision confidence is calculated from the step-level prediction confidence values. As aforementioned, for each SE, the *step-level* prediction confidence is calculated from outputs of one-against-one SVM or ITSVM in FootprintID. Therefore, for a k-class problem, a number of $k \times (k-1)$ pairwise decision confidence can be obtained. For an SE predicted as class i, there are k one-against-one (binary) SVM or ITSVM models. Each of these models predicts a class with the confidence c_j, where $j \in \{1, 2, \cdots k\}$. This confidence score c_j is calculated from the distance between the sample $\boldsymbol{x}$ and the hyperplane $\boldsymbol{w}$ of

j-th (binary) SVM or ITSVM model as $d_j = \frac{\boldsymbol{w}^T \phi(\boldsymbol{x}) + b}{|\boldsymbol{w}|}$, by linearly normalize the d_j. The confidence of this data point being classified as class j is $C^{\text{step}} = \sum_{j=1}^{k} c_j$.

Note that the scale range of different binary SVM or ITSVM's confidence c_j may vary for different $j \in \{1, 2, \cdots k\}$. Thus, in practice, for each binary SVM or ITSVM, their prediction confidences are linearly normalized into the same scale range $[0, 1]$. The *trace-level* prediction confidence is then calculated based on the *step-level* prediction confidence of partial steps in this trace. Assume the system predicts a trace with m steps as class i after the vote, and there are n steps in this trace predicted as class i. For each step in these n steps, the corresponding prediction confidence is C_p^{step}, where $p \in \{1, 2, \cdots, n\}$. Then the system calculates the *trace-level* prediction confidence as $C^{\text{trace}} = \sum_{p=1}^{n} C_p^{\text{step}}$.

6.3.1.5   Experimental Settings

Three experiments are conducted with varying parameter control levels to evaluate the system: 1) the load test with tennis ball-drops, 2) the human walking experiments with controlled step frequencies, and 3) the uncontrolled human walking experiments. The load test is designed to investigate the sensing area's vibration response characteristics, therefore, the impulse source and location are controlled. The analysis from the load test is used to select the clustering threshold introduced in Section 6.1 to take the structural variation into account. The human walking experiments are designed to investigate different system parameters and evaluate algorithm robustness. In the controlled walking experiment, the step frequency of each participant is the controlled variable. The uncontrolled experiment is designed to evaluate the overall system robustness in realistic scenarios.

6.3.1.5.1   Structural Variation: Load Test with Ball-drops

In the load test, ball drops are used as the source of consistent impulse excitation at different locations. These signals are used to understand location-based variation in the structural vibration responses. Human footsteps are subject to human behavior randomness, therefore ball drops are used here to isolate the structural factors.

To understand the location-based variation in structural vibration responses, the controlled impulses are generated by ball-drops from a designated height. At each location, five ball drops are conducted from approximately 2ft above the floor for
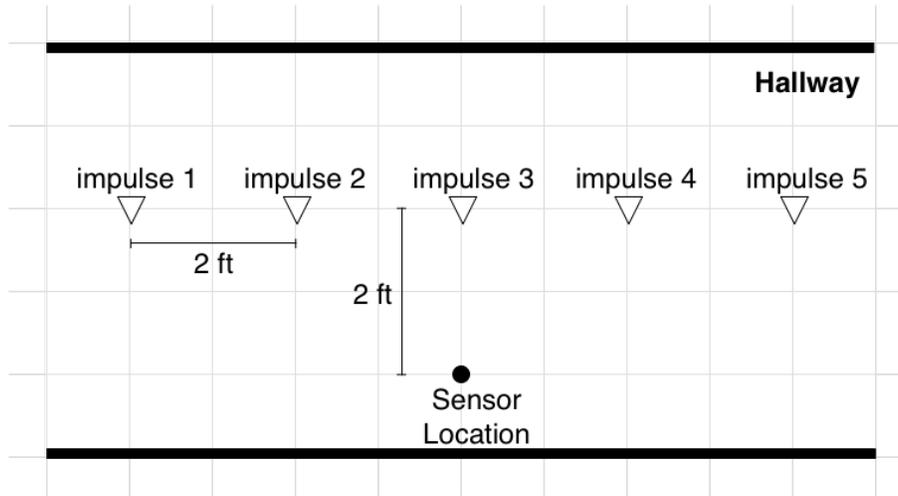
Figure 6.5: Experimental setting of load test with ball drops.

consistency. Therefore, when the impulse signals at different locations demonstrate variation, it is considered as the effects of the structural difference. Based on the structural difference-caused signal variation, the threshold to cluster the SEs in Section 6.1 are selected indicating negligible differences caused by structural variation.

Figure 6.5 shows the experimental setting for the load test, where five impulse locations that are two feet from the adjacent impulses and two feet from the sensor for the impulse 3. The structural variation analysis from the experiments are shown in Figure 6.6. The variation between signals from the same location is relatively low (with an average cross-correlation value of 0.96) as compared to signals from different locations. In different locations, the signal change from one location to another can be large, e.g., the impulse 1 and the impulse 5 have an average cross-correlation value of 0.58. Since the impulse 1 and the impulse 5 have the same distance to the sensor, the variation in the vibration signal is caused by structural variation instead of the dispersion effects.

The average value of the cross-correlation between impulse signals two feet away from each other is 0.87. The signals that are within two feet of each other (the adjacent testing locations) always have over 0.84 cross correlation value, which is then used as the threshold for the hierarchical clustering as aforementioned. This means that the SEs clustered together have less variation than the variation the same impulse applied a step-distance away might have from structural factors, i.e.,

133
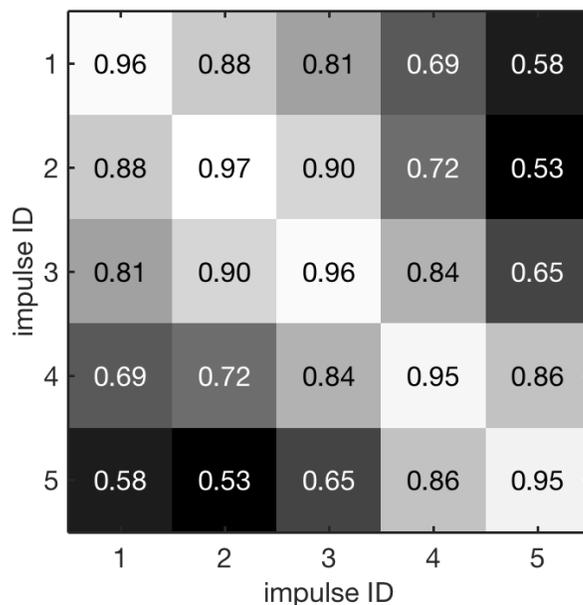
Figure 6.6: Load test results: structural similarity comparison at investigated locations.

the structural variation is negligible within a cluster.

These observations explain two situations observed from the signal characterization in Section 6.1: 1) when a person walks by a sensor, their steps at different locations within the sensing range look different, even when their gait is similar and 2) if a person passes by a location multiple times, the footstep signal from different traces at locations in close proximity look similar. These two observations are the foundation of the structural variation handler in the algorithm.

6.3.1.5.2 Step Frequency: Human Test with Controlled Step Frequency

In order to test the system in a practical setting, a sensing node is deployed in a hallway as shown in Figure 6.7. Participants are asked to walk through the hallway passing by the sensor. 10 students (between the ages of 20 to 29, eight males and two females) volunteered to participate the experiments and walked in a $30ft \times 6ft$ hallway. The deployed sensor can obtain between 10 to 15 steps' signals from each trace signal, and FootprintID selects the seven steps that are closest to the sensor from each trace for analysis.

The participants are told to to step with the metronome beats to control their

Gender	$\mu - 3\sigma$	$\mu - 2\sigma$	$\mu - \sigma$	μ	$\mu + \sigma$	$\mu + 2\sigma$	$\mu + 3\sigma$
Male	95	103	111	119	127	134	142
Female	98	107	116	125	134	143	152

Table 6.1: Metronome frequencies (beats/min)

step frequency. The metronome beats are selected based on the statistics from a survey [95], which are listed in Table 6.1. They were asked to wear flat bottom shoes that they are comfortable with when walking fast and for a long time. Note that in the controlled experiments, the participants are asked to follow the metronome beats as consistently as possible. If the beats are too fast for them to physically follow while walking, they are encouraged to walk as fast as they can instead of jogging or running. This is because that the jogging and running would cause different gait changes compared to walking at different speeds and hence fall out of the scope of this section.

Furthermore, even when the participants are asked to follow the frequency of the metronome, their step frequency might still vary due to the randomness in human behavior. Figure 6.8 (a) shows the detected average step frequency of all participants under different controlled step frequencies (x-axis). These step frequencies are listed in Table 6.1, where the first column is the person ID, the first row displays the relative frequency of the metronome under a setting calculated from prior works [95] to represent the general step frequency range people may apply.
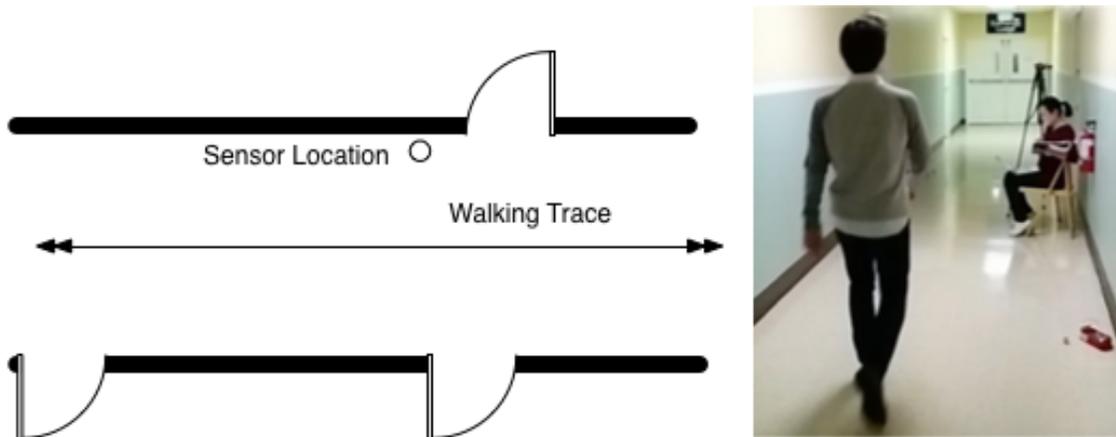
Figure 6.7: The human walking experiment conducted in a hallway.

135
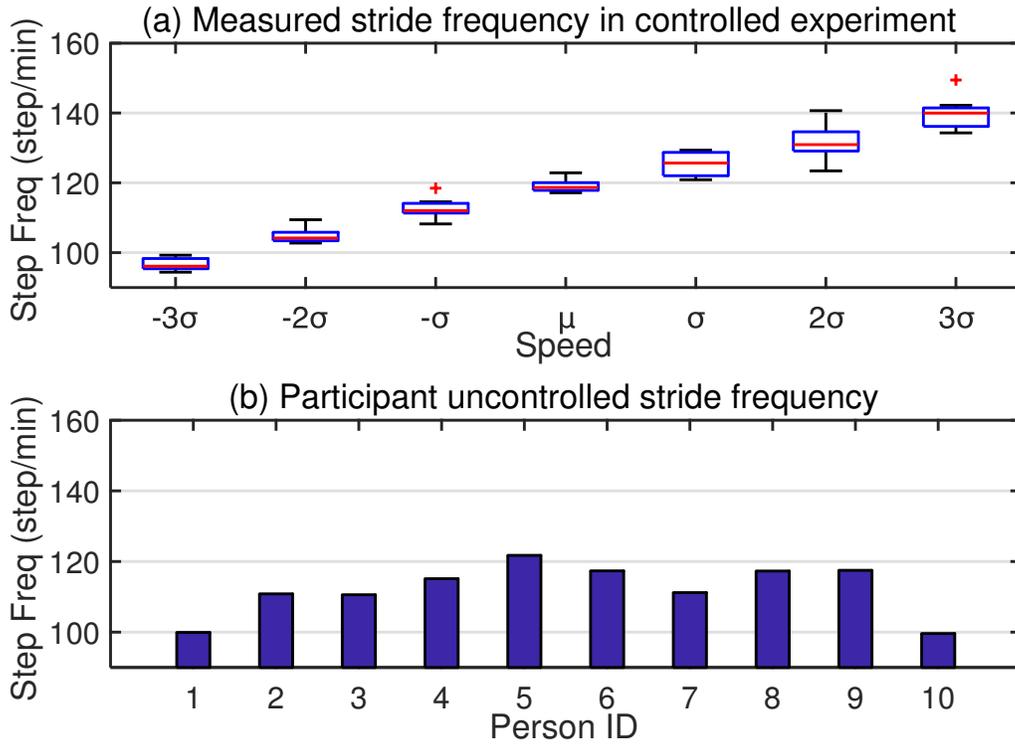
Figure 6.8: Human walking experiment statistics: measured step frequency.

6.3.1.5.3 Uncontrolled Experiments

The uncontrolled experiments were conducted to evaluate system performance when the pedestrians walk at their natural speeds. The uncontrolled experiments are conducted on the same day as the controlled experiments, so that the participant's physical condition as well as the shoes remain the same as the controlled experiments. The participants were asked to walk along the hallway in a way as comfortable as they would like to.

The experiment settings are the same as discussed in Section 6.3.1.5.2, except that the participants are asked to walk based on their natural step frequency. When they are asked to walk in their self-selected pace, their step frequency shows a different range as shown in Figure 6.8 (b). For example, Person 1's self-selected step frequency is around 100 step/min, which is a value between the controlled experiment frequency $\mu - 2\sigma$ and $\mu - 3\sigma$. As shown in Figure 6.1 (b), his SE cluster pattern for self-selected frequency (SpeedID 8) is similar to that of $\mu - 2\sigma$ and $\mu - 3\sigma$ (SpeedID

1 and SpeedID 2), meaning at a specific step frequency, a person's gait is stable possibly due to the individual physical character.

6.3.2   Metrics and Parameters

The prediction accuracy is a common metric used for classification problems. The parameters that may affect the prediction accuracy include 1) the amount of labeled training data and 2) the distribution of the labeled data. Therefore, in the baseline experiments, different training and testing data combinations with the SVM algorithm are compared to understand the properties of the data distribution. All the baseline experiments produce the trace level identification accuracy (majority vote of 7 consecutive SEs from the area that is closest to the sensor) as the final results. There are three main aspects that may affect the identification accuracy, which are respectively 1) the amount of the training data (Section 6.3.2.1); 2) the location of the selected Step Events (Section 6.3.2.2); and 3) the step frequency of the training/testing datasets (Section 6.3.2.3). In this section, these factors are analyzed with the supervised learning method SVM.

6.3.2.1   Amount of the Training Data

The amount of labeled training data affects the SVM classification accuracy. If the training and testing datasets share the same data distribution, the more labeled training data the system has access to, the higher accuracy the system achieves. Figure 6.9 (a) shows the identification accuracy of SVM when the number of labeled training data increases from 1 to 9. Three accuracies are presented under different amount of training data, including 1) step level accuracy, 2) selected step (closest to the sensor) accuracy, and 3) trace level accuracy. The results are of 10 fold cross validation.

When the amount of labeled dataset increases, the accuracies increase. When the model is trained with only 1 trace, the system achieves 60% for step level accuracy, and 80% for trace level accuracy. The selected step shows higher accuracy than that of the overall steps, however it is lower than the trace level decision. The result that considers multiple footsteps achieves significantly higher accuracy. When the amount of the training data increases to 6 traces, the step level accuracy rises to 83%, and the trace level accuracy achieves up to 97%. Both the step and trace level accuracies

137

(a) Relation between # labeled training data and ID accuracy

(b) Stride location variation

(c) Stride frequency variation
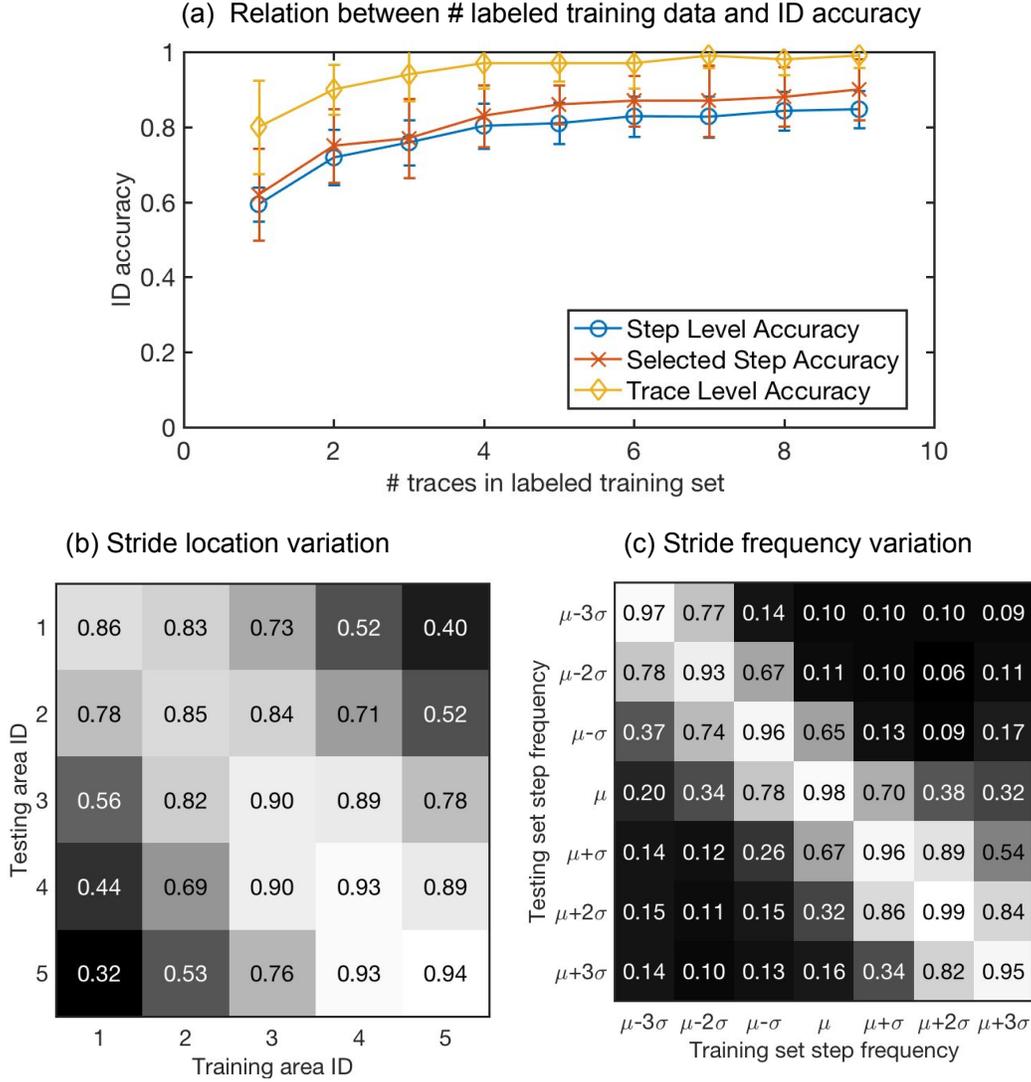
Figure 6.9: Baseline experiments with SVM investigating (a) the amount of labeled training data, (b) stride location variation, and (c) stride frequency variation.

shows no clear increment for the cases where the amount of labeled training data is beyond 6 traces.

6.3.2.2 Step Location of the Training Data

Due to the structural variation, the corresponding locations of the detected SEs affect the data distribution. As aforementioned in Load Test (Section 6.3.1.5.1), when excitation impulses happen in the same area, their signals show consistency

in waveforms. On the other hand, if they are in different areas, their signals show differences in waveforms. As a result, when the training and testing SEs are from the same area, the identification accuracy is higher than those from different areas due to the consistency in waveforms.

To investigate the SE location effect on the learning accuracy, the sensing area for each sensor is further separated into five sub-areas (with three consecutive SEs in each sub-area). Each two consecutive sub-areas are one SE away from each other. The experiments are carried out to train with SEs only from one sub-area and test on the SEs from each sub-area. The identification accuracies are shown in Figure 6.9 (b).

When the training and testing datasets are from the same area based on their approximate localization (diagonal values' average is 90%), the identification accuracy is higher than those when the areas are far away from each other (32% and 40% respectively). In addition, the SEs from areas closer to Area 5 shows a higher accuracy when train and tested on them, compared to those from Area 1. The average accuracy in Area 4 and Area 5 is 93.5% and the average accuracy in Area 1 and Area 2 is 85.5%. This could be caused by a more homogeneous structural condition in Area 5, which leads to a higher classification accuracy. To summarize, selecting the SEs from the same area for modeling and classification allows higher identification accuracy.

6.3.2.3   Step Frequency of the Training Data

Other than SE location, the corresponding step frequencies affects the data distribution as well. Experiments are carried out where the model is built with training data only come from one step frequency, while the test is on SEs from each step frequency separately. Figure 6.9 (c) shows the identification accuracy for SE trained on step frequency listed on x-axis, and tested on SEs from step frequencies listed on y-axis.

The results are shown in Figure 6.9 (c), where the diagonal values in the matrix are the highest through all the rows, reaching an average of 96% accuracy. This result verifies our assumption that the closer the training and testing datasets' step frequencies are, the higher the step identification accuracy is. For each evaluated step frequency, the one level lower/higher frequency shows slight decrease in the identification accuracy, which has an average value of 76%. As for the step frequencies

that fall outside of the $\mu \pm \sigma$ range, the accuracy drops significantly.

Those results indicate that the identification accuracy is higher when the step frequency gap between training and test data is smaller and further motivate us to build our model in an iterative way by ITSVM.

6.3.3 Experiment I: Algorithm Analysis

The controlled experiments evaluate the algorithm robustness especially through testing data with different step frequencies discussed in Section 6.3.1.5.2. In this section, the following aspects of the evaluation are explored: 1) how FootprintID outperforms traditional SVM and how well each component of the algorithm works (e.g., SVM v.s. TSVM v.s. RTSVM v.s. FootprintID), 2) what is the run-time for the aforementioned algorithms components, 3) how well each algorithm components perform with the dataset of the particular data distribution (of each step frequency), 4) how does the thresholding on decision confidence affect the final identification accuracy, and 5) how biased datasets affect the algorithm performances.

6.3.3.1 Algorithm Components Comparison

To understand the contribution of each component in the algorithm, four different settings of the algorithm and their identification accuracies are compared. For the traces of the average step frequency μ, 6 traces are used as the labeled training data and the remaining 4 as testing data. For the traces of other step frequencies, 4 out of 10 are used as the testing data to match the dataset size. The cross validation is conducted 10 times for each test case.

SVM^{light} [57] is used to run SVM and TSVM in the experiments. For SVM in Equation (6.4), the RBF kernel is used, where $\phi(\boldsymbol{x}_i)\phi(\boldsymbol{x}_j) = exp(-\gamma|\boldsymbol{x}_i - \boldsymbol{x}_j|^2)$. The key parameters γ and C are set as 1 and 16 respectively. This parameter set achieves the highest accuracy in the 5-fold cross validation on the labeled dataset.

For TSVM in Equation (6.5), the RBF kernel is used with the key parameters γ and C_1 are 1 and 16. The weight C_2 of the loss function introduced by the unlabeled data is set as 1, which follows the default setting in SVM^{light}. In this case, C_1 is larger than C_2, meaning that the model gives a higher weight to the loss of the labeled data than that of the unlabeled data. For ITSVM, the high confidence threshold of value 70% is used to determine if the unlabeled SEs in a trace can be labeled based

on `RTSVM` results. The threshold is selected empirically that indicates the expansion of the model is based on over 70% similarity between labeled and unlabeled data.

6.3.3.1.1 Case I: Support Vector Machine (SVM)

The system only applies SVM on the labeled training data, and the identification accuracies are shown as the dark blue bars in Figure 6.11 (a). When the model is trained on SEs of the average step frequency μ, the identification accuracy on SEs of the same step frequency achieves as high as 98%. On the other hand, when tested on SEs of rare step frequencies ($\mu \pm 2\sigma$ and $\mu \pm 3\sigma$) data demonstrate tremendous decrease, e.g., the rare step frequencies like $\mu \pm 3\sigma$ drop down to 10% and 16% accuracy. The overall accuracy on datasets of all step frequencies shows an average of 43%. As discussed in Section 6.3.2.3, the more similar the training and testing step frequency, the higher the identification accuracy.

6.3.3.1.2 Case II: Transductive SVM (`TSVM`)

The system applies the transductive SVM without refining/selecting unlabeled data in each pairwise decision. The identification accuracies are shown in Figure 6.11 with light blue bars. For datasets with step frequencies within the range of $\mu \pm \sigma$, the `TSVM` achieves 57% accuracy, which is 20% lower than that of the SVM. For datasets with step frequencies outside the range of $\mu \pm \sigma$, the `TSVM` achieves similar average accuracy compared to that of the SVM, which is 17%.

As aforementioned in Section 6.3.1.3.2, the multi-class SVM is built upon binary-class SVM, and the irrelevant cases overfit the model. Therefore, this inaccurate model reduces the accuracy of the testing identification compared to use SVM. Despite the overfitting, the SEs of rarer step frequencies achieves similar accuracy compared to the similar average accuracy due to the consideration of the unlabeled data in modeling.

6.3.3.1.3 Case III: Refined TSVM (`RTSVM`)

In this case, the system applies the transductive SVM and refines the binary-class `TSVM` by only taking the selective subset of unlabeled data. As aformentioned in Section 6.3.1.3.3, this subset of unlabeled data is selected by applying SVM and choosing the data with corresponding classes.

Figure 6.11 shows the identification accuracy using `RTSVM` with light green bars under different test data scenarios. For SEs with step frequencies within the range of $\mu \pm \sigma$, the `TSVM` achieves 77% accuracy, which is similar to that of the SVM. Since the selective unlabeled data in the binary-class transductive learning process has a higher precision than that in the `TSVM`, the final accuracy increases compared to the `TSVM` as well. For SEs with step frequencies outside the range of $\mu \pm \sigma$, the `RTSVM` outperforms the SVM by an average increment accuracy of 5% due to the incorporation of the unlabeled data characteristics.

6.3.3.1.4   Case IV: Iterative TSVM (`ITSVM`)

In this case, we evaluate the final setting of the presented algorithm. First of all, for traces with step frequencies in the range of the labeled training dataset, the system applies supervised learning (SVM) directly. Then, the system further selects the Step Events with higher confidence in the first iteration of transductive learning to improve the accuracy of the additional labels. The accuracy of the rare step frequency ($\mu \pm 3\sigma$) datasets increases tremendously from 10% to 36% and 16% to 38%, approximately 2.85$\times$ average improvement. The overall accuracy on datasets of all step frequencies shows an average of 62%, 1.5$\times$ improvement compared to that of the SVM.

6.3.3.2   Algorithm Run-time Comparison

The run-time of the aforementioned algorithm components are further compared to estimate the scalability of the system. Table 6.2 shows the average and the standard deviation of the run-time of the step level identification (10 times cross validation).

6.3.3.2.1   Experimental Run-time

The SVM algorithm shows sensitivity on the amount of labeled training data as shown in the first column (trained with only SEs of step frequency μ) and the second column (trained with SEs of all step frequencies). When the labeled training data size increased 7$\times$ (μ v.s. $\mu - 3\sigma...\mu + 3\sigma$), the run-time increases about 9.7$\times$.

When `TSVM` is applied, the amount of the labeled and unlabeled data used to train the model is similar to the SVM trained with SEs of all step frequencies. For this case, the run-time increased almost 40$\times$ compared to that in the second column.

Algorithms	SVM (train: μ)	SVM (train: all)	TSVM	RTSVM	ITSVM (FootprintID)
Runtime Avg. (s)	0.8724	9.7886	382.4303	74.7586	218.9920
Runtime Std. (s)	0.0496	0.9174	94.8886	8.3837	18.9385

Table 6.2: Runtime comparison for 196 testing footsteps over different algorithms [106].

RTSVM, on the other hand, decreases the run-time $5\times$ compared to TSVM by refining the unlabeled data used in modeling. ITSVM (FootprintID) takes about $2.8\times$ the run-time compared to RTSVM because of the iteration. Yet, the run-time of ITSVM is lower than that of the TSVM.

Then for ITSVM, the average run-time to identify each SE is evaluated in Figure 6.10. Figure 6.10 (a) compares the average run-time per SE when the number of classes increases. The average run-time increases almost linearly with the increase of the number of classes. When there are two iterations, the increase ratio is almost $3\times$ of that in the first iteration.

Figure 6.10 (b) shows that the total run-time increases when the number of unlabeled data increases. However, the average run-time for each SE decreases for the first iteration and remains stable for the case of two iterations. Although the total run-time is a monotonic increasing function of the amount of unlabeled data, the average run-time for unlabeled SEs is not. The increasing or decreasing trend is determined by the ratio of labeled and unlabeled data.

On the other hand, Figure 6.10 (c) shows the relation between the average run-time and the number of labeled data provided. For both the first iteration and two iterations cases, the average run-time per SE increases with the increase of the labeled training data.

6.3.3.2.2 Time Complexity Analysis

The time complexity of the aforementioned algorithm components are analyzed in this section. Note that the number of labeled data samples is l, the number of unlabeled data samples is u, and there are n users to be classified in the system. Each experiment, the local search algorithm runs T repetitions in TSVM and runs k iterations in ITSVM.

In SVM^{light}, optimizing the loss function of TSVM is a combinatorial optimization
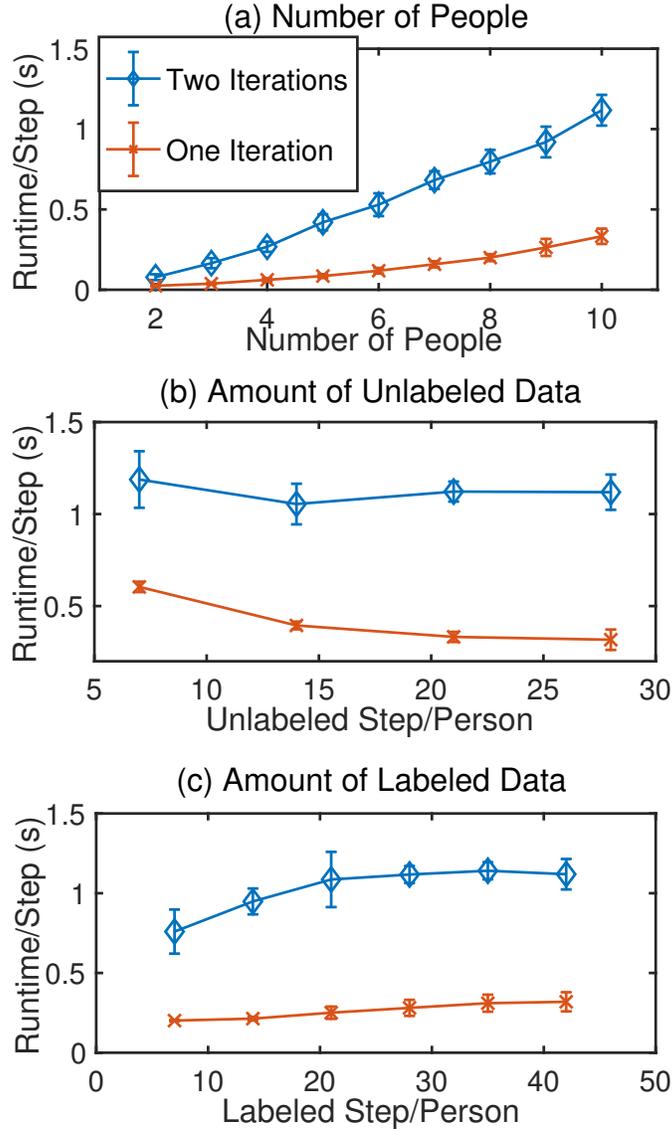
Figure 6.10: Average runtime for each tested footstep over different (a) number of class, (b) amount of unlabeled data, and (c) amount of labeled data.

problem, which is solved by a local search algorithm. In each repetition of the local search, an SVM is trained by the quadratic programming. In that case, the time complexity of `TSVM` and `RTSVM` is $\mathcal{O}(T(l+u)^2 n^2)$. The time complexity of `ITSVM` is $\mathcal{O}(kT(l+u)^2 n^2)$. In contrast, the time complexity of n-class SVM is $\mathcal{O}(l^2 n^2)$.

Based on this time complexity analysis, we further explain the runtime shown in
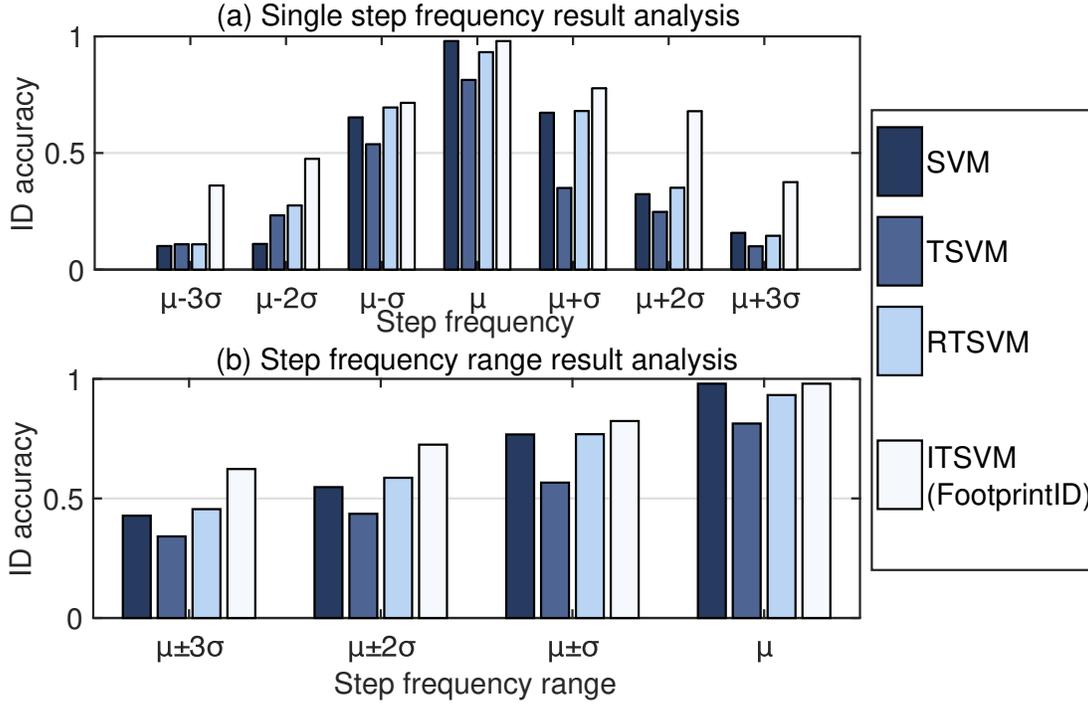
Figure 6.11: Comparison of identification accuracy between different algorithms [106].

Table 6.2. When compared to the n-class SVM, TSVM takes T repetitions as opposed to 1 for SVM. In addition, the amount of the unlabeled data u makes each repetition longer. Therefore, TSVM takes multiple times the run-time compared to that of SVM. RTSVM, on the other hand, has a smaller amount of unlabeled data compared to TSVM since its data is selected, resulting in a lower run-time.

Compared to RTSVM, ITSVM (FootprintID) has a higher run-time because ITSVM takes more iterations to train the final model. ITSVM (FootprintID) has a lower run-time compared to TSVM, because ITSVM also only uses partial unlabeled data. Despite the run-time increase due to the calculation repetitions introduced by the TSVM, the time and cost to collect labeled data, which takes hours, is more than the run-time difference shown here.

6.3.3.3 Step Frequency Accuracy Comparison

Since the step frequency is the main factor that affects the algorithm accuracy, the identification accuracies with testing SEs of different step frequencies are compared here. Figure 6.11 shows the ID accuracy of aforementioned algorithm components

under different testing SE distribution.

Figure 6.11 (a) shows a tremendous drop in identification accuracy from $\mu \pm \sigma$ to $\mu \pm 2\sigma$ when supervised learning method (SVM) is used. The TSVM and RTSVM shows similar accuracy decreasing trends as the SVM at the extreme step frequencies ($\mu \pm 3\sigma$). However, compared to SVM, RTSVM shows an improvement at the step frequencies of $\mu \pm 2\sigma$. The ITSVM (FootprintID) shows a significant improvement compared to all the other algorithm components expecially in the extreme step frequencies of $\mu \pm 3\sigma$.

Figure 6.11 (b) presents the average identification accuracy within the step frequency range listed in the x-axis as a demonstration of the overall identification accuracy on SEs of different step frequencies. The performance of the supervised learning (SVM) shows better accuracies when compared to RTSVM when the step frequencies of the datasets are closer to the labeled training data. This is because the supervised model describes the same type of dataset more accurately. The improvement of ITSVM (FootprintID) compared to SVM is larger when the difference between testing and training data's step frequencies is larger (e.g., testing on $\mu \pm 3\sigma$ and training on μ).

6.3.3.4   Decision Confidence and Thresholding

In many applications, identifying users as unknown is more tolerable than making wrong predictions. Therefore, FootprintID further utilizes the decision confidence calculated for each trace to determine whether to output an ID from the labeled dataset or as unknown. The *trace-level* decision confidence for a trace is calculated as discussed in Section 6.3.1.4. Figure 6.12 demonstrates the identification accuracy after thresholding on the *trace-level* decision confidence of the ITSVM prediction shown in Figure 6.11 (b). When the threshold is not applied (Confidence threshold is 0), all trace decisions are considered confident and the identification accuracy is as discussed earlier in *Case IV: Iterative TSVM (*ITSVM*)*. For the SEs within the distribution range of $\mu \pm 3\sigma$, when the threshold value is increased to 0.227, the ratio of the traces labeled 'unknown' goes up to 50%, and the accuracy increases to 83% from 62% without thresholding. When the unlabeled data distribution range decreases, i.e., when not considering rare step frequency data, the corresponding identification accuracy as well as the confident ID ratio increase. Figure 6.12 shows

146
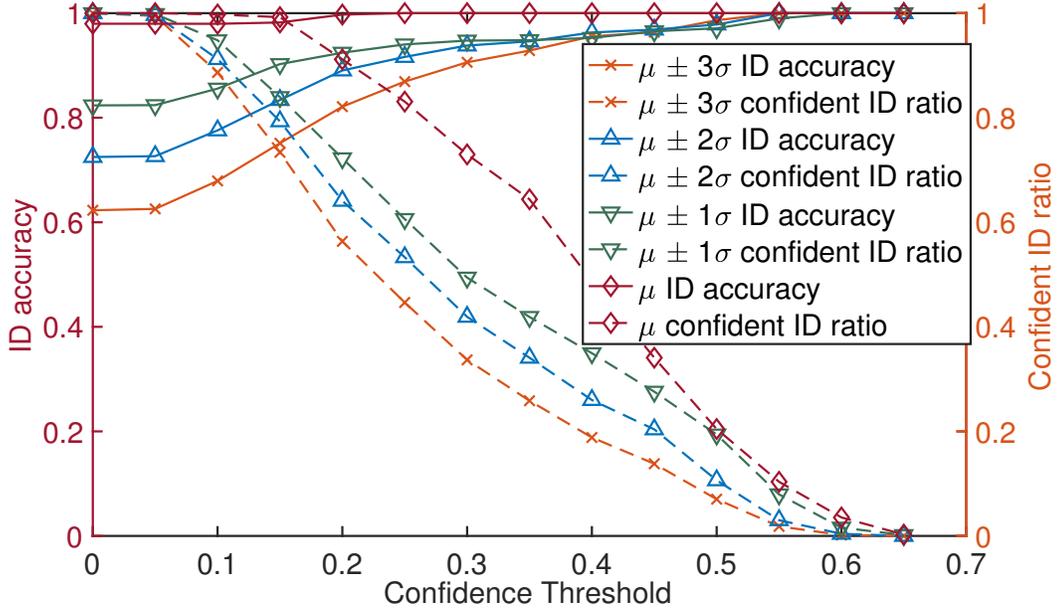
Figure 6.12: Decision confidence thresholding on `ITSVM` results from Figure 6.11 (b) [106].

the SEs within the distribution range of $\mu \pm 2\sigma$, $\mu \pm \sigma$, and μ with blue, green and purple lines respectively. It also means that the error estimation in the rare step frequency data is of low confidence, and the thresholding effectively mitigates the error estimation.

6.3.3.5 Dataset with Biased Size

In this section, we discuss the system robustness when the unlabeled data used for updating the transductive learning model has biased size. This scenario could happen when the participants appear in the sensing area at different frequencies. The baseline for this experiment is SVM, and all the available traces are used. We evaluate the bias dataset effects using the accuracy reduction ratio compared to this baseline.

In general, the dataset of a specific step frequency of a participant can be insufficient or sufficient. The definition of insufficient and sufficient indicates the *degree of bias*. When the *number of people with sufficient dataset* is high, the identification accuracy should increase. When the dataset of different *step frequencies* have a different bias, e.g., only the rare step frequencies are biased, the identification

accuracy varies too.

In the rest of this section, we, first of all, select the biased step frequencies based on different scenarios. Then we go through the different combination of degree of bias and the number of people with sufficient datasets.

6.3.3.5.1   Bias Variable I: Degree of Bias

We define the degree of bias as the definition of a sufficient and insufficient unlabeled dataset for each participant. That is, how much bias appears between the participants who show up in the sensing area most often? Here we investigate the cases when the unlabeled data ratios are are 1:8, 2:7, 3:6, and 4:5 for people who show up less often (having insufficient dataset sizes) v.s. those who show up often (having sufficient dataset sizes).

6.3.3.5.2   Bias Variable II: Number of People with Sufficient Dataset

The scenarios where some of the participants have a sufficient unlabeled dataset, while the rest have an insufficient unlabeled dataset are evaluated in this section. I.e., how many people show up often (having sufficient dataset sizes)? For example, when only a small number of people have a sufficient unlabeled dataset, the system may have lower accuracy comparing to when a large number of people have a sufficient unlabeled dataset.

6.3.3.5.3   Bias Variable III: Step Frequencies

When different step frequencies are biased differently, will the system performance be affected by the difference in the bias as well? That is, which iteration of biased datasets affects the system performance more?

6.3.3.5.4   Analysis on Data Bias

Figure 6.13 demonstrates the comparison of the identification accuracy regarding these three factors. As shown in both (a) and (b), when the gap between sufficient and insufficient are large (insufficient:sufficient dataset size is 1:8), the identification accuracy is lower for all cases. When even number people have insufficient and sufficient datasets, the accuracy actually drops when the degree of bias is large. This
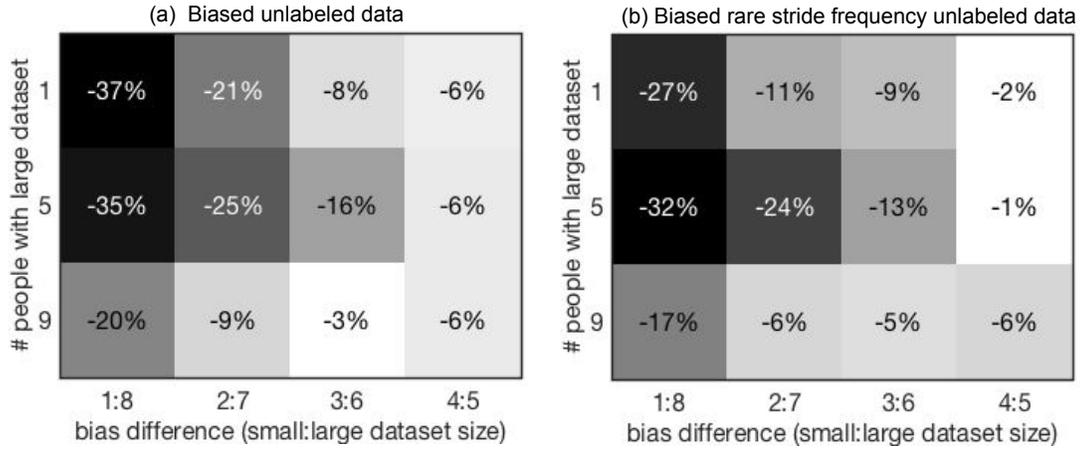
Figure 6.13: Identification accuracy drop of biased datasets compared to fair datasets [106]. (a) All step frequency unlabeled data is biased between participants. (b) The rare step frequency unlabeled data is biased between participants.

is because when the size of a dataset is too small, the estimation tends to have a higher error rate, which is expected from the analysis in Section 6.3.2.1.

The accuracy of insufficient and sufficient datasets respectively is further analyzed to made the following observations: 1) the sufficient dataset cases have similar accuracy; 2) for the insufficient dataset cases, the accuracy increases when the number of people with sufficient datasets increases. The later happens because when more people are identified accurately, the error for the rest of the decision decreases as well. Hence, due to the portion of the insufficient and sufficient dataset, the second row (5 people with insufficient datasets and 5 people with sufficient datasets) is lower than the first row (1 person with a sufficient dataset and 9 people with insufficient datasets).

On the other hand, the third column (bias 3:6) in the third row (1 person with an insufficient dataset and 9 people with sufficient datasets) outperforms the second (bias 2:7) and the fourth column (bias 4:5) in the third row. This could be caused by a higher accuracy of the majority population (9 people with sufficient datasets) when compared to the fourth column, and a higher accuracy of the minority population (1 people with insufficient datasets) when compared to the second column.

When comparing the cases in which all unlabeled datasets are biased by a person (Figure 6.13 (a)) v.s. only the rare step frequencies dataset are biased by a person (Figure 6.13 (b)), the latter achieves higher accuracy especially when the degree of

bias is high and the number of people with sufficient datasets is low. This is because when the first iteration of datasets is sufficient, the estimation has higher accuracy and therefore the new labeled data has a label with higher accuracy, hence the model describes the data more accurately.

When the dataset is biased, the identification accuracy can be further improved by tuning the confidence threshold. As discussed in Figure 6.12, when the unlabeled dataset is large enough, the threshold used to select additional labels from the RTSVM results can be increased. The more accurate the first iteration model is, the more accurate the final results will be, as shown in the comparison between cases in Figure 6.13 (a) and (b).

6.3.4   Experiment II: Algorithm Robustness

FootprintID is further evaluate with uncontrolled experiments where pedestrians walk through the sensing area one at a time with their natural walking form. The uncontrolled experiment is conducted with the unlabeled data used in the controlled experiment as the observed historical unlabeled data (Section 6.3.1.3.4) and add the uncontrolled data set as the new unlabeled data.

As discussed in Section 6.3.1.5.2, when the participants are asked to walk with their natural form, their step frequencies vary as shown in Figure 6.7. Four algorithms (SVM, TSVM, RTSVM and ITSVM) are studied on the uncontrolled experiment data. Average identification accuracy values are compared in Table 6.3.

Three modeling settings are investigated and listed in the table: 1) the system takes the labeled data with step frequency μ and unlabeled historical data with step frequency between $\mu \pm 3\sigma$ as well as the uncontrolled data to build the model, 2)

Models	SVM	TSVM	RTSVM	ITSVM (FootprintID)
labeled: μ, unlabeled: $\mu \pm \sigma$, $\mu \pm 2\sigma$, $\mu \pm 3\sigma$, uncontrolled	56%	52%	52%	67%
labeled: μ, unlabeled: $\mu \pm \sigma$, $\mu \pm 2\sigma$, $\mu \pm 3\sigma$	56%	54%	51%	67%
labeled: μ, unlabeled: uncontrolled	50%	22%	22%	45%

Table 6.3: Classification accuracy comparison for 196 testing footsteps over different algorithms [106].

the system takes the labeled data with step frequency μ and unlabeled historical data with step frequency between $\mu \pm 3\sigma$ to build the model, and 3) the system takes the labeled data with step frequency μ and the uncontrolled data to build the model. These three settings are representatives of different stages of the system. The first modeling represents the stage where the model is still growing. The second modeling is the closest to the system after a bootstrap phase. The third modeling is the stage when there is not enough historical data yet. When the model covers SEs with enough range of the walking step frequency, the extra unlabeled data does not necessarily increase the identification accuracy (67% for the first and the second model). In addition, when there is not enough gradually changing historical data like that shown in Figure 6.4, the system may end up with a lower accuracy when compared to SVM (45% for the third model).

When using `ITSVM`, the system determines the step frequency range that applies supervised SVM as between 116 and 122 steps/min, which is the range of the labeled data with walking speed μ shown in Figure 6.7 (c). Step frequency values of the SEs from uncontrolled traces spread out between $\mu - 3\sigma$ and $\mu + \sigma$. Therefore, the improvement on the `ITSVM` is mainly due to the higher identification accuracy on the low step frequency data.

The 'unknown thresholding' changing and its effects on the identification accuracy using `ITSVM` is demonstrated in Figure 6.14. For certain applications, the system can report the identify only when the prediction confidence score is beyond a pre-defined threshold. Otherwise, the system predicts the identify as "unknown" [104]. Figure 6.14 shows the corresponding identification accuracy under different thresholds. When FootprintID thresholds on half of the high confidence trace, the identification accuracy rises from 67% to 90%.

6.4   Related Work

In this section, prior works related to the ILPC algorithm and FootprintID system implementation, including gait kinetic (Section 6.4.1), person identification sensing methods (Section 6.4.2), and transductive transfer learning (Section 6.4.3), are discussed.
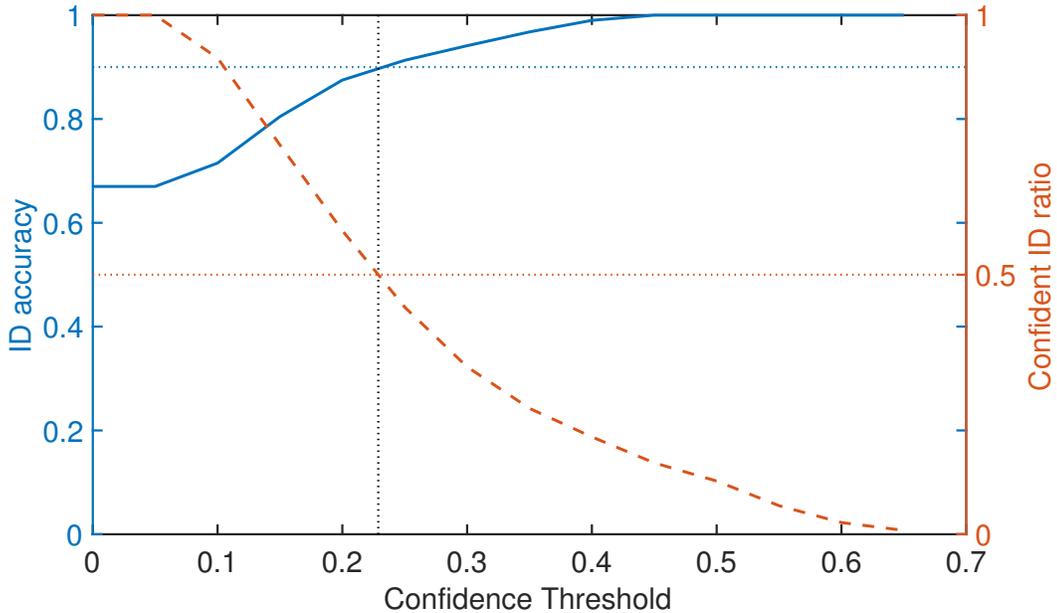
Figure 6.14: The accuracy of the uncontrolled experiment in which people walk at their natural frequency [106].

6.4.1 Gait Kinetics

Kinetics of human gaits usually refers to the study on 1) forces passing through the joints, 2) force plate embedded in the floor records, and 3) ground reaction force vectors. Study have shown that the force that a human subject applies to the ground can be used for identification of animals and human beings [141]. Further investigation on the ground's reaction to the force of indoor human footsteps also indicates the possibility of identifying pedestrians through the measurement of the force applied on the floor [83, 96]. The excitation force induces shape changing on the pressured part of the floor and vibrations of the structure [13]. The prior work focuses either on the gait's ground reaction force or the structural vibration induced by an excitation force, and the connection between these two are missing. We focus on the footstep-induced vibration sensing, which 1) allows sparse deployment due to the wave propagation properties in structure and 2) captures the gait kinetic information for identification purposes.

6.4.2 Person Identification Sensing Methods Comparison

Various of sensing methods have been explored to obtain person identification in the indoor environment. They mainly fall into a few categories, including vision [24, 98, 139, 151], RF [155, 164], mobile [90, 144], inertial sensors [37, 80, 124], acoustic [9, 39], and vibration [32, 104] based methods.

Vision-based methods extract visual biometrics of an individual, including facial structure [12, 24], hand geometry [98, 128], and gait [139, 151]. Although the accuracy of the identification is high, the systems often require a clear visual path. This makes it a a challenge to utilize them in a surveillance scenario or ubiquitous sensing without requiring them to interact with a sensing system. On the other hand, RF based methods overcome such problems, and have been explored to obtain a variety of information about people in an indoor environment [155, 164]. This type of method often requires the sensing target to be between a pair of transmitters, therefore they require dense deployment to achieve high accuracy information inquiry. Furthermore, such data-driven approaches usually require a large labeled training set, which makes the deployment of the system a challenge.

Mobile-based methods [90, 144] utilize the relation between the sensing target and the device with a unique ID, and assume each target will carry the registered device. Another way to identify pedestrians with mobile devices uses inertial sensor based methods [37, 80, 124], which is through the gait information reflected on the motion. However, these mobile device related methods require wearable devices on the pedestrian to obtain the gait information. Therefore, it might be difficult to use when deployed in applications such as elderly care or patient care, in which people may not carrying their phones all the time.

Acoustic- and vibration-based methods have their similarities [32, 104]. Compared to acoustic sensing, vibration sensing is less sensitive to the audible noise in the environment. However, both are data-driven approaches and require a large amount of the labeled training data to achieve robust identification. In this paper, we presented our system FootprintID, a vibration-based method that takes physical insights into account and allows robust identification with a limited amount of labeled training data.

6.4.3   Transductive Transfer Learning

Transductive learning is a technique that addresses the distribution change between training and test data, including self-training, low-density separation, and graph-based methods [20]. Low-density separation methods, such as Transductive SVM (`TSVM`) [20, 58], tend to place decision boundaries where the unlabeled data has low density. The graph-based methods (`GSSL`) [171] construct a graph where the nodes are the labeled and unlabeled data points. In addition to having good accuracy in the labeled training data, the graph-based methods learn a model where the prediction of unlabeled data is smooth in the constructed graph. Previous works [16, 72, 73, 74, 148] show that traditional transductive learning methods such as `TSVM` and `GSSL`, may have decreased accuracy with unlabeled data in the model training. To avoid accuracy decrease, unlabeled instances are selectively used in the transductive learning [72, 74, 148].

6.5   Chapter Summary

In this chapter, an iterative learning method that takes physical attributes into account is presented to tackle the challenge of data distribution gradually change between the training and testing datasets. The algorithm iterative applies the transfer learning that can handle limited data distribution changes to the selected unlabeled data. The selection of the unlabeled data is based on the measurable/sensible physical attributes, which is the key to achieving high learning accuracy with labeled data of limited data distributions.

An indoor pedestrian identification system – FootprintID – is then presented as an implementation of this physical attributes guided iterative learning methods. The physical attributes guided iterative learning achieves $1.5\times$ increase in identification accuracy on all footstep frequency data and $3\times$ increase on identification accuracy on extreme footstep frequency data.

Chapter 7

Conclusion

In conclusion, this thesis introduced a structural vibration based indoor human information acquisition system. The system utilizes the structures as the extension of the sensor to achieve non-intrusive sparse information acquisition on people in the indoor environment. The challenges of utilizing this sensing method comes from the complexity of the physical world and can be summarized to three folds: 1) obtain high fidelity sensing signal for detailed information learning, 2) conduct accurate signal characterization to extract precise features that describe the information, and 3) learning accurate human information with limited labeled data.

To handle these challenges, a methodology is introduced in this thesis: utilizing physical insights to guide the sensing and learning process – the understanding of the physical world (i.e., structure and human in our case) can be used to enhance each phase of the information acquisition process. For signal acquisition phase, the model of wave attenuation and human walking trajectory continuity are used to predict the footstep induced vibration signal amplitude and the optimal hardware configurations for it. For signal characterization phase, the wave properties are used to guide the selection of different algorithms that take the signal properties into account when extract features/information. For learning phase, the physical attributes behind the data distribution change are used to guide the learning order to ensure the accuracy of each iteration of learning. The physical insights play an important rule in each phase of the human information acquisition system and assist the system to achieve high accuracy.

Bibliography

[1] Heba Abdelnasser, Moustafa Youssef, and Khaled A Harras. Wigest: A ubiquitous wifi-based gesture recognition system. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 1472–1480. IEEE, 2015. Cited on page 12.

[2] GG Adams. Radiation of body waves induced by the sliding of an elastic half-space against a rigid surface. *Journal of applied mechanics*, 67(1):1–5, 2000. Cited on pages 66 and 67.

[3] Fadel Adib, Zachary Kabelac, Dina Katabi, and Robert C Miller. 3d tracking via body radio reflections. In *NSDI*, volume 14, pages 317–329, 2014. Cited on page 12.

[4] Yuvraj Agarwal, Bharathan Balaji, Rajesh Gupta, Jacob Lyles, Michael Wei, and Thomas Weng. Occupancy-driven energy management for smart building automation. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, pages 1–6. ACM, 2010. Cited on page 113.

[5] Ian F Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. A survey on sensor networks. *IEEE Communications magazine*, 40(8):102–114, 2002. Cited on pages 10 and 11.

[6] Ian F Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci. Wireless sensor networks: a survey. *Computer networks*, 38(4):393–422, 2002. Cited on page 11.

[7] Ibrahim Al-Naimi and Chi Biu Wong. Indoor human detection and tracking using advanced smart floor. In *Information and Communication Systems (ICICS), 2017 8th International Conference on*, pages 34–39. IEEE, 2017.

Cited on page 13.

[8] Pär-Anders Albinsson and Shumin Zhai. High precision touch screen interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 105–112. ACM, 2003. Cited on page 13.

[9] David T Alpert and Martin Allen. Acoustic gait recognition on a staircase. In *World Automation Congress (WAC), 2010*, pages 1–6, New York, USA, 2010. IEEE. Cited on page 153.

[10] Dawn An, Nam H Kim, and Joo-Ho Choi. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliability Engineering & System Safety*, 133:223–236, 2015. Cited on page 16.

[11] Th Arampatzis, John Lygeros, and Stamatis Manesis. A survey of applications of wireless sensors and wireless sensor networks. In *Intelligent Control, 2005. Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation*, pages 719–724. IEEE, 2005. Cited on page 10.

[12] WA Aulsebrook, MY İşcan, JH Slabbert, and P Becker. Superimposition and reconstruction in forensic facial identification: a survey. *Forensic Science International*, 75(2-3):101–120, 1995. Cited on page 153.

[13] Hugo Bachmann. Case studies of structures with man-induced vibrations. *Journal of Structural Engineering*, 118(3):631–647, 1992. Cited on page 152.

[14] Hugo Bachmann and Walter Ammann. *Vibrations in structures: induced by man and machines*, volume 3. Iabse, Zurich, Switzerland, 1987. Cited on pages 10, 66, and 69.

[15] Aaron Barzilai. Improving a geophone to produce an affordable broadband seisometer. *Mechanical Engineering, Stanford University Jan.*, 25, 2000. Cited on page 62.

[16] Mikhail Belkin and Partha Niyogi. Semi-supervised learning on Riemannian manifolds. *Machine learning*, 56(1-3):209–239, 2004. Cited on page 154.

[17] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992. Cited on pages 126

and 127.

[18] Michael S Brandstein and Harvey F Silverman. A robust method for speech signal time-delay estimation in reverberant rooms. In *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, volume 1, pages 375–378. IEEE, 1997. Cited on page 70.

[19] Chih-Chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011. Cited on page 127.

[20] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542, 2009. Cited on pages 128, 131, and 154.

[21] Ke-Yu Chen, Daniel Ashbrook, Mayank Goel, Sung-Hyuck Lee, and Shwetak Patel. Airlink: sharing files between multiple devices using in-air gestures. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 565–569. ACM, 2014. Cited on page 13.

[22] Tsong-Yi Chen, Chao-Ho Chen, Da-Jinn Wang, and Yi-Li Kuo. A people counting system based on face-detection. In *Genetic and Evolutionary Computing (ICGEC), 2010 Fourth International Conference on*, pages 699–702. IEEE, 2010. Cited on page 113.

[23] SungJung Cho, Jong Koo Oh, WonChul Bang, Wook Chang, Eunseok Choi, Yang Jing, Joonkee Cho, and Dong Yoon Kim. Magic wand: a hand-drawn gesture input device in 3-d space with inertial sensors. In *Frontiers in Handwriting Recognition, 2004. IWFHR-9 2004. Ninth International Workshop on*, pages 106–111. IEEE, 2004. Cited on page 12.

[24] Jeffrey S Coffin and Darryl Ingram. Facial recognition system for security access and identification, November 23 1999. US Patent 5,991,429. Cited on page 153.

[25] Yang Cong, Haifeng Gong, Song-Chun Zhu, and Yandong Tang. Flow mosaicking: Real-time pedestrian counting without scene-specific learning. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1093–1100. IEEE, 2009. Cited on page 113.

[26] Marco Costa. Interpersonal distances in group walking. *Journal of Nonverbal*

Behavior, 34(1):15–26, 2010. Cited on pages 76 and 103.

[27] Thomas M Cover and Peter E Hart. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27, 1967. Cited on page 103.

[28] Dwm 1000 ieee 802.15.4-2011 uwb transceiver module. `https://www.decawave.com/sites/default/files/dwm1000-datasheet-v1.6.pdf`, 2016. Accessed: 2018-04-02. Cited on page 25.

[29] Chasity DeLoney. Person identification and gender recognition from footstep sound using modulation analysis. Technical report, DRUM, 2008. Cited on page 15.

[30] Vasant Dhar. Data science and prediction. *Communications of the ACM*, 56(12):64–73, 2013. Cited on page 16.

[31] Klaus Dobbler, Moritz Fišer, Maria Fellner, and Bernhard Rettenbacher. Vibroacoustic monitoring: Techniques for human gait analysis in smart homes. In *Ambient Assisted Living*, pages 47–58. Springer, 2014. Cited on page 62.

[32] Alexander Ekimov and James M Sabatier. Vibration and sound signatures of human footsteps in buildingsa). *The Journal of the Acoustical Society of America*, 120(2):762–768, 2006. Cited on pages 125 and 153.

[33] Varick L Erickson, Yiqing Lin, Ankur Kamthe, Rohini Brahme, Amit Surana, Alberto E Cerpa, Michael D Sohn, and Satish Narayanan. Energy efficient building environment control strategies using real-time occupancy measurements. In *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, pages 19–24. ACM, 2009. Cited on page 113.

[34] Jonathon Fagert, Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. Characterizing left-right gait balance using footstep-induced structural vibrations. In *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2017*, volume 10168, page 1016819. International Society for Optics and Photonics, 2017. Cited on page 115.

[35] Francois Fleuret, Jerome Berclaz, Richard Lengagne, and Pascal Fua. Multi-camera people tracking with a probabilistic occupancy map. *IEEE transactions on pattern analysis and machine intelligence*, 30(2):267–282, 2008. Cited on

pages 12 and 13.

[36] Vc mems vs piezo sensors. https://techzone.colibrys.com/sensor-performance/vc-mems-vs-piezo-sensors/, 2015. Accessed: 2017-01-24. Cited on page 14.

[37] Davrondzhon Gafurov, Einar Snekkenes, and Patrick Bours. Gait authentication and identification using wearable accelerometer sensor. In *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on*, pages 220–225, Alghero, Italy, 2007. IEEE. Cited on page 153.

[38] Alexander Gammerman, Volodya Vovk, and Vladimir Vapnik. Learning by transduction. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 148–155, Madison, Wisconsin, USA, 1998. Morgan Kaufmann Publishers Inc. Cited on page 126.

[39] Jürgen T Geiger, Maximilian Kneißl, Björn W Schuller, and Gerhard Rigoll. Acoustic gait-based person identification using hidden markov models. In *Proceedings of the 2014 Workshop on Mapping Personality Traits Challenge and Workshop*, pages 25–30. ACM, 2014. Cited on pages 15 and 153.

[40] Farhad Ghassemi and Vikram Krishnamurthy. A cooperative game-theoretic measurement allocation algorithm for localization in unattended ground sensor networks. In *Information Fusion, 2008 11th International Conference on*, pages 1–7. IEEE, 2008. Cited on page 63.

[41] Mayank Goel, Brendan Lee, Md Tanvir Islam Aumi, Shwetak Patel, Gaetano Borriello, Stacie Hibino, and Bo Begole. Surfacelink: using inertial and acoustic sensing to enable multi-device interaction on a surface. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 1387–1396, New York, NY, USA, 2014. ACM. Cited on page 15.

[42] King-Shy Goh, Edward Chang, and Kwang-Ting Cheng. Svm binary classifier ensembles for image classification. In *Proceedings of the Tenth International Conference on Information and Knowledge Management (CIKM)*, pages 395–402, Atlanta, GA, USA, 2001. ACM. Cited on page 128.

[43] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org. Cited on page 126.

[44] RD Gregory and I Gladwell. The reflection of a symmetric rayleigh-lamb wave

161

at the fixed or free edge of a plate. *Journal of Elasticity*, 13(2):185–206, 1983. Cited on page 10.

[45] Sidhant Gupta, Daniel Morris, Shwetak Patel, and Desney Tan. Soundwave: using the doppler effect to sense gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1911–1914. ACM, 2012. Cited on page 13.

[46] RF Guratzsch and S Mahadevan. Sensor placement design for shm under uncertainty. In *Proc. 3rd European Workshop: Structural Health Monitoring*, pages 1168–75, 2006. Cited on page 62.

[47] Jun Han, Madhumitha Harishankar, Xiao Wang, Albert Jin Chung, and Patrick Tague. Convoy: Physical context verification for vehicle platoon admission. In *18th International Workshop on Mobile Computing Systems and Applications (HotMobile)*, pages 73–78, Sonoma California, Feb 2017. ACM. Cited on page 113.

[48] Chris Harrison and Scott E Hudson. Scratch input: creating large, inexpensive, unpowered and mobile finger input surfaces. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pages 205–208, New York, NY, USA, 2008. ACM. Cited on page 15.

[49] Chris Harrison, Julia Schwarz, and Scott E Hudson. Tapsense: enhancing finger interaction on touch surfaces. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 627–636, Santa Barbara, California, 2011. ACM. Cited on page 15.

[50] Chikio Hayashi. What is data science? fundamental concepts and a heuristic example. In *Data Science, Classification, and Related Methods*, pages 40–51. Springer, 1998. Cited on page 16.

[51] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. *The International Journal of Robotics Research*, 31(5):647–663, 2012. Cited on page 69.

[52] Timothy W Hnat, Erin Griffiths, Ray Dawson, and Kamin Whitehouse. Door-jamb: unobtrusive room-level tracking of people in homes using doorway sensors. In *Proceedings of the 10th ACM Conference on Embedded Network*

Sensor Systems, pages 309–322. ACM, 2012. Cited on page 113.

[53] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 13(2):415–425, 2002. Cited on pages 127 and 128.

[54] Geophone sm-24. https://www.sparkfun.com/products/11744, 2006. Accessed: 2017-01-24. Cited on pages 22 and 44.

[55] AJEM Janssen, R Veldhuis, and L Vries. Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 34(2):317–330, 1986. Cited on page 62.

[56] Zhenhua Jia, Musaab Alaziz, Xiang Chi, Richard E Howard, Yanyong Zhang, Pei Zhang, Wade Trappe, Anand Sivasubramaniam, and Ning An. Hb-phone: a bed-mounted geophone-based heartbeat monitoring system. In *Information Processing in Sensor Networks (IPSN), 2016 15th ACM/IEEE International Conference on*, pages 1–12, Vienna, Austria, 2016. IEEE. Cited on pages 15, 113, and 115.

[57] T. Joachims. Making large-scale SVM learning practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, chapter 11, pages 169–184. MIT Press, Cambridge, MA, 1999. Cited on page 140.

[58] Thorsten Joachims. Transductive inference for text classification using support vector machines. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pages 200–209. Morgan Kaufmann Publishers Inc., 1999. Cited on page 154.

[59] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967. Cited on page 116.

[60] Philo Juang, Hidekazu Oki, Yong Wang, Margaret Martonosi, Li Shiuan Peh, and Daniel Rubenstein. Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet. *ACM SIGARCH Computer Architecture News*, 30(5):96–107, 2002. Cited on page 11.

[61] Amir Kale, Naresh Cuntoor, B Yegnanarayana, AN Rajagopalan, and Rama Chellappa. Gait analysis for human identification. In *International Conference*

on Audio-and Video-Based Biometric Person Authentication, pages 706–714. Springer, 2003. Cited on pages 12 and 13.

[62] Jon Kerridge, Alistair Armitage, David Binnie, Lucy Lei, and Neil Sumpter. Monitoring the movement of pedestrians using low-cost infrared detectors: initial findings. In *Proceedings US Transport Research Board Annual Meeting*, 2004. Cited on page 113.

[63] Zaheer Khan, Janne Lehtomaki, Kenta Umebayashi, and Johanna Vartiainen. On the selection of the best detection performance sensors for cognitive radio networks. *IEEE Signal Processing Letters*, 17(4):359–362, 2010. Cited on page 63.

[64] Sukun Kim, Shamim Pakzad, David Culler, James Demmel, Gregory Fenves, Steven Glaser, and Martin Turon. Health monitoring of civil infrastructures using wireless sensor networks. In *Information Processing in Sensor Networks, 2007. IPSN 2007. 6th International Symposium on*, pages 254–263. IEEE, 2007. Cited on page 14.

[65] Srdjan Kitic, Laurent Jacques, Nilesh Madhu, Michael Peter Hopwood, Ann Spriet, and Christophe De Vleeschouwer. Consistent iterative hard thresholding for signal declipping. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 5939–5943. IEEE, 2013. Cited on page 62.

[66] Andrew C Krochmal, David P Stewart, and John E Whitecar. Variable distortion limiter using clip detect predictor, April 10 2007. US Patent 7,202,731. Cited on page 62.

[67] Mike Lam, Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. Robust occupant detection through step-induced floor vibration by incorporating structural characteristics. In *IMAC XXXIV A Conference and Exposition on Structural Dynamics*, volume 54, pages 1–13. SEM, 2016. Cited on pages 14, 15, and 28.

[68] Edward A Lee. Cyber physical systems: Design challenges. In *Object oriented real-time distributed computing (isorc), 2008 11th ieee international symposium on*, pages 363–369. IEEE, 2008. Cited on page 9.

[69] Gwang-Gook Lee, Hyeong-ki Kim, Ja-Young Yoon, and Jae-Jun Kim. Pedes-

trian counting using an ir line laser. In *Convergence and Hybrid Information Technology, 2008. ICHIT'08. International Conference on*, pages 482–485. IEEE, 2008. Cited on page 113.

[70] Heyoung Lee, Jung Wook Park, and Abdelsalam Sumi Helal. Estimation of indoor physical activity level based on footstep vibration signal measured by mems accelerometer in smart home environments. In *Mobile Entity Localization and Tracking in GPS-less Environnments*, pages 148–162. Springer, 2009. Cited on pages 14, 15, and 76.

[71] RI Leine, DH Van Campen, A De Kraker, and L Van Den Steen. Stick-slip vibrations induced by alternate friction models. *Nonlinear dynamics*, 16(1):41–54, 1998. Cited on page 66.

[72] YF Li, SB Wang, and ZH Zhou. Graph quality judgement: a large margin expedition. *Proceedings of the 25th International Joint Confernece on Artificial Intelligence, New York, NY*, 2016. Cited on page 154.

[73] Yu-Feng Li, Ivor W Tsang, James T Kwok, and Zhi-Hua Zhou. Convex and scalable weakly labeled svms. *Journal of Machine Learning Research*, 14(1):2151–2188, 2013. Cited on page 154.

[74] Yu-Feng Li and Zhi-Hua Zhou. Improving semi-supervised support vector machines through unlabeled instances selection. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, pages 386–391. AAAI Press, 2011. Cited on page 154.

[75] Hongbo Liu, Yu Gan, Jie Yang, Simon Sidhom, Yan Wang, Yingying Chen, and Fan Ye. Push the limit of wifi based localization for smartphones. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 305–316. ACM, 2012. Cited on pages 12 and 13.

[76] W Loetwassana, R Punchalard, A Lorsawatsiri, J Koseeyaporn, and P Wardkein. Adaptive howling suppressor in an audio amplifier system. In *2007 Asia-Pacific Conference on Communications*, pages 445–448. IEEE, 2007. Cited on page 62.

[77] Rail-to-Rail Low-Voltage. *LMV321, LMV358, LMV324*, 2009. Cited on page 46.

[78] Carlos A Luna, Cristina Losada-Gutierrez, David Fuentes-Jimenez, Alvaro

Fernandez-Rincon, Manuel Mazo, and Javier Macias-Guarasa. Robust people detection using depth information from an overhead time-of-flight camera. *Expert Systems with Applications*, 71:240–256, 2017. Cited on pages 12 and 13.

[79] Alan Mainwaring, David Culler, Joseph Polastre, Robert Szewczyk, and John Anderson. Wireless sensor networks for habitat monitoring. In *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 88–97. Acm, 2002. Cited on page 11.

[80] Jani Mantyjarvi, Mikko Lindholm, Elena Vildjiounaite, SM Makela, and HA Ailisto. Identifying users of portable devices from gait pattern with accelerometers. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 2, pages ii–973. IEEE, 2005. Cited on pages 12, 13, and 153.

[81] Giulio Marin, Fabio Dominio, and Pietro Zanuttigh. Hand gesture recognition with leap motion and kinect devices. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 1565–1569. IEEE, 2014. Cited on page 12.

[82] How do microphones work? https://www.mediacollege.com/audio/microphones/how-microphones-work.html, 2015. Accessed: 2018-01-24. Cited on page 14.

[83] Lee Middleton, Alex Buss, Alex Bazin, Mark S Nixon, et al. A floor sensor system for gait recognition. In *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*, pages 171–176, Buffalo, NY, USA, USA, 2005. IEEE. Cited on page 152.

[84] Microphones: Dynamics, ribbons, and condensers. https://www.acousticfields.com/microphones-dynamics-ribbons-and-condensers/, 2012. Accessed: 2018-01-24. Cited on pages 14 and 15.

[85] Mostafa Mirshekari, Shijia Pan, Adeola Bannis, Yan Pui Mike Lam, Pei Zhang, and Hae Young Noh. Step-level person localization through sparse sensing of structural vibration. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, pages 376–377. ACM, 2015. Cited on pages 14, 15, and 57.

[86] Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. Characterizing wave propagation to improve indoor step-level person localization using floor

166

vibration. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*, pages 90611O–90611O. International Society for Optics and Photonics, 2016. Cited on pages 13, 57, 65, 74, 85, and 115.

[87] Mostafa Mirshekari, Shijia Pan, Pei Zhang, and Hae Young Noh. Characterizing wave propagation to improve indoor step-level person localization using floor vibration. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*, pages 980305–980305, Las Vegas, NV, USA, 2016. International Society for Optics and Photonics. Cited on pages 14, 67, 68, 72, and 73.

[88] Mostafa Mirshekari, Pei Zhang, and Hae Young Noh. Non-intrusive occupant localization using floor vibrations in dispersive structure: Poster abstract. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, pages 378–379. ACM, 2016. Cited on pages 15 and 113.

[89] Shin Miura, Hirofumi Nakajima, Shigeki Miyabe, S Makino, T Yamada, and K Nakadai. Restoration of clipped audio signal using recursive vector projection. In *TENCON 2011-2011 IEEE Region 10 Conference*, pages 394–397. IEEE, 2011. Cited on page 62.

[90] Le T Nguyen, Yu Seung Kim, Patrick Tague, and Joy Zhang. Identitylink: user-device linking through visual and rf-signal cues. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 529–539, Seattle, WA, USA, 2014. ACM. Cited on page 153.

[91] H Noh, R Rajagopal, and AS Kiremidjian. Sequential structural damage diagnosis algorithm using a change point detection method. *Journal of Sound and Vibration*, 332(24):6419–6433, 2013. Cited on page 14.

[92] Hae Young Noh, Dimitrios G Lignos, K Krishnan Nair, and Anne S Kiremidjian. Development of fragility functions as a damage classification/prediction method for steel moment-resisting frames using a wavelet-based damage sensitive feature. *Earthquake Engineering & Structural Dynamics*, 41(4):681–696, 2012. Cited on page 14.

[93] Hae Young Noh, K Krishnan Nair, Anne S Kiremidjian, and CH Loh. Application of time series based damage detection algorithms to the benchmark experiment at the national center for research on earthquake engineering (ncree)

in taipei, taiwan. *Smart Structures and Systems*, 5(1):95–117, 2009. Cited on page 14.

[94] Hae Young Noh, Ram Rajagopal, and Anne S Kiremidjian. Damage diagnosis algorithm using a sequential change point detection method with an unknown distribution for damage. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*, pages 834507–834507. International Society for Optics and Photonics, 2012. Cited on page 14.

[95] T Öberg, Alek Karsznia, and K Öberg. Basic gait parameters: reference data for normal subjects, 10-79 years of age. *Journal of rehabilitation research and development*, 30:210–210, 1993. Cited on page 135.

[96] Robert J Orr and Gregory D Abowd. The smart floor: A mechanism for natural user identification and tracking. In *CHI'00 extended abstracts on Human factors in computing systems*, pages 275–276. ACM, 2000. Cited on pages 13, 69, and 152.

[97] Shijia Pan, Amelie Bonde, Jie Jing, Lin Zhang, Pei Zhang, and Hae Young Noh. Boes: building occupancy estimation system using sparse ambient vibration monitoring. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*, pages 90611O–90611O. International Society for Optics and Photonics, 2014. Cited on pages xv, 12, 13, 14, 15, 19, 26, 27, 28, 31, 32, 35, 36, 37, 38, 41, 42, 76, 123, and 125.

[98] Shijia Pan, An Chen, and Pei Zhang. 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*, pages 99–104. ACM, 2013. Cited on page 153.

[99] Shijia Pan, Kent Lyons, Mostafa Mirshekari, Hae Young Noh, and Pei Zhang. Multiple pedestrian tracking through ambient structural vibration sensing: Poster abstract. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, pages 366–367, Stanford, California, 2016. ACM. Cited on pages xix, 20, 28, 65, 112, and 115.

[100] Shijia Pan, Mostafa Mirshekari, Jonathon Fagert, Ceferino Gabriel Ramirez, Albert Jin Chung, Chih Chi Hu, John Paul Shen, Pei Zhang, and Hae Young Noh. Characterizing human activity induced impulse and slip-pulse excitations

through structural vibration. *Journal of Sound and Vibration*, 414:61–80, 2018. Cited on pages xvi, xix, 20, 79, 81, 82, and 84.

[101] Shijia Pan, Mostafa Mirshekari, Hae Young Noh, and Pei Zhang. Structural sensing system with networked dynamic sensing configuration. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, pages 344–345. ACM, 2015. Cited on page 15.

[102] Shijia Pan, Mostafa Mirshekari, Pei Zhang, and Hae Young Noh. Occupant traffic estimation through structural vibration sensing. In *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*, pages 980306–980306, Las Vegas, NV, 2016. International Society for Optics and Photonics. Cited on page 15.

[103] Shijia Pan, Ceferino Ramirez, Mostafa Mirshekari, Jonathon Fagert, Albert Jin Chung, Chih Chi Hu, John Paul Shen, Hae Young Noh, and Pei Zhang. Surfacevibe: Vibration-based tap & swipe tracking on ubiquitous surfaces. In *Information Processing in Sensor Networks (IPSN), 2017 15th ACM/IEEE International Conference on*, pages 1–12. IEEE, 2017. Cited on pages 13, 15, 20, 28, 65, 73, 74, 79, and 113.

[104] Shijia Pan, Ningning Wang, Yuqiu Qian, Irem Velibeyoglu, Hae Young Noh, and Pei Zhang. Indoor person identification through footstep induced structural vibration. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, pages 81–86. ACM, 2015. Cited on pages 12, 13, 15, 62, 65, 76, 77, 115, 116, 126, 127, 131, 151, and 153.

[105] Shijia Pan, Susu Xu, Mostafa Mirshekari, Pei Zhang, and Hae Young Noh. Collaboratively adaptive vibration sensing system for high fidelity monitoring of structural responses induced by pedestrians. *Frontiers in Built Environment*, 3:28, 2017. Cited on pages xv, 19, 31, 32, and 37.

[106] Shijia Pan, Tong Yu, Mostafa Mirshekari, Jonathon Fagert, Amelie Bonde, Ole J Mengshoel, Hae Young Noh, and Pei Zhang. Footprintid: Indoor pedestrian identification through ambient structural vibration sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):89, 2017. Cited on pages xv, xviii, xix, 20, 23, 65, 115, 116, 118, 126, 143, 145, 147, 149, 150, and 152.

169

[107] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, 2011. Cited on page 121.

[108] Alexandros Pantelopoulos and Nikolaos G Bourbakis. A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(1):1–12, 2010. Cited on page 11.

[109] Joseph A Paradiso and Che King Leo. Tracking and characterizing knocks atop large interactive displays. *Sensor Review*, 25(2):134–143, 2005. Cited on pages 15, 67, and 74.

[110] Shwetak N Patel, Khai N Truong, and Gregory D Abowd. Powerline positioning: A practical sub-room-level indoor location system for domestic use. In *UbiComp*, volume 6, pages 441–458. Springer, 2006. Cited on page 12.

[111] Jacquelin Perry, Jon R Davids, et al. Gait analysis: normal and pathological function. *Journal of Pediatric Orthopaedics*, 12(6):815, 1992. Cited on page 10.

[112] Bo Persson. *Sliding friction: physical principles and applications*. Springer Science & Business Media, Berlin, 2013. Cited on pages 66 and 67.

[113] Duc Truong Pham, Ze Ji, Ming Yang, Zuobin Wang, and Mostafa Al-Kutubi. A novel human-computer interface based on passive acoustic localisation. In *International Conference on Human-Computer Interaction*, pages 901–909, Beijing, China, 2007. Springer. Cited on page 15.

[114] Ronald Poppe. A survey on vision-based human action recognition. *Image and vision computing*, 28(6):976–990, 2010. Cited on page 11.

[115] Jeffrey D Poston, R Michael Buehrer, and Pablo A Tarazaga. Indoor footstep localization from structural dynamics instrumentation. *Mechanical Systems and Signal Processing*, 88:224–239, 2017. Cited on page 57.

[116] Jeffrey D Poston, Javier Schloemann, R Michael Buehrer, VVN Malladi, Americo G Woolard, and Pablo A Tarazaga. Towards indoor localization of pedestrians via smart building vibration sensing. In *Localization and GNSS (ICL-GNSS), 2015 International Conference on*, pages 1–6. IEEE, 2015. Cited on pages 14, 15, and 62.

[117] Lion Precision. Capacitive sensor operation and optimization. *Tech Note. Jul,* 2006. Cited on page 13.

[118] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th annual international conference on Mobile computing & networking*, pages 27–38. ACM, 2013. Cited on page 12.

[119] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, and Kyle Jamieson. Widar: Decimeter-level passive tracking via velocity monitoring with commodity wi-fi. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, page 6. ACM, 2017. Cited on pages 12, 13, and 69.

[120] Weijun Qin, Jiadi Zhang, Bo Li, and Limin Sun. Discovering human presence activities with smartphones using nonintrusive wi-fi sniffer sensors: the big data prospective. *International Journal of Distributed Sensor Networks*, 9(12):927940, 2013. Cited on pages 12 and 13.

[121] Lawrence R Rabiner and Ronald W Schafer. *Digital processing of speech signals*. Prentice Hall, 1978. Cited on page 70.

[122] Ragunathan Raj Rajkumar, Insup Lee, Lui Sha, and John Stankovic. Cyber-physical systems: the next computing revolution. In *Proceedings of the 47th Design Automation Conference*, pages 731–736. ACM, 2010. Cited on page 9.

[123] Zhou Ren, Junsong Yuan, Jingjing Meng, and Zhengyou Zhang. Robust part-based hand gesture recognition using kinect sensor. *IEEE transactions on multimedia*, 15(5):1110–1120, 2013. Cited on page 12.

[124] Liu Rong, Duan Zhiguo, Zhou Jianzhong, and Liu Ming. Identification of individual walking patterns using gait acceleration. In *Bioinformatics and Biomedical Engineering, 2007. ICBBE 2007. The 1st International Conference on*, pages 543–546, Wuhan, China, 2007. IEEE. Cited on page 153.

[125] Osvaldo A Rosso, Susana Blanco, Juliana Yordanova, Vasil Kolev, Alejandra Figliola, Martin Schürmann, and Erol Başar. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of neuroscience methods*, 105(1):65–75, 2001. Cited on page 78.

[126] Hosam Rowaihy, Sharanya Eswaran, Matthew Johnson, Dinesh Verma, Amotz Bar-Noy, Theodore Brown, and Thomas La Porta. A survey of sensor selection

schemes in wireless sensor networks. In *Proc. SPIE*, volume 6562, pages A1–A13, 2007. Cited on page 63.

[127] James M Sabatier and Alexander E Ekimov. A review of human signatures in urban environments using seismic and acoustic methods. In *Technologies for Homeland Security, 2008 IEEE Conference on*, pages 215–220. IEEE, 2008. Cited on page 15.

[128] Raul Sanchez-Reillo. Hand geometry pattern recognition through gaussian mixture modelling. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 2, pages 937–940, Barcelona, Spain, Spain, 2000. IEEE. Cited on page 153.

[129] Cem Sertatıl, Mustafa A Altınkaya, and Kosai Raoof. A novel acoustic indoor localization system employing cdma. *Digital Signal Processing*, 22(3):506–517, 2012. Cited on page 69.

[130] Yuan Shi and Fei Sha. Information-theoretical learning of discriminative clusters for unsupervised domain adaptation. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 1079–1086, 2012. Cited on page 121.

[131] Ben Shneiderman. *Designing the user interface: strategies for effective human-computer interaction*. Pearson Education India, 2010. Cited on page 10.

[132] M Sifuzzaman, MR Islam, and MZ Ali. Application of wavelet transform and its advantages compared to fourier transform. *Journal of Physical Sciences*, 13, 2009. Cited on page 42.

[133] Martin L Smith and FA Dahlen. The azimuthal dependence of love and rayleigh wave propagation in a slightly anisotropic medium. *Journal of Geophysical Research*, 78(17):3321–3333, 1973. Cited on page 118.

[134] Jie Song, Yong-Feng Dong, Xin-Wei Yang, Jun-Hua Gu, and Pei-Pei Fan. Infrared passenger flow collection system based on rbf neural net. In *Machine Learning and Cybernetics, 2008 International Conference on*, volume 3, pages 1277–1281. IEEE, 2008. Cited on page 113.

[135] Sparkfun. *Sparkfun OpAmp Breakout LMV358*, 5 2015. Cited on pages xv and 45.

[136] Thad Starner, Bastian Leibe, David Minnen, Tracy Westyn, Amy Hurst, and Justin Weeks. The perceptive workbench: Computer-vision-based gesture tracking, object tracking, and 3d reconstruction for augmented desks. *Machine Vision and Applications*, 14(1):59–71, 2003. Cited on page 13.

[137] Allan Stisen, Henrik Blunck, Sourav Bhattacharya, Thor Siiger Prentow, Mikkel Baun Kjærgaard, Anind Dey, Tobias Sonne, and Mads Møller Jensen. Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 127–140. ACM, 2015. Cited on page 11.

[138] Arun Subramanian, Kishan G Mehrotra, Chilukuri K Mohan, Pramod K Varshney, and Thyagaraju Damarla. Feature selection and occupancy classification using seismic sensors. In *Trends in Applied Intelligent Systems*, pages 605–614. Springer, 2010. Cited on page 15.

[139] Budi Sugandi, Hyoungseop Kim, Joo Kooi Tan, and Seiji Ishikawa. Real time tracking and identification of moving persons by using a camera in outdoor environment. *Int. J. Innov. Comput. Inf. Control*, 5(5):1179–1188, 2009. Cited on page 153.

[140] Zheng Sun, Aveek Purohit, Kaifei Chen, Shijia Pan, Trevor Pering, and Pei Zhang. Pandaa: physical arrangement detection of networked devices through ambient-sound awareness. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 425–434. ACM, 2011. Cited on pages 13, 70, and 73.

[141] David H Sutherland. The evolution of clinical gait analysis part iii–kinetics and energy assessment. *Gait & Posture*, 21(4):447–461, 2005. Cited on page 152.

[142] Stephen P Tarzia, Robert P Dick, Peter A Dinda, and Gokhan Memik. Sonar-based measurement of user presence and attention. In *Proceedings of the 11th international conference on Ubiquitous computing*, pages 89–92. ACM, 2009. Cited on page 13.

[143] Thiago Teixeira, Gershon Dublon, and Andreas Savvides. A survey of human-sensing: Methods for detecting presence, count, location, track, and identity.

ACM Computing Surveys, 5(1):59–69, 2010. Cited on pages 12 and 69.

[144] Thiago Teixeira, Deokwoo Jung, and Andreas Savvides. Tasking networked cctv cameras and mobile phones to identify and localize multiple people. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 213–222, Copenhagen, Denmark, 2010. ACM. Cited on page 153.

[145] Acoustical Terminology. American national standard. *ANSI S1*, pages 1–1994, 2006. Cited on page 14.

[146] HP Verhas. Prediction of the propagation of train-induced ground vibration. *Journal of Sound and Vibration*, 66(3):371–376, 1979. Cited on page 67.

[147] Igor Aleksandrovich Viktorov. *Rayleigh and Lamb waves: physical theory and applications*. Plenum press, Berlin, Germany, 1970. Cited on pages 10, 35, 56, and 67.

[148] Hai Wang, Shao-Bo Wang, and Yu-Feng Li. Instance selection method for improving graph-based semi-supervised learning. In *Pacific Rim International Conference on Artificial Intelligence*, pages 565–573. Springer, 2016. Cited on page 154.

[149] Hanbiao Wang, Kung Yao, Greg Pottie, and Deborah Estrin. Entropy-based sensor selection heuristic for target localization. In *Proceedings of the 3rd international symposium on Information processing in sensor networks*, pages 36–45. ACM, 2004. Cited on page 62.

[150] Jue Wang, Deepak Vasisht, and Dina Katabi. Rf-idraw: virtual touch screen in the air using rf signals. In *ACM SIGCOMM Computer Communication Review*, volume 44, pages 235–246. ACM, 2014. Cited on pages 12 and 13.

[151] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. Silhouette analysis-based gait recognition for human identification. *IEEE transactions on pattern analysis and machine intelligence*, 25(12):1505–1518, 2003. Cited on page 153.

[152] Hiroki Watanabe, Tsutomu Terada, and Masahiko Tsukamoto. Ultrasound-based movement sensing, gesture-, and context-recognition. In *Proceedings of the 2013 International Symposium on Wearable Computers*, pages 57–64. ACM, 2013. Cited on page 13.

[153] Geoffrey Werner-Allen, Konrad Lorincz, Mario Ruiz, Omar Marcillo, Jeff

Johnson, Jonathan Lees, and Matt Welsh. Deploying a wireless sensor network on an active volcano. *IEEE internet computing*, 10(2):18–25, 2006. Cited on page 11.

[154] Robert Xiao, Greg Lew, James Marsanico, Divya Hariharan, Scott Hudson, and Chris Harrison. Toffee: enabling ad hoc, around-device interaction with acoustic time-of-arrival correlation. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pages 67–76. ACM, 2014. Cited on page 15.

[155] Chenren Xu, Bernhard Firner, Robert S Moore, Yanyong Zhang, Wade Trappe, Richard Howard, Feixiong Zhang, and Ning An. Scpl: indoor device-free multi-subject counting and localization using radio signal strength. In *Information Processing in Sensor Networks (IPSN), 2013 ACM/IEEE International Conference on*, pages 79–90. IEEE, 2013. Cited on pages 12, 113, and 153.

[156] Chenren Xu, Bernhard Firner, Yanyong Zhang, Richard Howard, Jun Li, and Xiaodong Lin. Improving rf-based device-free passive localization in cluttered indoor environments through probabilistic classification methods. In *Proceedings of the 11th international conference on Information Processing in Sensor Networks*, pages 209–220. ACM, 2012. Cited on pages 12 and 113.

[157] Huazhong Xu, Pei Lv, and Lei Meng. A people counting system based on head-shoulder detection and tracking in surveillance video. In *Computer Design and Applications (ICCDA), 2010 International Conference on*, volume 1, pages V1–394. IEEE, 2010. Cited on page 113.

[158] Susu Xu, Lin Zhang, Pei Zhang, and Hae Young Noh. An indirect traffic monitoring approach using building vibration sensing system: Poster abstract. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, pages 374–375, Stanford, CA, 2016. ACM. Cited on page 113.

[159] Vladislav A Yastrebov. Sliding without slipping under coulomb friction: opening waves and inversion of frictional force. *Tribology Letters*, 62(1):1–8, 2016. Cited on page 67.

[160] Hae Young Noh, K Krishnan Nair, Dimitrios G Lignos, and Anne S Kiremidjian. Use of wavelet-based damage-sensitive features for structural damage diagnosis

using strong motion data. *Journal of Structural Engineering*, 137(10):1215–1228, 2011. Cited on page 14.

[161] Moustafa Youssef, Matthew Mah, and Ashok Agrawala. Challenges: device-free passive localization for wireless environments. In *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pages 222–229. ACM, 2007. Cited on pages 12 and 13.

[162] Tong Yu, Shijia Pan, Susu Xu, Xinlei Chen, Mostafa Mirshekari, Jonathon Fagert, Pei Zhang, Hae Young Noh, and Ole J. Mengshoel. Ilpc: Iterative learning using physical constraints in real-world sensing data. *SmartIoT 2018 Workshop: AI enhanced IoT data processing for Intelligent Applications*, 2018. Cited on pages 120 and 122.

[163] Chengbin Zeng and Huadong Ma. Robust head-shoulder detection by pca-based multilevel hog-lbp detector for people counting. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 2069–2072. IEEE, 2010. Cited on page 113.

[164] Yunze Zeng, Parth H Pathak, and Prasant Mohapatra. Wiwho: wifi-based person identification in smart spaces. In *Proceedings of the 15th International Conference on Information Processing in Sensor Networks*, page 4. IEEE Press, 2016. Cited on pages 12, 13, and 153.

[165] Jingjing Zhang, Rui Wang, Shilong Lu, Jibing Gong, Ze Zhao, Haiming Chen, Li Cui, Nanyue Wang, and Youhua Yu. Easicprs: design and implementation of a portable chinese pulse-wave retrieval system. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, pages 149–161. ACM, 2011. Cited on page 62.

[166] Pei Zhang, Christopher M Sadler, Stephen A Lyon, and Margaret Martonosi. Hardware design experiences in zebranet. In *Proceedings of the 2nd international conference on Embedded networked sensor systems*, pages 227–238. ACM, 2004. Cited on page 11.

[167] Ying Zhang, Chuanjiang Luo, and Juan Liu. Walk&sketch: create floor plans with an rgb-d camera. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 461–470. ACM, 2012. Cited on page 69.

[168] Zhaonian Zhang and Andreas G Andreou. Human identification experiments

using acoustic micro-doppler signatures. In *Micro-Nanoelectronics, Technology and Applications, 2008. EAMTA 2008. Argentine School of*, pages 81–86. IEEE, 2008. Cited on page 13.

[169] Yuanqing Zheng, Guobin Shen, Liqun Li, Chunshui Zhao, Mo Li, and Feng Zhao. Travi-navi: Self-deployable indoor navigation system. *IEEE/ACM Transactions on Networking*, 2017. Cited on page 12.

[170] Datong Zhou, Qie Hu, and Claire J Tomlin. Model comparison of a data-driven and a physical model for simulating hvac systems. *arXiv preprint arXiv:1603.05951*, 2016. Cited on page 16.

[171] Xiaojin Zhu, Zoubin Ghahramani, John Lafferty, et al. Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, volume 3, pages 912–919, 2003. Cited on page 154.