



# *AutoQual*: task-oriented structural vibration sensing quality assessment leveraging co-located mobile sensing context

Yue Zhang<sup>1</sup> · Zhizhang Hu<sup>1</sup> · Susu Xu<sup>2</sup> · Shijia Pan<sup>1</sup>

Received: 24 March 2021 / Accepted: 9 June 2021 / Published online: 6 July 2021  
© The Author(s) 2021

## Abstract

In this paper, we introduce *AutoQual*, a mobile-based assessment scheme for infrastructure sensing task performance prediction under new deployment environments. With the growth of the Internet-of-Things (IoT), many non-intrusive sensing systems have been explored for various indoor applications, such as structural vibration sensing. This indirect sensing approach's learning performance is prone to deployment variance when signals propagate through the environment. As a result, current systems heavily rely on expert knowledge and manual assessment to achieve effective deployments and high sensing task performance. In order to mitigate this expert effort, we propose to systematically study factors that reflect deployment environment characteristics and methods to measure them autonomously. We present *AutoQual* that measures a series of assessment factors (AFs) reflecting how the deployment environment impacts the system performance. *AutoQual* outputs a task-oriented sensing quality (TSQ) score by integrating measured AFs trained from known deployments as a prediction of untested system's performance. In addition, *AutoQual* achieves this assessment without manual effort by leveraging co-located mobile sensing context to extract structural vibration signal for processing automatically. We evaluate *AutoQual* by using it to predict untested systems' performance over multiple sensing tasks. We conduct real-world experiments and investigate 48 deployments in 11 environments. *AutoQual* achieves less than 0.10 average absolute error when auto-assessing multiple tasks at untested deployments, which shows a  $\leq 0.018$  absolute error difference compared to the manual assessment approach.

**Keywords** Sensing quality · Deployment characterization · Mobile sensing · Performance prediction · Structural vibration sensing · Task-oriented integrated score

## 1 Introduction

IoT systems are becoming more and more pervasive in people's daily life. Due to their increasing applications and advantages in deployment (e.g., sparse, privacy preserving),

many non-intrusive indirect sensing techniques are developed for indoor human information acquisition, including RF-, vibration-, light-based methods. However, the indirect sensing mechanisms of these systems also induce large variances of the acquired data quality over deployment environment conditions and configurations, which reduces the system performance. We focus on structural vibration-based indoor sensing due to its passiveness, non-intrusiveness, room-level sensing range enabling extraction of fine-grained information (Pan et al. 2014; Bales et al. 2016; Clemente et al. 2019). The system's information inference performance (e.g., detection rate, learning accuracy) is impacted by the deployment environment. To systematically understand these deployment environment impacts, we define **sensing quality** as a series of measurable factors/models reflecting how they impact a given information inference task. Quantifying sensing quality allows further

---

✉ Yue Zhang  
yzhang58@ucmerced.edu

Zhizhang Hu  
zhu42@ucmerced.edu

Susu Xu  
susu.xu@stonybrook.edu

Shijia Pan  
span24@ucmerced.edu

<sup>1</sup> University of California Merced, 5200 N Lake Rd., Merced, CA 95340, USA

<sup>2</sup> Stony Brook University, 100 Nicolls Rd, Stony Brook, NY 11794, USA

enhancement of deployment efficiency to improve IoT sensing systems' performance.

Compared to prior work on signal quality assessment, which mainly used in the domains of communication (Srinivasan and Levis 2006; Islam et al. 2008; Baccelli and Błaszczyszyn 2010; Boano et al. 2009) and computer vision (Van den Branden Lambrecht 1998; Li and Bovik 2009; Wang and Bovik 2002), our proposed *AutoQual* reflects effects of deployment environment on the sensing task performance. The data quality assessments (Pipino et al. 2002; Cai and Zhu 2015) target the evaluation of the existing dataset and provide comparisons between multiple acquired datasets. However, they do not quantify the environmental impacts to acquired data characteristics. Our prior work on application-oriented sensing signal quality (SSQ) proposes a system-level signal quality assessment scheme with a set of metrics (Zhang et al. 2019) and demonstrate the possibility of using measurements of these metrics for optimal sensor placement selection (Yu et al. 2021). However, this sensing signal quality assessment requires manual calibration with known excitation of dense coverage. As a result, the approach is labor-intensive and impractical for large-scale sparse deployment assessment. In addition, the SSQ model combines different factors' measurements heuristically, which makes the approach difficult to generalize.

In this paper, we present *AutoQual*, an autonomous sensing quality assessment framework to quantify impacts of deployment environment on IoT sensing system performance. We take the structural vibration-based indoor human sensing system as an example and apply *AutoQual* for an autonomous sensing quality assessment on multiple sensing tasks. The main **challenges** include (1) how to identify and quantify environmental characteristics that impact the system performance, (2) how to integrate these AFs to assess the system for a given sensing task—different sensing tasks may be sensitive to different AFs, and (3) how to achieve (1) and (2) autonomously without manual efforts. We tackle these challenges by (1) utilizing domain knowledge on wave propagation and structural properties to identify a set of AFs and design the measuring method accordingly (Sect. 3.1), (2) adopting a data-driven approach to estimate the relationship between measured AFs and sensing tasks' performance (Sect. 3.3), (3) automating AF measurements using human-induced vibration signals extracted by co-located mobile devices (Sect. 3.2). The contributions of this work are as follows:

- We present *AutoQual*, a framework of autonomous task-oriented sensing quality assessment that predicts the IoT system performance utilizing the mobility of ambient occupants.
- We identify a set of measurable environmental factors that determine the sensing quality.
- We propose an auto-assessment scheme via human-induced signals enabled by co-located mobile sensing context.
- We evaluate *AutoQual* through real-world experiments at 48 deployments in 11 environments on multiple sensing tasks.

The rest of the paper is organized as follows. First, Sect. 2 introduces related works on IoT system quality assessment. Next, Sect. 3 presents the auto-assessing system design leveraging the co-located mobile devices. Then, Sects. 4 and 5 explain the evaluation experiments and result analysis. Furthermore, we discuss the limitation of this work and future directions in Sect. 6. Finally, we conclude the paper in Sect. 7.

## 2 Related work

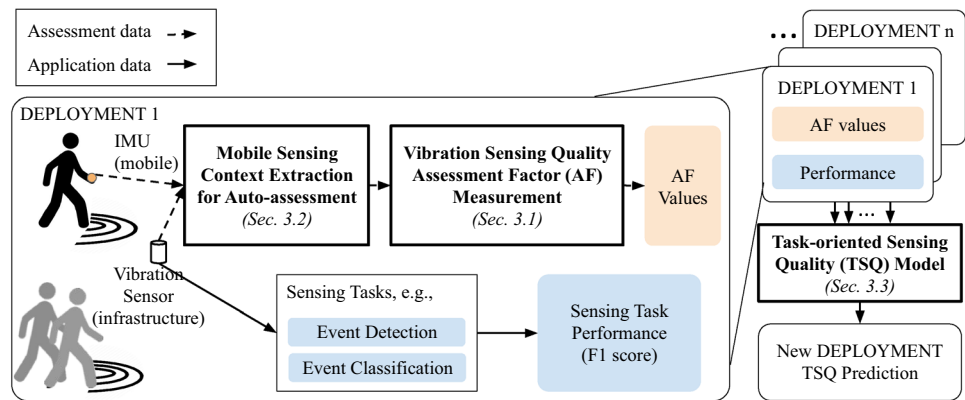
We consider the following aspects of related work including cross-modal system with infrastructural and mobile sensing (Sect. 2.1), sensing data quality measuring metrics (Sect. 2.2), and CPS-IoT system performance quantification methods (Sect. 2.3).

### 2.1 Infrastructural and mobile cross-modality sensing

Prior work has been done on combining infrastructural and mobile sensing to acquire target information, such as human activity recognition (Hu et al. 2020; Fortin-Simard et al. 2014), and air quality monitoring (Devarakonda et al. 2013; Gao et al. 2016; Jiang et al. 2011; Chen et al. 2020; Xu et al. 2019a, b). The infrastructure- and mobile-based subsystems often provide complementary data for each other to achieve a higher accuracy or a finer granularity of information (Pan and Nguyen 2020).

Although they combine the infrastructural and mobile sensing to improve system performance, they are usually focus on a specific sensing task or application instead of assess sensing quality or cross-modal system characterization. To the best of our knowledge, our work is the first sensing quality assessment framework that fusing the infrastructure and mobile sensing to assess sensing quality. Our work utilized the event label information from the mobile data to achieve autonomous sensing quality assessment for structural sensing.

**Fig. 1** *AutoQual* overview. The dash line arrows represent assessment data (mobile + infrastructural) and the solid line arrows represent sensing task data (infrastructural only)



## 2.2 Signal quality metrics

Prior work on signal quality measurements mainly focuses on measuring general signal properties, such as signal-to-noise ratio (SNR) (Oppenheim and Schaffer 1975), signal structural similarity (SSIM) (Li and Bovik 2009). When it comes to different types of signals for particular tasks, specific metrics have been developed. For example, wireless communication quality measurements include and not limited to Received Signal Strength Indicator (RSSI) (Srinivasan and Levis 2006), Carrier-to-Noise Ratio (CNR) (Islam et al. 2008), Signal-to-Interference-plus-Noise Ratio (SINR) (Baccelli and Błaszczyszyn 2010), and Link Quality Indicator (LQI) (Boano et al. 2009). Similarly, computer vision quality measurements include and not limited to image quality index (Van den Branden Lambrecht 1998), Structure SIMilarity (SSIM) (Li and Bovik 2009), and universal image quality index (Wang and Bovik 2002). These metrics, when applied to structural vibration-based approaches inferring information indirectly, do not model the deployment characteristics that impacts the performance of tasks with different requirements. In this paper, we focus on modeling (1) the deployment characteristics via ambient sensing signal and (2) the relationship between these deployment characteristics and different sensing tasks (which we refer to as the sensing quality metrics).

## 2.3 CPS/IoT performance quantification

Various frameworks or metrics have been proposed to quantify CPS/IoT systems' performance by measuring their data quality. For example, Karkouch et al. define the IoT data quality with a multi-dimensional definition including accuracy, confidence, completeness, data volume, and timeliness (Karkouch et al. 2016). While Banerjee et al. define the application-driven IoT quality from the computer system perspective (Banerjee and Sheth 2017). Other CPS/IoT performance assessments include communication performance and RFID-based health care application (Sellitto et al. 2007;

van der Togt et al. 2011). However, none of these prior works systematically evaluate the sensing (or data acquisition) process and identify the environment characteristics that impact this process, and eventually determine the CPS/IoT application performance.

## 3 AutoQual system design

We present *AutoQual*, a cyber-physical system sensing quality assessment framework that leverages indoor occupant mobility for automating the assessment of structural vibration-based indoor human sensing systems. Figure 1 shows an overview of *AutoQual*, consisting of three major components (1) vibration sensing assessment factor measurements, (2) mobile sensing context extraction for auto-assessment, and (3) Task-oriented Sensing Quality (TSQ) scoring model.

When occupants walk in a target sensing area, they are sensed by both vibration- and mobile-based systems, assuming a set of sensing systems are deployed on various indoor environments for data acquisition. This ambient sensing data of occupants is then used to train a model that describes the deployment environment characteristics and predicts sensing tasks' performance under a new deployment (i.e., assessing the sensing quality of the new deployment).

During the assessment data collection, footstep-induced vibrations are detected and grouped by traces based on the shared-context from mobile sensing. These selected assessment data are then sent to the AF measurement module and used to calculate the system's key impact factors—AF values. For the training deployments, the AF values and the system performance are sent to the TSQ scoring model module to establish a data-driven model that predicts the sensing tasks' performance of new deployments.

### 3.1 Vibration sensing quality assessment factor measurement

In order to quantify environmental impacts to sensing task performance, we use three physics models to describe the deployment environment characteristics. These models are (1) the attenuation model ( $AF_1, AF_2$ ), (2) the structural homogeneity model ( $AF_3$ ), and (3) the structure-sensor coupling model ( $AF_4, AF_5$ ). *AutoQual* autonomously measures parameters of these models based on the temporal association between mobile- and vibration-based sensing data.

#### 3.1.1 Attenuation model

When a vibration wave propagates through a solid, its attenuation is determined by the combined effects of geometric wave spreading, intrinsic attenuation, and scattering attenuation (Shapiro and Kneib 1993; Stein and Wyssession 2009; Fagert et al. 2021). These effects are modeled as a function of propagation distance  $d$

$$Amp(d) = \frac{Amp_0}{\sqrt{d}} e^{-2\alpha d} \quad (1)$$

where the decay rate  $\alpha = \alpha_s + \alpha_a$ ,  $\alpha_a$  is the coefficient of absorption and  $\alpha_s$  is the coefficient describing mean-field attenuation due to scattering (Shapiro and Kneib 1993).  $Amp_0$  is the initial amplitude of the vibration wave created by an excitation.

*AutoQual* measures the decay rate  $\alpha$  and the initial amplitude  $Amp_0$  of a deployment. However, Eq. 1 only describes the ideal attenuation without taking into account ambient noise, which cannot be directly measured in practice. We subtract the logarithmic amplitude of the background noise  $\lg Amp_N$  on both side of Eq. 1 and get

$$\lg(Amp(d)) = \lg Amp_0 - 2\alpha d - \frac{1}{2} \lg d \quad (2)$$

$$20 \lg \left( \frac{Amp(d)}{Amp_N} \right) = -40\alpha d + 20 \lg \left( \frac{Amp_0}{Amp_N} \right) - 10 \lg d \quad (3)$$

The term  $20 \lg \left( \frac{Amp(d)}{Amp_N} \right)$  is the Signal-to-Noise Ratio (SNR) (Oppenheim and Schaffer 1975) of the signal generated by excitation at distance  $d$  to the sensor, which can be directly measured. We estimate  $AF_1 = -40\alpha$ , and  $AF_2 = \lg \left( \frac{Amp_0}{Amp_N} \right)$  by conducting a linear fitting with SNR values measured at locations with multiple excitation-sensor distance  $d$ .

#### 3.1.2 Structural homogeneity model

The homogeneity of the structure directly impacts the data distribution of sensing signals. When a vibration wave propagates through a solid, the waveform distortion occurs and can be represented as

$$\mathbf{Y} = \mathbf{H}(d, \mathbf{dir})\mathbf{X} \quad (4)$$

where  $\mathbf{X}$  is the input force spectrum and  $\mathbf{Y}$  is the vibration frequency representation (i.e., the spectrum of the acquired signal). The function  $\mathbf{H}$  is a frequency response function of the structure (Mirshekari et al. 2020). The function is impacted by (1) distance  $d$  due to the dispersion effects of the Rayleigh–Lamb waves (Viktorov 1970), and the signal propagation direction  $\mathbf{dir}$  due to the structural homogeneity difference. In a homogeneous structure, the structural distortion effects at different directions are the same, i.e.,

$$\mathbf{H}(d) = \mathbf{H}(d, \mathbf{dir}).$$

A sensing system deployed in a homogeneous structure has a higher data efficiency because we can use signals propagated from any direction to establish the model  $\mathbf{H}$ .

*AutoQual* measures the structural homogeneity of a deployment as the similarity of signals frequency response  $\mathbf{Y}(d, \mathbf{dir})$  from different propagation directions  $\mathbf{dir}$  with controlled (same) input force spectrum  $\mathbf{X}$  and sensor-excitation distance  $d$ . *AutoQual* calculates this signal similarity as,

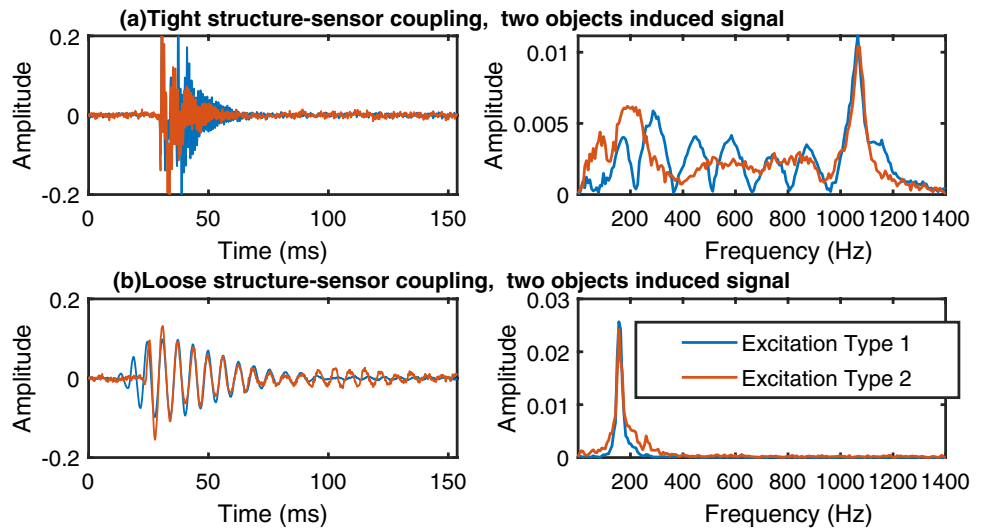
$$AF_3 = \max \text{xcorr}(\mathbf{Y}(d, \mathbf{dir1}), \mathbf{Y}(d, \mathbf{dir2})),$$

where signals  $\mathbf{Y}(d, \mathbf{dir1})$  and  $\mathbf{Y}(d, \mathbf{dir2})$  are normalized by their signal energy.  $AF_3$  reflects the directional distortion of signal propagation media.

#### 3.1.3 Structure-sensor coupling model

The structure-sensor coupling condition varies over different deployments, especially for different surface materials. When there is a tight coupling between the structure and the sensor, the sensor captures distinct structural response frequencies caused by different excitation. However, for surfaces with a loose structure-sensor coupling, the excitation's frequency response is dominated by the interaction between the structure and the sensor showing a less distinguishable waveform for different excitation. Figure 2 demonstrates an example of structural response frequency over different structure-sensor coupling conditions. The signals shown in Fig. 2a are acquired with a tight structure-sensor coupling condition, and the signals in (b) are acquired with a loose structure-sensor coupling. A tight coupling condition ensures that the acquired signal is dominated by the effects of structure-excitation interaction. These types of signals are

**Fig. 2** Frequency response over different structure-sensor coupling conditions. **a, b** show the vibration signals (normalized by signal energy) induced by different types excitation with different structural-sensor coupling condition. The excitation frequency responses in tight structural-sensor coupling environment are more distinguishable



less impacted by effects induced by structure-sensor interaction. Therefore, a sensing system with a tight structure-sensor coupling is more informative of different excitation sources, giving better sensing quality in event classification task.

*AutoQual* measures the excitation vibration signal's frequency component distribution as  $x\%$  energy concentration bandwidth (ECB).

$$ECB(x) = \arg \min_b \left( \sum_{f=b_0}^{b_0+b} PSD_{norm}(f) \geq x\% \right) \quad (5)$$

where the  $PSD_{norm}$  is the power spectral density (PSD) normalized by signal energy.  $b_0$  is traversed from 0 to  $F_s/2 - b$ .  $F_s$  is the sampling rate. Specifically, *AutoQual* measures  $AF_4 = ECB(75)$ ,  $AF_5 = ECB(50)$  to reflect structural modes that can be excited by the excitation.

### 3.2 Mobile sensing context extraction for vibration sensing assessment

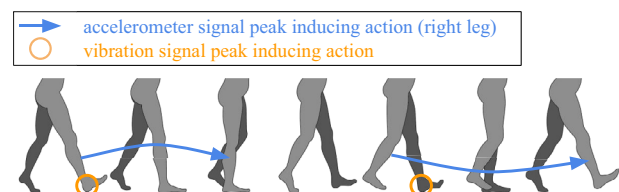
To reduce human efforts for collecting data for assessing environmental impacts (Zhang et al. 2019), *AutoQual* utilizes occupants' mobility and shared-context between mobile- and vibration-sensing to achieve autonomous assessment.

The properties that allow pedestrian's footstep-induced signals to substitute manually generated standard excitation are twofold: (1) the same person's footfall generates consistent excitation, which is equivalent to standard excitation for assessment in prior work, (2) when a pedestrian passing by, their footfall locations change, generating signals with different sensor-excitation distances needed for AF measurements. The challenges of using ambient human footstep-induced vibration signals to measure AFs are twofold: (1) there are

ambient non-footstep events that can be detected and should not be used for AF measurement, (2) the human footfall, compared to the standard excitation, are less consistent (e.g., left/right foot difference) and more complicated (e.g., toe push-off induces damped free vibrations).

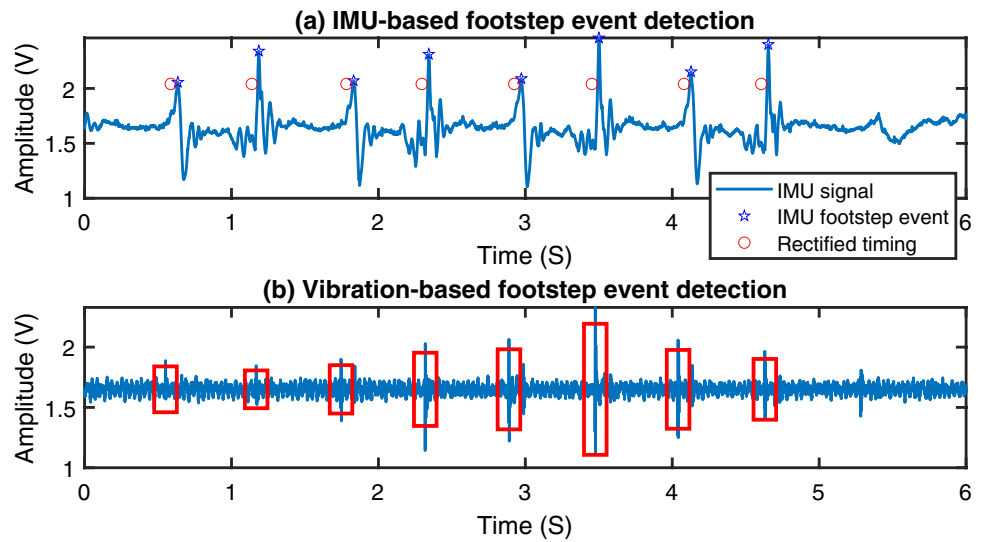
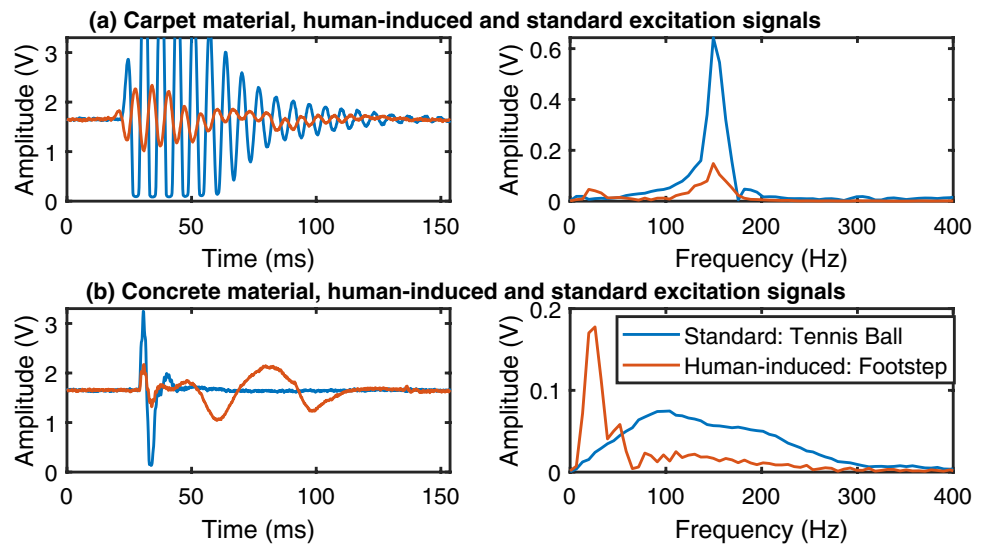
#### 3.2.1 Gait-based mobile-infrastructure sensing signal temporal association

When people walk in their natural form, their gait/footstep can be detected by mobile devices carried on them (Hanon and Anderson 2009) through the Inertial Measurement Unit (IMU). *AutoQual* leverages co-located mobile device to detect the timing of the footfall, which has been explored for position and gait cycle detection (Mokaya et al. 2013; Grimmer et al. 2019; Moon et al. 2019). In this work, we use a three-axis accelerometer sensor on the calf to measure the pedestrian's footstep timing. *AutoQual* conducts peak detection on the accelerometer signal by finding local maxima as footfall timestamp. This is done on the axis with the highest signal amplitude. For vibration signals, *AutoQual* conducts the anomaly detection to extract the excitation event signal segments (Pan et al. 2014).



**Fig. 3** Cross-modal gait-cycle-based timing association



**Fig. 4** Multi-modal gait-cycle-based timing rectification**Fig. 5** An example of the difference between human-induced signal and standard excitation signal. **a, b** show the time domain and frequency domain (normalized) of tow signals that induced by human and standard excitation in two environments. The surface of environment **(a)** is a soft carpet and the surface of environment **(b)** is concrete which has a higher stiffness. We can observe a clear damped free vibration with the footstep-induced signal in **(b)**

The challenge to associate the footstep timing between the structural vibration and accelerometer data is that the detectable gait patterns have a *cycle offset* between these two sensing modalities, illustrated in Fig. 3. The *initial strike* of the investigated leg (marked in the orange circle) would induce a footstep-induced vibration signal event at time  $T_{vib}$ . However, the calf motion that induces the cycle phase shift from the *initial strike* to the *mid stance* would induce a detectable peak in the acceleration measurement at time  $T_{acc}$ . As a result, the gait phase detected by the IMU is slightly ( $\sim 150$ ms) lagged compared to that of the vibration signal.

To rectify this gait cycled offset for robust event signal association (e.g., when there are detectable ambient events

occur at a similar time of the footstep), *AutoQual* utilizes the average of the gait cycle offset between event timing of two modalities within a trace.<sup>1</sup> *AutoQual* first associates the IMU footstep timing to the vibration footstep event with the closest timing. Given  $q^{th}$  detected vibration event, we find the  $p^{th}$  IMU event  $\arg \min_p |T_{acc}(p) - T_{vib}(q)|$  as a pair. For the  $Q$  pairs of associated IMU and vibration event in a trace, *AutoQual* estimates the gait cycle offset as

$$T_{offset} = \frac{1}{Q} \sum_{i=1}^Q T_{acc}(q) - T_{vib}(p).$$

Then, *AutoQual* rectifies the  $T_{acc}$  as  $T_{acc} - T_{offset}$ . Figure 4 shows an example of the two modalities event timing respectively as well as the rectified timing.

<sup>1</sup> We define a trace as a series of footstep events when a person walks by.

### 3.2.2 Signal processing for AF measurements

Unlike the standard excitation, using footstep-induced vibration signals is challenging because of the following reasons (1) the randomness of human behavior makes the excitation less consistent compared to that of the standard excitation. (2) the footstep-sensor distance  $d$  used in the attenuation model is unknown. (3) the footstep-induced vibration signal has a heavy damped free vibration due to the toe push-off motion, which may directly impact the signal's frequency characteristics for the structure-sensor coupling model calculation. As shown in Fig. 5b, the footstep-induced vibration signal contains strong damped free vibration components, i.e., the tail of the signal (Shi et al. 2019), which makes the ECB measurements ambiguous for different structure-sensor coupling conditions.

To address the first challenge, *AutoQual* leverages the associated information from IMU data to identify corresponding footstep-induced vibration signals to calculate the AFs. To address the second challenge, we consider an average foot stride length *strLen* of people approximately 2 ft at a normal walking speed for men and women between 20 and 39 years old (Öberg et al. 1993). When a pedestrian passes by, *AutoQual* first conducts the event detection with an anomaly detection algorithm (Pan et al. 2014). Then the system selects the detected footstep signal with the highest signal energy as the reference footstep and assign a fixed reference distance  $d_{ref}$ . For a footstep that is  $k$  step away from the reference footstep, we consider its distance to the target sensor  $d_k$  can be calculate as

$$d_k = \sqrt{d_{ref}^2 + (k * strLen)^2} \quad (6)$$

We use the estimated distances and the footstep induced vibration signals to further calculate the attenuation model factors as discussed in Sect. 3.1. For the associated events of a sequence of footstep-induced vibration signals, *AutoQual* calculates the AF values and estimates the TSQ score. For the deployment that no events are associated, *AutoQual* reports as not a valid assessment. To address the third challenge, we only extract the onset of the footstep signal to avoid the push-off induced damped free vibration when calculating the structure-sensor coupling model factors.

### 3.3 Task-oriented sensing quality (TSQ) scoring model

Different sensing tasks are sensitive to different environmental factors and have distinct requirements. To ensure the fairness of system sensing quality comparison, a Task-oriented sensing quality score is necessary. Compared to

conventional once-for-all models, the TSQ score provides a fine-grained representation of how AFs influence the system performance under different tasks. It further enables a comprehensive understanding of sensing system qualities under different task scenarios and thus achieves a more precise prediction of system performance on a set of tasks. To calculate the TSQ score, we first model the projection from the individual raw AF measurements to a saturation function ranged between 0 and 1 (Sect. 3.3.1). Then we integrate all AF's saturation functions as the TSQ score (Sect. 3.3.2).

#### 3.3.1 Saturation function for individual AF

The impacts of AFs on the sensing task performance are often constrained. For example,  $AF_1$  is calculated from the signal decay rate  $AF_1 = -40\alpha$ . For the same excitation, the higher the decay rate (the lower the  $AF_1$  value), the lower the event detection rate. When  $AF_1$  is lower than a threshold, there will be no footstep-induced vibration signal detected, and the sensing task performance's decreasing trend flattens. Similarly, when the value of  $AF_1$  is higher than a threshold, the detection rate increase in the target sensing range is no longer obvious. In both cases, we consider the impact of  $AF_1$  is saturated, which we model with a saturation function—sigmoid.

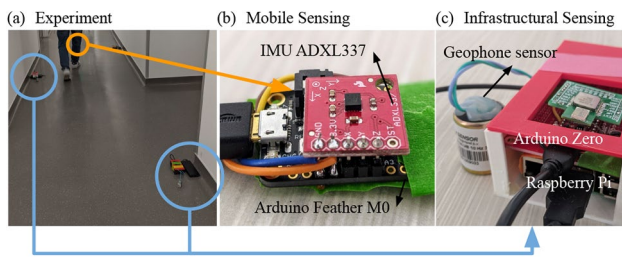
$$S(AF) = \frac{1}{1 + e^{(aAF+b)}} \quad (7)$$

As shown in E.q. 7, we quantify the impact of individual AF on system performance in the range of 0 to 1. In addition, we further constraint the saturation range by setting an upper and lower threshold ( $T_u$  and  $T_l$ ) on the AF measurement where  $S(T_u) = 0.9$  and  $S(T_l) = 0.1$ .

#### 3.3.2 Task-oriented integration model

A holistic model is built to integrate all AFs and provide an overall assessment (TSQ score) of the deployment. Different AFs may have different impacts on the sensing task performance. For example, the attenuation model AFs may play more important roles for the event detection task than the structure-sensor coupling model. **The integration model should configure a unique weight for each AF.** On the other hand, the same AF has different impacts for different sensing tasks. For example, the attenuation model AFs may have a more essential influence on event detection than event classification. **The weight of each AF should varies for different tasks.**

To enable a uniform and quantified sensing quality assessment criteria, we build a task-oriented integration model to calculate the TSQ score. In real-world scenarios, there are often a limited number of sensing system deployments as well as noisy AF measurements to train the integration model.



**Fig. 6** Hardware and an example deployment

Therefore, a model with high capacity and complexity is easy to get over-fitting with limited training sets and results in low performance of final TSQ score prediction. Meanwhile, the interpretability of the integration model is important for people to understand how different AFs impact the system performance and response accordingly. Therefore, we choose linear regression model (Wang et al. 2004) on top of the saturation function for individual AFs to build the integration model. Given a deployment environment  $j$ , the TSQ score of  $j$  is modeled as

$$TSQ^j = \sum_{i=1}^M w_i S(AF_i^j) + c, \quad (8)$$

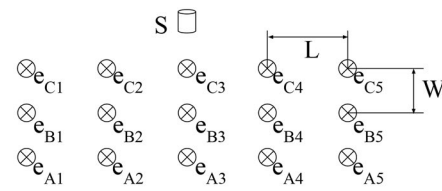
where  $AF_i^j$  is the  $i$ th AF of the deployment environment  $j$ ,  $w_i$  is the weight of  $AF_i^j$ , and the  $c$  is the estimated constant offset,  $M$  is the number of all AFs.

The goal is to predict the real-world sensing system performance. The least squares approach is employed here to estimate the parameters. The objective function for obtaining an optimal TSQ score model is as follows:

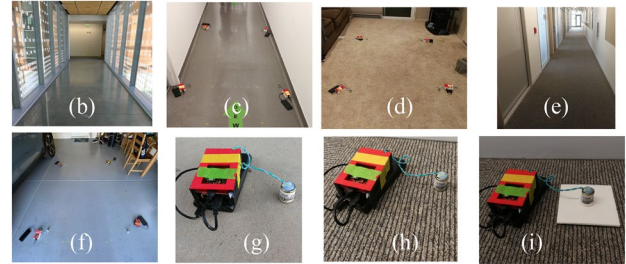
$$\min_{w, c, T_l, T_u} \sum_{j=1}^N \left[ \left( \sum_{i=1}^M w_i S(AF_i^j) + c \right) - p^j \right]^2 \quad (9)$$

where  $p^j$  is the sensing task performance (classification accuracy, detection rate, etc.) at deployment  $j$ .  $T_l$  and  $T_u$  is a vector consists of the lower and upper thresholds of the saturation function.  $N$  is the number of training deployments.

We use the gradient descent algorithm (Boyd et al. 2004) to find the optimal solution of the objective function 9. To avoid the local minimum, we randomly select the initial point multiple times and select the best loss solution as the final optimal solution. Two stopping conditions are set to exit the iteration, (1) when the difference between two consecutive objectives is less than a given threshold  $\delta_o$ ; (2) when the Euclidean distance between two consecutive solutions is less than a given threshold  $\delta_s$ . The selection of  $\delta_o$  and  $\delta_s$  are described in Sect. 4.2.



**(a) deployment configuration: sensor and excitation locations**



**Fig. 7** Experiment deployment environments and configurations. **a** Deployment configuration. The cylinders represent sensor locations, the circles with crosses indicate excitation locations. **b–i** Deployment environment examples

**Table 1** Deployment environment details

Environment ID	Floor material	Layout
1	Epoxy	Rooms
2	Carpet (hard)	Hallway
3	Epoxy	Bridge (connect two second floor)
4	Epoxy	Hallway
5	Epoxy	Hallway (with beam)
6	Carpet (soft)	Rooms
7	Carpet (soft)	Rooms (with mounting device)
8	Concrete	Rooms
9	Concrete	Open space
10	Epoxy	Rooms
11	Epoxy	Open space

## 4 Experiments

Experiments are conducted to evaluate the introduced assessment framework, including two sensing tasks over 48 deployments at 11 different environments. We obtain sensing task datasets (Sect. 4.1) and sensing quality assessment datasets (Sect. 4.2) separately.

*AutoQual* consists of two sensing modalities, a structural vibration sensing system and a wearable system. Figure 6 shows the two system hardware implementation—(b) demonstrates the mobile sensing system with an three-axis accelerometer and (c) illustrates the infrastructural sensing system



with structural vibration sensor geophone SM-24 (SM-24 Geophone Element 2006). The mobile sensing node consists of the accelerometer module ADXL 337 (Small, Low Power 2010) to capture the footfall motion, and an Arduino Feather M0 board (Adafruit feather m0 bluefruit le 2021) to digitize the signal and store the data. We sample the three axes of the accelerometer at 800 Hz to ensure the high temporal resolution of the signals. The infrastructural sensing node consists of an operational amplifier LVM385 (LMV3xx Low-Voltage 2020) and an Arduino board with a 32-bit ARM Cortex-M0+ processor and a 10–12 bit ADC module (Sparkfun samd21 mini breakout 2021). We sample the geophone at a sampling rate of 6500 Hz.

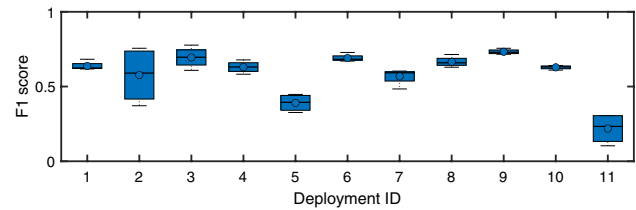
We conduct experiments at 11 different environments varying over floor materials (e.g., concrete, carpet, epoxy), and layout (e.g. hallways, rooms, and open space). Table 1 lists the characteristics of each deployment environment. We setup in total 48 deployments at 11 environments over different locations. Figure 7a shows an example deployment configuration, where the sensor and relative locations of excitation used for assessment are marked. Since the narrowest hallway we deploy has a width of 5 ft, we set the relative location of the excitation in the environment with  $W = 1$  ft. In addition, because people's strike length is approximately 2 ft, we set  $L = 2$  ft. We define a set of excitation across the sensing area as one **trace**, e.g., the set of five excitation at  $e_{A1}, e_{A2}, \dots, e_{A5}$  is considered as one trace referred to as trace  $e_A$ .

#### 4.1 Sensing task dataset collection and processing

We collected floor vibration signals when people walk through sensing areas to establish the sensing task dataset. For each participant, we collect her/his footstep-induced floor vibration with two types of shoes—sneakers (soft-soled) and hiking shoes (hard-soled) over every deployment. In each deployment, the task dataset contains four types of footstep-induced excitation from two participants with two shoe types each. For each scenario (distinguished by person, shoe type, environment, sensor), a 3-min vibration signal is collected and in total we collect more than 9 h (576 min) worth of floor vibration data. To demonstrate the task-oriented assessment, we evaluate *AutoQual* over two common sensing task representatives: event detection (Sect. 4.1.1) and event classification (Sect. 4.1.2).

##### 4.1.1 Event detection

The event detection task is to detect footstep-induced vibration signal events from ambient floor vibration. We apply



**Fig. 8** F1 scores of event detection in 11 different deployment environments. The x-axis is the deployment environment ID. The y-axis is the event detection F1 score. The event detection task shows varying performances in different environments

an anomaly detection based algorithm to achieve the event detection (Pan et al. 2014). The anomaly detection is done on the sliding window (with a window size of 0.415 s and an offset size of 0.015 s). For each noise window, we calculate the signal energy and build a Gaussian noise model  $\mathcal{N}(\mu, \sigma^2)$ . For each testing window, the algorithm compares the sliding windows' signal energy to the Gaussian noise model. If the signal energy is higher than a threshold (here set to  $8\sigma$  empirically), we consider the testing window contains a detected event.

We acquire the label for each footstep occurred during the experiment with the accelerometer on the mobile device (placed on the calf). A peak detection algorithm (Ying et al. 2007) is used to detect the timestamp of a footfall with the time series data of accelerometer. In total, more than 27,000 individual footstep events are collected in the task dataset. F1 score is used as the evaluation metric to measure the event detection performance. For each labeled timestamp  $T_{gt}$ , if an event is detected within the time window  $T_{gt} \pm Th_{gt}$ , we consider it as a true positive (TP), otherwise a false negative (FN). If an event is detected outside the time window of  $T_{gt} \pm Th_{gt}$ , we consider it a false positive (FP). The threshold  $Th_{gt}$  is empirically set as 0.0769 s (500 samples). The precision and recall are calculated as

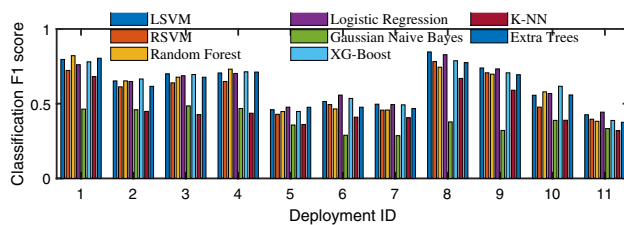
$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

The F1 score is a function of precision and recall

$$F1\ score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

Figure 8 shows the detection F1 score in different deployment environments. The F1 score of event detection over 48 deployments vary from 0.044 to 0.777, with an average of 0.600 and a standard deviation of 0.186.



**Fig. 9** F1 scores of event classification using eight learning algorithms in 11 different deployment environments. The x-axis is the deployment environment ID. The y-axis is the event classification F1 score. Different color bars represent results using different algorithms. The event classification task shows varying performance in different environments

#### 4.1.2 Event classification

The event classification task is a multi-class classification task to recognize four types of footsteps, i.e., two participants' footstep-induced vibration wearing two types of shoes. In order to mitigate the impact of the classification algorithm and enable a fair comparison between deployments, we apply eight commonly used classification algorithms, including Support Vector Machine (SVM) with Linear Kernel (LSVM) and RBF Kernel (RSVM), Gaussian Naive Bayes (GNB), Random Forest (RF), Extra Trees (ET), K-nearest Neighbors (k-NN), Logistic Regression (LR), and XG-Boost (Boser et al. 1992; Rish 2001; Kleinbaum et al. 2002; Liaw and Wiener 2002; Freund et al. 1999; Chen et al. 2016; Goldberger et al. 2005; Geurts et al. 2006; Chen et al. 2015). The classifier inputs are the frequency components ranging from 10 to 400 Hz of a 1000-sample window containing the investigated signal, which is a 391 dimension feature. The model is trained with bootstrap aggregating. The training and evaluation is conducted in a nested 5-fold cross-validation fashion. In each fold, we randomly sample 80% data from the training set each time, and aggregate the learned models to obtain the final model hyperparameters. The ground truth of event classification is the type of excitation distinguished by the pedestrian identities and their footwear types, which are labeled manually. The averaging F1-score is employed as the metric to evaluate the performance of multi-class classification tasks (Schütze et al. 2008). Figure 9 shows the classification F1 score of each algorithm in different environments. Different algorithms have different performances using the same dataset. To eliminate the performance variation induced by the classification algorithm, the **event classification** performance at each deployment is calculated as the mean value of the averaging F-1 scores of eight different classification algorithms. In this way, we can eliminate the assessment bias introduced by

algorithms. The event classification performance over 48 deployments varies from 0.302 to 0.780, with a mean value of 0.555 and a standard deviation of 0.125.

## 4.2 Sensing quality assessment

The goal of the sensing quality assessment is to provide a fair and comparable measurement of the deployment environments regardless of the applications/tasks. As a result, the data used for assessment should be independent of the sensing tasks, so we collected the assessment dataset separately from the sensing task dataset. In order to compare to prior work on manual sensing signal quality assessment (Zhang et al. 2019), we collect both the manual assessment dataset with standard excitation (tennis ball drops from a consistent height) and *AutoQual* assessment dataset with ambient footstep excitation.

The locations of excitation are marked as circles with crosses in Fig. 7a. For the manual assessment dataset, we collected 10 excitation samples at each marked excitation location. For human excitation, we collected 18 traces (people walking by) including six along  $e_A$ ,  $e_B$ , and  $e_C$  respectively.

### 4.2.1 *AutoQual* auto-assessment via ambient footstep excitation

In order to calculate a stable TSQ score of a deployment, we use 18 traces from the sensing quality assessment dataset at each deployment to calculate the AFs. The AFs are calculated as follows. (1) Attenuation Model ( $AF_1, AF_2$ ): *AutoQual* selects the footstep-induced vibration signal with the highest SNR in each trace as the reference footstep and set the parameter  $d_{ref} = 2ft$  in Eq. 6 to estimate each signal's sensor-excitation distance. Then we use Eq. 3 to conduct linear regression. We remove outliers beyond  $2\sigma$  of the model expectation, where the  $\sigma$  is the standard deviation of the model variance. (2) Structural Homogeneity Model  $AF_3$ : *AutoQual* selects pairs of footstep-induced signal that have the same distance to the reference footstep signal to calculate  $AF_3$ . We report the maximum of all the calculated values as the measurement of  $AF_3$  in this deployment. (3) Structure-sensor Coupling Model ( $AF_4, AF_5$ ): *AutoQual* uses the reference footstep signal in each trace to calculate  $ECB(x)$ , where  $x = 75\%, 50\%$  to calculate the 3rd quartile, 2nd quartile of the energy distribution. We use the average value over multiple traces as the final  $AF_4, AF_5$  of this deployment.

For each training deployment, we calculate its AFs and acquire the corresponding ground-truth system performance  $p$ . Given  $N$  training deployments, the data-driven model is calculated by solving the objective function in Eq. 9. We apply the gradient descent algorithm to obtain the optimal TSQ score model parameters. The stopping conditions are

setup as:  $\delta_o = 10^{-6}$  and  $\delta_s = 10^{-10}$ . To avoid overfitting, we set the maximum number of iterations to 200. The gradient descent algorithm exits the iteration if one of the stopping conditions is satisfied. We randomly select 50 initial points and report the solution with the lowest loss as the final solution.

#### 4.2.2 Baseline 1 manual assessment vs. auto assessment

To understand the efficiency of *AutoQual* with mobile-enabled assessment, we conduct a manual assessment with standard excitation as a baseline. In each deployed environment, we drop tennis ball at the marked 15 locations in Fig. 7a as the standard excitation to calculate the AFs. (1) Attenuation Model AFs: we use the mean SNR of all excitations at each locations as the SNR of each location, and then calculate the distance between each location and the vibration sensor. Finally, SNR and distance values to calculate  $AF_1$  and  $AF_2$ . (2) Structural Homogeneity Model AF: we select pairwise signals with the same distance to the sensor in each trace and calculate their waveform similarity. The highest value from all traces in a deployment is used as  $AF_3$ . (3) Structure-sensor Coupling Model AFs: we utilizes signals with the shortest excitation-sensor distance in a trace ( $e_{A3}$ ,  $e_{B3}$ , or  $e_{C3}$  in Fig. 7a) to calculate  $AF_4$  and  $AF_5$ .

#### 4.2.3 Baseline 2 AF saturation function: sigmoid v.s. piecewise

*AutoQual* utilizes data-driven approaches to determine the saturation bound of each AF, i.e., when an AF value is out of the saturation bound, its impact on the system performance is saturated. To demonstrate the efficiency of the data-driven approach adopted by *AutoQual*, we compare *AutoQual*'s saturation function to a piecewise function as the baseline

$$f(AF) = \begin{cases} 1 & T_{max} < AF \\ (AF - T_{min}) / (T_{max} - T_{min}) & T_{min} \leq AF \leq T_{max} \\ 0 & AF < T_{min} \end{cases} \quad (10)$$

where  $T_{min}$  and  $T_{max}$  are the minimum and maximum values of each AF from the training deployments.

#### 4.2.4 Baseline 3 assessment factor selection

To understand the effectiveness of the AFs for representing the sensing quality, we predict the system performance using the same framework with two traditional signal quality metrics: SNR (Yi et al. 2012) and SSIM (Clifton et al. 1938;

Chen and Bovik 2011). For the SNR, we calculate the mean SNR value of all footstep-induced signals in a deployment. For the SSIM, we calculate the mean value of all signal pairs with equal distance to the sensor in a deployment. The SSIM value of two signals  $x$  and  $y$  is calculated as

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)(\sigma_{xy} + C_3)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)(\sigma_x\sigma_y + C_3)},$$

where  $\mu_x$  is the mean value and  $\sigma_x$  is the standard deviation of the signal  $x$ , respectively.  $\sigma_{xy}$  is the covariance of  $x$  and  $y$ ;  $C_1 = (K_1L)^2$ ,  $C_2 = (K_2L)^2$ , and  $C_3 = C_2/2$  are constants that ensure stability when the denominator approaches zero;  $L$  is the dynamic range of the digitized signal values. In our implementation, We select SSIM parameters  $K_1 = 0.01$ ,  $K_2 = 0.02$  based on recommended parameter ranges (Chen and Bovik 2011). The dynamic range of the signal values is  $L = 2^{10}$  for a 10-bits ADC module. We project the SNR and SSIM values in the training set to the range of 0 and 1 linearly, and predict the TSQ score as  $TSQ_{SNR} = f(SNR)$  and  $TSQ_{SSIM} = f(SSIM)$ , where  $f(\cdot)$  is defined in Eq. 10,  $T_{min}$  and  $T_{max}$  are the minimum and maximum of SNR or SSIM values from the training set.

#### 4.2.5 Baseline 4 integration methods: task oriented vs. equal weight

*AutoQual* provides a task-oriented integration model by selecting different weights for different AFs. We consider a non-task-oriented and non-data-driven integration model (Equal Weight) as the baseline to evaluate the efficiency of our integration model. For each deployment, we use the mean value of all AFs to calculate the TSQ score, which is calculated as

$$AF_{avg} = \frac{1}{N} \sum_{i=1}^N AF_i.$$

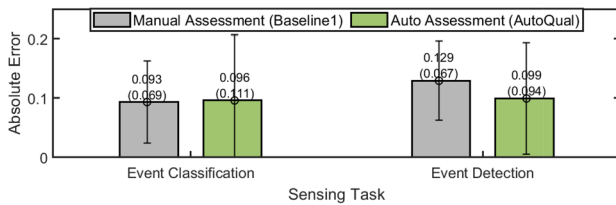
The TSQ score is calculated as

$$TSQ = f(AF_{avg}),$$

where  $f(\cdot)$  is defined in Eq. 10,  $T_{min}$  and  $T_{max}$  are the minimum and maximum of  $AF_{avg}$  from the training set.

## 5 *AutoQual* evaluation

We conduct a detailed performance analysis on *AutoQual* over baselines from multiple perspectives in this section. Implementation details and multiple types of baseline methods are first presented. We further compare our framework



**Fig. 10** *AutoQual* evaluation with manual assessment and mobile-enabled auto assessment. The mobile-enabled auto assessment method has a similar performance comparing with the manual assessment. The x-axis is the sensing task. The y-axis is the absolute error between the predicted sensing quality score and the task performance of the testing deployment

with baseline methods and characterize relationships between assessment factors, optimal weights of each factor, and final assessment performance. Finally, a robustness test is conducted to evaluate how the performance varies with multiple human objects.

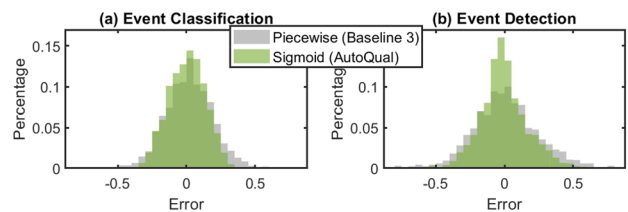
Since the TSQ score represents the prediction of the system performance at a given deployment, we calculate the absolute error between the predicted TSQ score and the task performance of each testing deployment to evaluate the performance of *AutoQual*. For each analysis, we randomly sample some deployments as the training deployments from the aforementioned datasets (Sect. 4) to train the TSQ scoring model, and use all the other deployments as the testing deployments to evaluate the model performance. The final result of each analysis is the evaluation of 2000 testing deployments. We repeat the random sample of the training and testing multiple times until the testing deployments number is not less than 2000. The default training size is 24.

### 5.1 *AutoQual* Comparisons to manual assessment (Baseline 1)

We compare *AutoQual*'s performance to the manual assessment approach using standard excitation as described in Sect. 4.2.2. Figure 10 shows these two approaches performance with absolute error of task performance prediction. The grey bars present manual assessment results using standard excitation, and the green bars present our *AutoQual* results with ambient footstep excitation. The manual assessment achieves an average absolute error of 0.093 and 0.129 for the two sensing tasks—event classification and event detection. *AutoQual* achieves a similar performance compared to the manual assessment using standard excitation, where it achieves an average absolute error of 0.096 and 0.099 for the two sensing tasks.



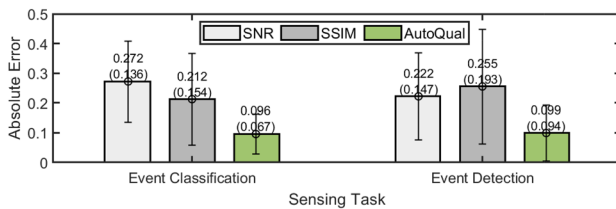
**Fig. 11** AF impact quantification analysis. The x-axis is the number of training deployments that we used to select the parameters. The y-axis is the absolute error between the TSQ score and task performance of the testing deployments. **a, b** Identification the performance of two methods for event classification and event detection, respectively



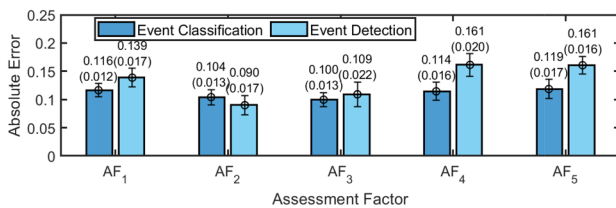
**Fig. 12** Analysis on error distribution of our method and the baseline method when only use eight deployments as the training deployment. **a, b** Show the result for event classification and event detection, respectively. The x-axis is the error between the TSQ score and the task performance of the testing deployment. The y-axis is the percentage of the error range in all errors. The error distribution of our method is more concentrated than the baseline method for in **(a, b)**, which identifies that our method has a better performance than the baseline method

In order to further illustrate the advantage of *AutoQual*, we compare the time duration needed for manual assessment and using *AutoQual* to achieve autonomous assessment with ambient footstep excitation. For the manual assessment using standard excitation, the user needs to mark the excitation locations on the floor (Fig. 7a) and generate the standard excitation (e.g., drop a tennis ball) at each location. We time the manual deployment assessment procedure five times, which reaches an average of 622 s ( $\sim 10$  min) with a standard deviation of 23.4 s. On the other hand, *AutoQual* does not require additional time for assessment data collection. In summary, *AutoQual* demonstrates the robustness of the assessment accuracy over different sensing tasks and the efficiency in terms of time consumption.





**Fig. 13** Comparing with SNR and SSIM baseline methods, the absolute error of *AutoQual* is at least 2X lower than them for the two sensing tasks. The x-axis is the sensing task. The y-axis is the absolute error between the predicted sensing quality score and the task performance of the testing deployment

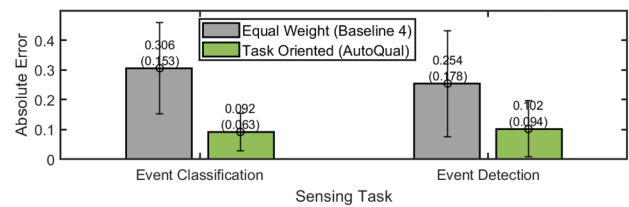


**Fig. 14** Single AF model analysis. The x-axis is the AF and the y-axis is the absolute error between the predicted TSQ score and sensing task F1 score. The models are trained with 24 deployments.  $AF_2$  and  $AF_3$  have a stronger impact for event classification than other AFs.  $AF_2$  has a stronger impact for event detection than other AFs

## 5.2 AF saturation function: sigmoid vs. piecewise (Baseline 2)

In order to verify the effectiveness of the non-linear projection for *AutoQual*, we compare our non-linear model to the Min–Max baseline as discussed in Sect. 4.2.3. Figure 11 demonstrates the absolute error of our approach (green bars) and the baseline method (grey bars). The x-axis is the training size for event classification and event detection assessment shown in (a) and (b), respectively. *AutoQual* achieves a lower error rate than the baseline method, especially when the number of training deployments is less than 10. With eight training deployments, the mean absolute error decreased 35% and 38% for event classification and event detection, respectively. When the training size increased from 8 to 24, the mean absolute error of our method decreased less than 0.031, which is 3X better than the decrease of the baseline method (0.081).

To further understand the performance of these two approaches with limited training deployments, we also analyze their error distribution. Figure 12a, b show the error distribution of the two methods with eight training deployments for event classification and event detection task performance prediction, respectively. We observe that our approach has



**Fig. 15** Multiple AFs integration model analysis. Our task-oriented data-driven model achieves at least 2.5X lower absolute error than the baseline method. The x-axis is the sensing tasks. The y-axis is the absolute error between the TSQ score and the sensing task performance over the testing deployments

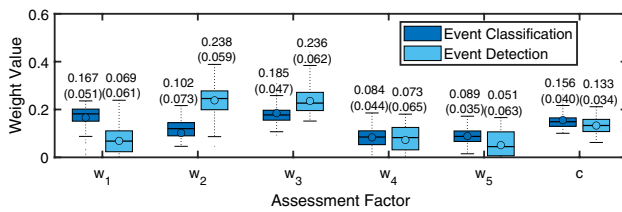
fewer outliers and presents a more concentrated distribution, which leads to a lower absolute error. The performance is improved due to two reasons. First, when the training size is small, the maximum and minimum measurements of each AF in the training set may not be sufficient to represent the model. Second, the baseline approach assumes that the relationship between each AF and system performance is linear in all ranges, which is not true as discussed in Sect. 3.3.1.

## 5.3 Assessment factor selections (Baseline 3)

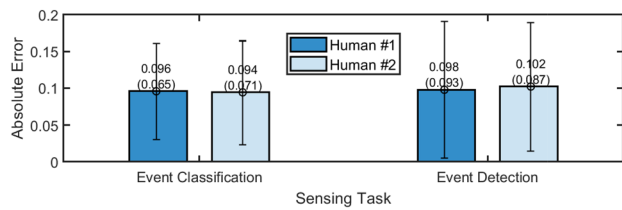
We compare the AFs used in *AutoQual* to the traditional signal quality metrics SNR and SSIM as discussed in Sect. 4.2.4. Figure 13 shows the absolute error of the system performance prediction over different assessment metrics. The results for baseline metrics are marked in grey bars, and the results for *AutoQual* are marked in green bars. We observe that for the event classification, the mean absolute error for *AutoQual* is 0.096, and the corresponding mean absolute error for event detection task is 0.99, which is 2X better than those of the SNR (0.27 and 0.22) and SSIM (0.207 and 0.25). In addition, when using *AutoQual* to predict the system performance, the standard deviation of the prediction is also lower than those of the baselines. In summary, *AutoQual* achieves a more accurate and stable sensing quality estimation compared to the baseline methods.

We further investigate the impact of each AF on the sensing task. To do so, we use single AF to implement *AutoQual*. Figure 14 shows the absolute error between the TSQ score and the task performance. We observe that, for the event classification task,  $AF_2$  and  $AF_3$  demonstrate lower error compared to other AFs. For the event detection task,  $AF_2$  achieves the best performance, which indicates that  $AF_2$  has the strongest impact.  $AF_2$  measures the ratio of the initial amplitude of the signal over the amplitude of background noise. The event detection algorithm is based on the signal to noise ratio (SNR). As a result, it is more sensitive to





**Fig. 16** Analysis on task-oriented integration model variation over different tasks. The x-axis is the AFs and the y-axis is the weight values. The models are trained with 24 deployments. The weight values of the same AF varies over different tasks, and the different AFs have varying weights for the same task



**Fig. 17** Analysis on human variation. The x-axis is the sensing task, and the y-axis is the absolute error between the TSQ score and sensing task performance. *AutoQual* demonstrate similar performance when the assessment is done on Human #1 and Human #2, which indicates robustness over different users

the SNR related AFs, i.e.,  $AF_1$  and  $AF_2$ .  $AF_3$  measures the directional distortion of signal propagation media. A higher  $AF_3$  value indicates that the same type of signals collected from different directions have less variation. This may lead to less structure variance induced data distribution shift within each class for the event classification task. In addition, when compared between tasks, the event classification task is more sensitive to  $AF_1$ ,  $AF_4$  and  $AF_5$  compared to the event detection task. This is because the event detection task does not require the signal to contain rich frequency components that reflect the structural mode excitation for fine-grained characterization.

#### 5.4 Integration methods: task-oriented vs. equal weight (Baseline 4)

In order to evaluate the efficiency of the AFs integration model, we compare our task-oriented data-driven integration model to the Equal Weight baseline as introduced in Sect. 4.2.5. Figure 15 shows the absolute error of our approach (green bars) and the baseline method (grey bars). For event classification assessment, the mean absolute error of our method (0.092) is near 3X lower than the baseline method (0.306); For event detection assessment, the mean

absolute error of our method (0.102) is near 2.5X lower than the baseline method (0.254). On the other hand, the standard deviation of our method also lower than the baseline method for both sensing tasks, which indicates that our approach is more stable. In summary, the task-oriented data-driven integration model is more stable and at least 2.5X better than the baseline method.

We further analyze the selected parameters of our integration model for these two sensing tasks. The weight values of the integration model (Eq. 8) reflect the contribution of different AFs to the assessment of sensing tasks. Figure 16 presents the weight values of the integration model with a training size of 24. We observe that different AFs have different weights for the same task. This indicates that the deployment characteristics (AF) affect the system performance. On the other hand, most AFs have different weights for different tasks. This indicates that different tasks favor distinct criteria for assessing sensing quality.

#### 5.5 Assessment robustness over different users

The human variation experiment evaluates the robustness of *AutoQual* when it utilizes multiple persons' footstep excitation for assessment. Prior work on structural vibration-based human sensing indicates that different people's footstep excitation have distinct characteristics (Pan et al. 2015, 2017). To validate the robustness of *AutoQual* using different people's footstep excitation, we collect two persons' footstep excitation in each deployment for auto-assessment. Figure 17 shows the assessment performance of *AutoQual* for event classification and event detection with two people's footstep excitation. We observe that *AutoQual* has a similar performance when using different people's footstep excitation as the assessment excitation. The mean absolute error of event classification assessment is 0.0096 and 0.0094 from human #1 and human #2, respectively. The mean absolute error of event detection assessment is 0.0098 and 0.0102 from human #1 and human #2, respectively. The difference of mean absolute error between two people's assessment performance is 0.002 and 0.004 for event classification and event detection, respectively. In summary, *AutoQual* is robust when using different people's footstep excitation to achieve the assessment.

## 6 Discussion

In this section, we further discuss the limitation of this work and possible future directions.

## 6.1 Robust gait cycle detection and gait-based temporal association

In this work, we rely on a mobile device on the calf for gait-based temporal association between different sensing modalities. However, mobile devices or wearables may come in various form factors, e.g., smartwatch (Genovese et al. 2017), earable (Prakash et al. 2019), on-cloth/limbs (Mokaya et al. 2013). Because these form factors are placed on different body parts, they may be sensitive to capturing different phases of gait cycles. In our future work, we plan to explore schemes that allow automatic calibration of this gait cycle phase offset between mobile devices/wearables and infrastructural sensors for robust temporal association. Furthermore, for some form factors, the detecting gait cycle phase may change over different contexts (e.g., smartwatch detecting people walking with/without carrying loads). To ensure robust gait-based temporal association, the system may adapt or selectively conduct temporal association based on the context.

## 6.2 Assessing a group of sensors on collaborative tasks

In this work, we only investigated applications/sensing tasks done by an individual sensor. However, there are sensing tasks that require multiple sensors to conduct collaborative computation, such as multilateration based localization (Mirshekari et al. 2018; Fagert et al. 2021). We plan to extend the current Assessment Factor (AF) concepts for individual sensors to sensor groups. This can be done by measuring the group factors over multiple devices assigned for the collaborative task. The group factors could be clock synchronization resolution, environment-circuit interference over different locations, etc. The TSQ score model will take both group and individual factors into account for collaborative task's sensing quality assessment.

## 6.3 Integration model

In this paper, we focus on individual AF's modeling, i.e., the saturation function. The individual AF model reflects

the specific causes of high/low learning accuracy. However, the linear regression model we adopted is limited in terms of representing the relation and dependency between different AFs. The advantage of our linear integration model is the selected parameters (weight) intuitively reveal the relationship between the AF and the task performance. The limitation of the linear relationship is that it may not be the precise representative for the relationship between individual AF and task performance. When two AFs are dependent, the relationship between the task performance and two AFs is no longer a linear combination, even the relationship between the task performance and the individual AF may be linear. To further improve the robustness of the integrated score, we plan to explore more approaches to estimate the integration models, such as kernel-based regression, general regression neural network, etc. in our future work.

## 7 Conclusion

In this paper, we present *AutoQual*, a sensing quality assessment framework. We introduce our Task-oriented Sensing Quality (TSQ) assessment model as well as the auto-assessment scheme that utilizes a co-located mobile sensing system to assess the performance of structural vibration sensing systems. Ambient human-induced signals are combined to infer the deployment's environmental and hardware factors that impact system performance. Real-world experiments are conducted with 48 deployments at 11 environments. *AutoQual* achieves less than 0.10 absolute error, which is an 2× improvement to baselines. Our auto-assessment approach predicting multiple tasks' performance at untested deployments shows a  $\leq 0.018$  absolute error difference to the manual assessment.

## 1. F1 Score of the sensing tasks in Sect. 4.1

See Table 2.

**Table 2** Sensing task F1 score in 48 deployments

Env-ID	Dep-ID	Event Classification								Event Det.
		LSVM	RSVM	RF	LR	GNB	XG-Boost	K-NN	ET	
1	1	0.87	0.75	.88	0.81	0.48	0.82	0.74	0.88	0.68
	2	0.79	0.71	0.81	0.76	0.43	0.81	0.67	0.79	0.62
	3	0.71	0.68	0.77	0.71	0.46	0.71	0.58	0.74	0.62
	4	0.81	0.74	0.82	0.77	0.48	0.78	0.73	0.80	0.62
2	5	0.60	0.54	0.61	0.59	0.46	0.69	0.48	0.63	0.46
	6	0.70	0.67	0.74	0.68	0.46	0.75	0.49	0.71	0.72
	7	0.72	0.68	0.73	0.73	0.47	0.74	0.51	0.66	0.76
	8	0.59	0.56	0.52	0.59	0.45	0.48	0.31	0.46	0.37
3	9	0.72	0.64	0.66	0.69	0.53	0.71	0.38	0.74	0.78
	10	0.64	0.63	0.67	0.62	0.44	0.71	0.49	0.64	0.71
	11	0.71	0.66	0.69	0.74	0.48	0.66	0.43	0.70	0.68
	12	0.73	0.63	0.69	0.70	0.48	0.71	0.42	0.62	0.61
4	13	0.70	0.65	0.74	0.73	0.39	0.70	0.42	0.72	0.68
	14	0.64	0.59	0.66	0.62	0.42	0.71	0.35	0.68	0.64
	15	0.75	0.69	0.80	0.74	0.56	0.73	0.52	0.76	0.58
	16	0.73	0.66	0.73	0.72	0.50	0.72	0.45	0.68	0.62
5	17	0.42	0.41	0.44	0.46	0.29	0.44	0.39	0.46	0.36
	18	0.51	0.46	0.42	0.51	0.34	0.47	0.36	0.47	0.43
	19	0.44	0.42	0.46	0.45	0.37	0.45	0.33	0.50	0.45
	20	0.47	0.43	0.47	0.50	0.43	0.43	0.36	0.48	0.33
6	21	0.46	0.44	0.45	0.54	0.30	0.50	0.38	0.47	0.67
	22	0.54	0.52	0.47	0.59	0.28	0.56	0.44	0.44	0.73
	23	0.47	0.46	0.44	0.48	0.26	0.46	0.41	0.47	0.68
	24	0.59	0.56	0.51	0.62	0.32	0.62	0.41	0.52	0.68
7	25	0.49	0.43	0.46	0.50	0.25	0.48	0.44	0.47	0.60
	26	0.47	0.47	0.48	0.49	0.29	0.52	0.44	0.48	0.59
	27	0.56	0.50	0.48	0.54	0.33	0.53	0.40	0.49	0.59
	28	0.47	0.42	0.41	0.45	0.28	0.43	0.35	0.43	0.48
8	29	0.84	0.78	0.69	0.82	0.37	0.77	0.64	0.72	0.71
	30	0.83	0.81	0.75	0.82	0.35	0.84	0.71	0.79	0.66
	31	0.87	0.77	0.81	0.83	0.41	0.83	0.65	0.83	0.63
	32	0.84	0.77	0.72	0.83	0.38	0.71	0.68	0.75	0.65
9	33	0.67	0.68	0.67	0.69	0.34	0.74	0.55	0.76	0.76
	34	0.75	0.71	0.69	0.72	0.34	0.68	0.56	0.67	0.72
	35	0.74	0.70	0.73	0.75	0.29	0.69	0.59	0.65	0.73
	36	0.80	0.73	0.71	0.78	0.33	0.72	0.66	0.70	0.72
10	37	0.56	0.45	0.57	0.55	0.38	0.59	0.41	0.57	0.61
	38	0.52	0.48	0.55	0.53	0.39	0.64	0.35	0.54	0.63
	39	0.62	0.55	0.63	0.64	0.40	0.70	0.42	0.58	0.64
	40	0.53	0.43	0.55	0.55	0.39	0.54	0.39	0.55	0.64
11	41	0.63	0.58	0.74	0.66	0.47	0.68	0.54	0.69	0.62
	42	0.56	0.52	0.48	0.58	0.31	0.53	0.35	0.49	0.18
	43	0.45	0.44	0.41	0.49	0.39	0.43	0.36	0.41	0.30
	44	0.54	0.49	0.48	0.53	0.36	0.44	0.35	0.47	0.31
	45	0.32	0.30	0.32	0.34	0.28	0.33	0.26	0.26	0.16
	46	0.40	0.36	0.32	0.41	0.30	0.35	0.31	0.36	0.10
	47	0.41	0.42	0.43	0.45	0.32	0.44	0.38	0.45	0.29
	48	0.37	0.34	0.31	0.37	0.25	0.36	0.29	0.29	0.04

**Acknowledgements** This research was supported by a 2020 Seed Fund Award from CITRIS and the Banatao Institute at the University of California.

## Declarations

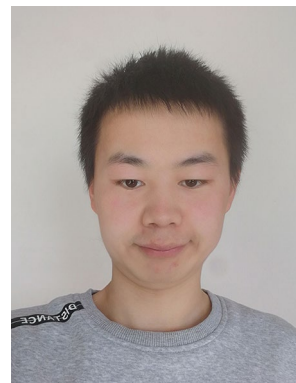
**Conflict or interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adafruit feather m0 bluefruit le, <https://www.adafruit.com/product/2995>, Accessed 14 Mar 2021. (2021)
- Baccelli, F., Błaszczyszyn, B., et al.: Stochastic geometry and wireless networks: volume ii applications. *Found. Trends® Netw.* **4**(1–2), 1–312 (2010)
- Bales, D., Tarazaga, P.A., Kasarda, M., Batra, D., Woolard, A.G., Poston, J.D., Malladi, V.S.: Gender classification of walkers via underfloor accelerometer measurements. *IEEE Internet Things J.* **3**(6), 1259–1266 (2016)
- Banerjee, T., Sheth, A.: Iot quality control for data and application needs. *IEEE Intell. Syst.* **32**(2), 68–73 (2017)
- Boano, C.A., Voigt, T., Dunkels, A., Osterlind, F., Tsiftes, N., Mottola, L., Suarez, P.: Exploiting the lqi variance for rapid channel quality assessment. In: *Proceedings of the 2009 International Conference on Information Processing in Sensor Networks*. IEEE Computer Society, pp. 369–370 (2009)
- Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152 (1992)
- Boyd, S., Boyd, S.P., Vandenberghe, L.: *Convex Optimization*. Cambridge University Press, Cambridge (2004)
- Cai, L., Zhu, Y.: The challenges of data quality and data quality assessment in the big data era. *Data Sci. J.* **14**, 1–10 (2015)
- Chen, M.-J., Bovik, A.C.: Fast structural similarity index algorithm. *J. Real-Time Image Process.* **6**(4), 281–287 (2011)
- Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785–794 (2016)
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y.: Xgboost: extreme gradient boosting. R package version **4–2**, 1–4 (2015)
- Chen, X., Xu, S., Liu, X., Xu, X., Noh, H.Y., Zhang, L., Zhang, P.: Adaptive hybrid model-enabled sensing system (hmss) for mobile fine-grained air pollution estimation. *IEEE Trans. Mob. Comput.* (2020)
- Clemente, J., Li, F., Valero, M., Song, W.: Smart seismic sensing for indoor fall detection, location, and notification. *IEEE J. Biomed. Health. Inf.* **24**(2), 524–532 (2019)
- Clifton, W., Frank, A., Freeman, S.-M.: Osteopetrosis (marble bones). *Am. J. Dis. Child.* **56**, 1020 (1938)
- Devarakonda, S., Sevusu, P., Liu, H., Liu, R., Iftode, L., Nath, B.: Real-time air quality monitoring through mobile sensing in metropolitan areas. In: *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*, pp. 1–8 (2013)
- Fagert, J., Mirshekari, M., Pan, S., Lowes, L., Iammarino, M., Zhang, P., Noh, H.Y.: Structure-and sampling-adaptive gait balance symmetry estimation using footstep-induced structural floor vibrations. *J. Eng. Mech.* **147**(2), 04020151 (2021)
- Fortin-Simard, D., Bilodeau, J.-S., Gaboury, S., Bouchard, B., Bouzouane, A.: Human activity recognition in smart homes: Combining passive rfid and load signatures of electrical devices. In: *2014 IEEE symposium on intelligent agents (IA)*. IEEE pp. 22–29 (2014)
- Freund, Y., Schapire, R., Abe, N.: A short introduction to boosting. *J. Jpn. Soc. For. Artif. Intell.* **14**, 771–780 (1999)
- Gao, Y., Dong, W., Guo, K., Liu, X., Chen, Y., Liu, X., Bu, J., Chen, C.: Mosaic: A low-cost mobile sensing system for urban air quality monitoring. In: *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*. IEEE pp. 1–9 (2016)
- Genovese, V., Mannini, A., Sabatini, A.M.: A smartwatch step counter for slow and intermittent ambulation. *IEEE Access* **5**, 13 028–13 037 (2017)
- Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. *Mach. Learn.* **63**(1), 3–42 (2006)
- Goldberger, J., Hinton, G.E., Roweis, S.T., Salakhutdinov, R.R.: Neighbourhood components analysis. *Adv. Neural Inf. Process. Syst.* **17**, 513–520 (2005)
- Grimmer, M., Schmidt, K., Duarte, J.E., Neuner, L., Koginov, G., Riener, R.: Stance and swing detection based on the angular velocity of lower limb segments during walking. *Front. Neurobot.* **13**, 57 (2019)
- Hanlon, M., Anderson, R.: Real-time gait event detection using wearable sensors. *Gait Posture* **30**(4), 523–527 (2009)
- Hu, Z., Yu, T., Zhang, Y., Pan, S.: Fine-grained activities recognition with coarse-grained labeled multi-modal data. In: *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, pp. 644–649 (2020)
- Islam, A.N., Lohan, E. S., Renfors, M.: Moment based cnr estimators for BOC/BPSK modulated signal for Galileo/GPS, In: *5th Workshop on Positioning, Navigation and Communication*. IEEE, pp. 129–136 (2008)
- Jiang, Y., Li, K., Tian, L., Piedrahita, R., Yun, X., Mansata, O., Lv, Q., Dick, R. P., Hannigan, M., Shang, L.: Maqs: a personalized mobile sensing system for indoor air quality monitoring. In: *Proceedings of the 13th International Conference on Ubiquitous Computing*, pp. 271–280 (2011)
- Karkouch, A., Mousannif, H., Al Moatassime, H., Noel, T.: Data quality in internet of things: a state-of-the-art survey. *J. Netw. Comput. Appl.* **73**, 57–81 (2016)
- Kleinbaum, D.G., Dietz, K., Gail, M., Klein, M., Klein, M.: *Logistic Regression*. Springer, New York, USA (2002)
- Li, C., Bovik, A. C.: Three-component weighted structural similarity index, In: *Image quality and system performance VI*, vol. 7242. International Society for Optics and Photonics, p. 72420Q (2009)
- Liaw, A., Wiener, M., et al.: Classification and regression by randomforest. *R News* **2**(3), 18–22 (2002)
- LMV3xx Low-Voltage Rail-to-Rail Output Operational Amplifier, Texas Instruments Incorporated, 5 (2020)
- Mirshekari, M., Pan, S., Fagert, J., Schooler, E.M., Zhang, P., Noh, H.Y.: Occupant localization using footstep-induced structural vibration. *Mech. Syst. Signal Process.* **112**, 77–97 (2018)
- Mirshekari, M., Fagert, J., Pan, S., Zhang, P., Noh, H.Y.: Step-level occupant detection across different structures through footstep-induced floor vibration using model transfer. *J. Eng. Mech.* **146**(3), 04019137 (2020)

- Mokaya, F., Nguyen, B., Kuo, C., Jacobson, Q., Rowe, A., Zhang, P.: Mars: a muscle activity recognition system enabling self-configuring musculoskeletal sensor networks. In: 2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). IEEE, pp. 191–202 (2013)
- Moon, K.S., Lee, S.Q., Ozturk, Y., Gaidhani, A., Cox, J.A.: Identification of gait motion patterns using wearable inertial sensor network. *Sensors* **19**(22), 5024 (2019)
- Öberg, T., Karsznia, A., Öberg, K.: Basic gait parameters: reference data for normal subjects, 10–79 years of age. *J. Rehabil. Res. Dev.* **30**, 210–210 (1993)
- Oppenheim, A.V., Schaffer, R.W.: Digital Signal Processing(book). Research supported by the Massachusetts Institute of Technology, Bell Telephone Laboratories, and Guggenheim Foundation, Englewood Cliffs, 9. 598 (1975). (**Prentice-Hall, Inc.**)
- Pan, S., Nguyen, P.: Opportunities in the cross-scale collaborative human sensing of 'developing' device-free and wearable systems. In: Proceedings of the 2nd ACM Workshop on Device-Free Human Sensing, pp. 16–21 (2020)
- Pan, S., Bonde, A., Jing, J., Zhang, L., Zhang, P., Noh, H.Y.: Boes: building occupancy estimation system using sparse ambient vibration monitoring. In: Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2014, vol. 9061. International Society for Optics and Photonics, pp. 90611O-1–90611O-16 (2014)
- Pan, S., Wang, N., Qian, Y., Velibeyoglu, I., Noh, H. Y., Zhang, P.: Indoor person identification through footprint induced structural vibration. In: Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications, pp. 81–86 (2015)
- Pan, S., Yu, T., Mirshekari, M., Fagert, J., Bonde, A., Mengshoel, O.J., Noh, H.Y., Zhang, P.: Footprintid: indoor pedestrian identification through ambient structural vibration sensing. *Proc. ACM Interact. Mob. Wear. Ubiquit. Technol.* **1**(3), 1–31 (2017)
- Pipino, L.L., Lee, Y.W., Wang, R.Y.: Data quality assessment. *Commun. ACM* **45**(4), 211–218 (2002)
- Prakash, J., Yang, Z., Wei, Y.-L., Choudhury, R. R.: Stear: Robust step counting from earables. In: Proceedings of the 1st International Workshop on Earable Computing, pp. 36–41 (2019)
- Rish, I., et al.: An empirical study of the naive bayes classifier. In: IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, 3(22), pp. 41–46 (2001)
- Schütze, H., Manning, C.D., Raghavan, P.: Introduction to Information Retrieval, vol. 39. Cambridge University Press, Cambridge (2008)
- Sellitto, C., Burgess, S., Hawking, P.: Information quality attributes associated with rfid-derived benefits in the retail supply chain. *Int. J. Retail Distrib. Manag.* **35**(1), 69–87 (2007)
- Shapiro, S., Kneib, G.: Seismic attenuation by scattering: theory and numerical results. *Geophys. J. Int.* **114**(2), 373–391 (1993)
- Shi, L., Mirshekari, M., Fagert, J., Chi, Y., Noh, H. Y., Zhang, P., Pan, S.: Device-free multiple people localization through floor vibration. In: Proceedings of the 1st ACM International Workshop on Device-Free Human Sensing, pp. 57–61 (2019)
- SM-24 Geophone Element, Input/Output, Inc., 4 2006, rev. 3
- Small, Low Power, 3-Axis  $\pm 3$  g Accelerometer, Analog Devices, Inc., (2010)
- Sparkfun samd21 mini breakout, <https://www.sparkfun.com/products/13664>. Accessed 14 Mar 2021. (2021)
- Srinivasan, K., Lewis, P.: Rssi is under appreciated, In: Proceedings of the Third Workshop on Embedded Networked Sensors (EmNets), vol. 2006. Cambridge, USA, MA. pp. 1–5 (2006)
- Stein, S., Wyssession, M.: An Introduction to Seismology, Earthquakes, and Earth Structure. Wiley, Hoboken, New Jersey, USA (2009)
- Van den Branden Lambrecht, C.: Special issue on image and video quality metrics. *Signal Process.* **70**(3), 153–154 (1998)
- van der Togt, R., Bakker, P.J., Jaspers, M.W.: A framework for performance and data quality assessment of radio frequency identification (rfid) systems in health care settings. *J. Biomed. Inf.* **44**(2), 372–383 (2011)
- Viktorov, I.A.: Rayleigh and Lamb Waves: Physical Theory and Applications. Plenum Press, Berlin (1970)
- Wang, Z., Bovik, A.C.: A universal image quality index. *IEEE Signal Process. Lett.* **9**(3), 81–84 (2002)
- Wang, Z., Lu, L., Bovik, A.C.: Video quality assessment based on structural distortion measurement. *Signal Process. Image Commun.* **19**(2), 121–132 (2004)
- Xu, S., Chen, X., Pi, X., Joe-Wong, C., Zhang, P., Noh, H. Y.: Vehicle dispatching for sensing coverage optimization in mobile crowdsensing systems. In: 2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN). IEEE, pp. 311–312 (2019a)
- Xu, S., Chen, X., Pi, X., Joe-Wong, C., Zhang, P., Noh, H.Y.: ilocus: incentivizing vehicle mobility to optimize sensing distribution in crowd sensing. *IEEE Trans. Mob. Comput.* **19**(8), 1831–1847 (2019b)
- Yi, T.-H., Li, H.-N., Zhao, X.-Y.: Noise smoothing for structural vibration test signals using an improved wavelet thresholding technique. *Sensors* **12**(8), 11 205–11 220 (2012)
- Ying, H., Silex, C., Schnitzer, A., Leonhardt, S., Schiek, M.: Automatic step detection in the accelerometer signal. In: 4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007). Springer, pp. 80–85 (2007)
- Yu, T., Zhang, Y., Hu, Z., Xu, S., Pan, S.: Vibration-based indoor human sensing quality reinforcement via Thompson sampling. In: Proceedings of the First International Workshop on Cyber-Physical-Human System Design and Implementation, pp. 33–38 (2021)
- Zhang, Y., Zhang, L., Noh, H. Y., Zhang, P., Pan, S.: A signal quality assessment metrics for vibration-based human sensing data acquisition. In: Proceedings of the 2nd Workshop on Data Acquisition To Analysis, pp. 29–33 (2019)



**Yue Zhang** received the Master's degree and Bachelor's degree in the Department of Electronic Engineering at Tsinghua University in 2019 and 2016, respectively. He is an exchange scholar since the 2019 fall at the University of California Merced and will join PANSLab at the University of California Merced as a Ph.D. student in the 2021 fall. His research interests include sensor networks, embedded systems, indoor human information acquisition.



**Zhizhang Hu** received the Master's degree from the Carnegie Mellon University 2020. Currently, He is a Ph.D. Student majoring electrical engineering and computer science at the University of California Merced. His research interests include multi-modal human sensing and learning, and artificial intelligence for IoT systems.





**Susu Xu** is an assistant professor at Stony Brook University. She received Ph.D. in Advanced Infrastructure Systems and M.S. in Machine Learning from Carnegie Mellon University and Bachelor's degree in Engineering from Tsinghua University. The foci of her research include mobile sensing, machine learning, urban computing, and smart infrastructure systems. She works on developing and deploying crowdsensing systems, machine learning algorithms, and incentive mechanisms to

help understand urban dynamics and enable a self-adaptive smart city. From 2019 to 2020, she was a Postdoctoral Research Fellow at Stanford University and a researcher at Qualcomm AI research.



**Shijia Pan** is an assistant professor in the Computer Science and Engineering Department at the University of California Merced. She received her Bachelor's degree in Computer Science and Technology from the University of Science and Technology of China and her Ph.D. degree in Electrical and Computer Engineering at Carnegie Mellon University. Her research interests include cyber-physical systems, Internet-of-Things, and ubiquitous computing. She worked in multiple disciplines and focused

on self-assessing and self-adaptive heterogeneous cyber-physical systems for accurate human information inference.