MM-Tap: Adaptive and Scalable Tap Localization on Ubiquitous Surfaces with mm-level Accuracy

Yandao Huang, *Student Member, IEEE*, Cong Li, Fuwen Chen, Qian Zhang, *Fellow, IEEE*, and Kaishun Wu, *Fellow, IEEE*

***Abstract*—Transforming physical surfaces into virtual interfaces can extend the interaction capability of many exciting metaverse applications in the future. Recent advances in vibration-based tap sensing show promise for this vision using passive vibration signals. However, current approaches based on Time-Difference-of-Arrival (TDoA) triangulation suffer the impact of fluctuant wave velocity due to the dispersive and heterogeneous nature of solid mediums, failing to meet the performance requirement for practical use. In this paper, we present MM-Tap, a vibration-based tap localization system that can transform ubiquitous surfaces into virtual touch screens with low overhead. A novel localization scheme is proposed based on the finding of spatio-temporal mapping between tap locations and TDoA values, which pushes the accuracy limits of vibration-based tap sensing from unstable cm-level to mm-level. We investigate the geometry of the sensor layout and design a model-based method to synthesize tap data, which enables MM-Tap to adapt to various surface materials and respond to arbitrary sensing scales after a few seconds of calibration. We combine MM-Tap with a COTS projector and facilitate a digitally augmented surface where users can play video games with low latency.**

***Index Terms*—Tap Sensing, Vibration-based Sensing, Localization, Human-Computer Interaction**

I. Introduction

The emergence of touch screens on electronic devices has largely improved our daily productivity. However, large-sized touchscreens incur high production costs and require complex installation processes, making them impractical for ubiquitous deployment in the real world. Flat surfaces such as tables, walls, and floors are abundant in our living spaces. What if we can convert these physical surfaces into interactive interfaces at a low cost? This innovation has enabled numerous novel and exciting applications that enhance our daily life experience, ranging from distributed interactive tables in a meeting room or a classroom, ubiquitous location-based control panels, to immersive gaming in the metaverse.

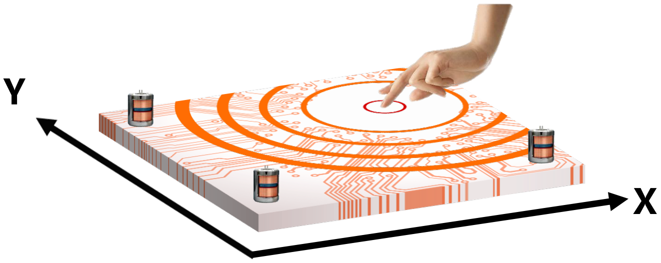


Fig. 1: MM-Tap can localize finger tap on ubiquitous surfaces with three geophone sensors. sensors.

Much research has been developed to sense finger touch on the physical surface. Vision-based approaches [10]–[13] are prevalent but require line-of-sight condition and high computation overhead. Recently, SurfaceVibe [16] utilizes TDoA triangulation to build a vibration-based sensing system that supports finger tap and swipe on solid surfaces using four geophone sensors. However, it can only provide cm-level accuracy, and the localization error is unacceptable when enlarging the sensing area. UbiTap [17] proposes a small-scale tap localization system with mm-level accuracy. It exploits the acoustic dispersion properties and infers the tap location based on the TDoA between air-borne and solid-borne sound signals. However, it requires a delicate position setting of three smartphones, which harms the usability. In addition, acoustic-based systems are sensitive to burst noise and fail to work when enlarging the sensing area.

In this paper, we propose MM-Tap, a vibration-based tap localization system with mm-level accuracy. As shown in Figure 1, MM-Tap builds upon the analysis of tap-induced vibration waves collected by three geophone sensors. It can transform ubiquitous surfaces into digitally augmented surfaces. Unlike previous work that applies TDoA triangulation for tap localization, MM-Tap proposes an advanced localization scheme to increase accuracy and stability. The key observation is that the TDoA values between sensors reflect the spatial information of finger tap, and we characterize this spatio-temporal mapping relationship using probabilistic regression models. MM-Tap can adapt to varying surface material and scale to arbitrary sensing area size after calibration within

Manuscript received September 21, 2022; revised April 21, 2023; accepted May 24, 2023. This research is supported in part by China NSFC Grant U2001207, Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things, the Project of DEGP (No. 2021ZDZX1068), RGC under Contract CERG 16203719, 16204820, 16206122, Contract R8015, and FS112. (Corresponding authors: Qian Zhang; Kaishun Wu.)

Yandao Huang is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong and also with the IoT Thrust, Information Hub, Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511466, China (e-mail: yhuangfg@connect.ust.hk).

Cong Li and Fuwen Chen is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: 2100271009@email.szu.edu.cn, 2018151017@email.szu.edu.cn).

Qian Zhang is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: qianzh@cse.ust.hk).

Kaishun Wu is with the DSA IoT Thrust, Information Hub, Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511466, China (e-mail: wuks@ust.hk).

a few seconds. Compared with the capacitive touch screen, MM-Tap reduces the cost of transforming a physical plane into a virtual interface by 10x while having more installation flexibility.

However, it is non-trial to implement a tap sensing system with high accuracy, stability, and usability using passive tap-induced vibration only. First, it is unclear how we can localize a finger tap with mm-level accuracy. The dispersion effect and heterogeneity nature of different materials lead to velocity fluctuation during the propagation of vibration waves. The state-of-the-art (SOTA) vibration-based sensing systems [16] mitigate this effect by optimizing a global wave velocity but can only provide unstable cm-level localization. Second, since the TDoA triangulation does not meet our precision requirement, it seems that we have to resort to learning-based methods for tap localization. However, it is impractical to let users collect location fingerprints for system initialization at each re-deployment. How can our system adapt to a new surface and scale to an arbitrary sensing layout with low human effort? Third, infrastructure-based localization requires users to measure the precise coordinate of sensors which harms the usability and flexibility of tap sensing systems on ubiquitous surfaces. Is it possible to deploy the system without measuring the 2D coordinate of sensors on the surface?

To address the above challenges, we first investigate how the fluctuant wave velocity affects the localization accuracy of TDoA triangulation and find that a mere time delay pattern between sensors can still reveal spatial information of tap location, even if we ignore the velocity parameter. We design a two-stage time delay estimator to obtain precise TDoA values. In order to project the measured TDoA values into the coordinate of the tap location, we employ a probabilistic regression model based on the linearity characteristics of the TDoA pattern. To avoid collecting a large number of location fingerprints for training, we design a synthetic data generator based on the observation of data structure. The users only need to tap on a few calibration points within seconds for initialization. In addition, a non-Euclidean distance metric is defined to release the burden of measuring sensor layout. Under this new distance metric, we can exploit the geometrical relationship between sensor intervals and sensing area and form a new deployment scheme that users no longer need to measure the exact coordinate of ambient sensors.

We evaluate MM-Tap across three types of common surface materials and four settings of sensing scales. For an 80 cm × 80 cm sensing area, MM-Tap can achieve median localization errors of 0.9 mm, 2.9 mm, and 3.6 mm for glass, acrylic, and wooden board, respectively. The experiment results show that our proposed methods enable MM-Tap to quickly adapt to new environments and various sensing scales without sacrificing accuracy and usability. Combining MM-Tap with a COTS projector, a digitally augmented surface is created on a real-world table for the user study, which shows that users can play exciting 3D shooting games with low latency (Demo: https://youtu.be/nQBnXOpntsc).

The main contributions of MM-Tap are summarized as follows:

- We propose MM-Tap to transform ubiquitous physical surfaces into interactive surfaces. We push the limits of vibration-based tap sensing on solid surfaces. The interaction powered by MM-Tap can achieve mm-level accuracy with low latency.
- We investigate the properties of solid-borne vibration signals and propose a new localization scheme for tap sensing. Our localization scheme relieves the pain of measuring the exact coordinates of ambient sensors. We design a two-stage time delay estimation method to improve the measurement accuracy and stability of TDoA values.
- We establish a model-based synthetic data generator that successfully synthesizes sufficient training data to support the regression-based localization. The regression model can be arbitrarily calibrated for a new environment and new sensing scale with extremely low effort.
- We conduct extensive experiments to validate the accuracy, adaptability, and generalizability of MM-Tap. In addition, we prototype a low-cost interactive projector using MM-Tap and conduct a real-world user study to prove its usability.

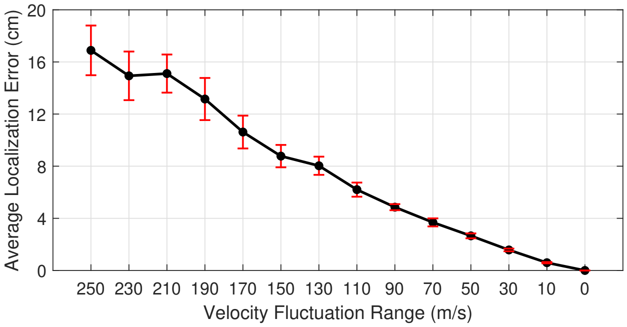


Fig. 2: The simulated localization errors of TDoA triangulation under different velocity fluctuation ranges.

II. Preliminary

A. TDoA Error Analysis

TDoA triangulation has been widely used for localization [19]–[24], [32]. Suppose Δt_{ij} is the TDoA value measured between the i-th and the j-th sensor (i = {1,2,3}, j = i % 3 + 1), we can derive the distance difference as

$$\Delta r_{ij} = r_i - r_j = v_g \cdot \Delta t_{ij}, \tag{1}$$

where v_g is the global wave velocity. The intersections of three hyperbolas indicated by Eq (1) are possible target locations.

However, TDoA-based systems can not provide stable and fine-grained results when localizing the passive vibration source (e.g., tap on a surface). The reason is that previous work assumes the wave propagation velocity on the surface is consistent while it is not [16]. The surface structure is heterogeneous even if made in the same material with an ideal craft process. The heterogeneous nature of the propagation medium leads to inconsistent wave velocity at different locations [51], [52]. If we set a global velocity for TDoA localization, then errors will be introduced when solving the above TDoA equations.

We measure the wave velocity (m/s) of different touchpoints on three types of boards, including toughened glass, acrylic and , wood. These boards have the same dimension of 120 cm × 120 cm × 1.5 cm. The velocity fluctuation ranges (VFR) of glass, acrylic, and wood boards fall in the range of [108, 171], [57,137], and [141, 269], respectively. In order to better understand the impact of fluctuant velocity, we conduct a simulation experiment. We consider a solid surface with a size of 100 $cm \times 100\ cm$. The solid-borne vibration wave propagates at different velocities in the range from 50 to 300 m/s [16], [25]. We evenly distribute 100 locations on the surface and assign a random velocity $v \in [v_{min}, v_{max}]$ for each location. The degree of velocity fluctuation is controlled by simultaneously increasing v_{min} and decreasing v_{max} with a step size of 10 m/s. The measurement of TDoA values for each location is assumed to be perfect and calculated based on the ground truth coordinate and the assigned velocity. Therefore, the only influence factor will be the setting of $v_g = (v_{min} + v_{max})/2$ when solving TDoA equations. The Chan algorithm [23] is used to find the solutions (i.e., intersections).

Figure 2 shows the simulation results regarding different velocity fluctuation ranges (VFR). The average localization error is 16.89 cm when VFR = 250 m/s, and the error decreases to 0.60 cm when VFR = 10 m/s. In the real world, the fluctuation may not be as high as 250 m/s. Previous work [16] reports localization errors ranging from 5.1 cm to 18.4 cm under a comparable surface size, which is consistent with our simulation.

Summary of observation 1: Due to the heterogeneous nature of solid mediums, wave velocity becomes an error term when using TDoA triangulation. We need to eliminate the velocity parameter to improve localization accuracy.

B. Spatio-temporal Mapping

If the wave velocity is not applicable, TDoA triangulation cannot be used to solve the target coordinates. But what if we only consider the TDoA values as the location hint? We conduct an experiment in the real world and measure TDoA values from 25 touchpoints (i.e., the blue area in Figure 4(a)) on a glass surface. Each point is tapped five times, and the corresponding TDoA values are plotted in Figure 3. Interestingly, we observe that the TDoA values have a linear relationship with the locations (e.g., the row with location index from 1 to 5). We further depict the 2D scatter map of TDoA value pairs at the same location in Figure 4(b). The scatter point forms an irregular quadrilateral using the sensor layout in Figure 4(a).

Summary of observation 2: It seems like mere TDoA values can provide special insight about tap locations. But the challenge is how we can model the spatio-temporal mapping relationship between tap coordinates and TDoA values while generalizing to different surface materials and sensing scales with low user effort.

III. System Overview

A. Design Goals and Challenges

MM-Tap is designed to meet the following goals.

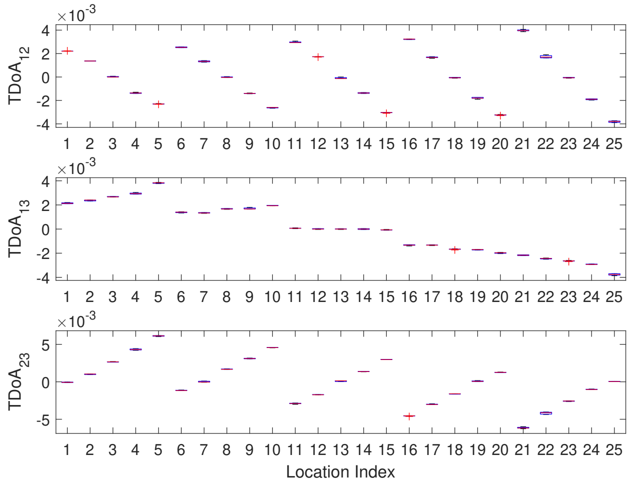
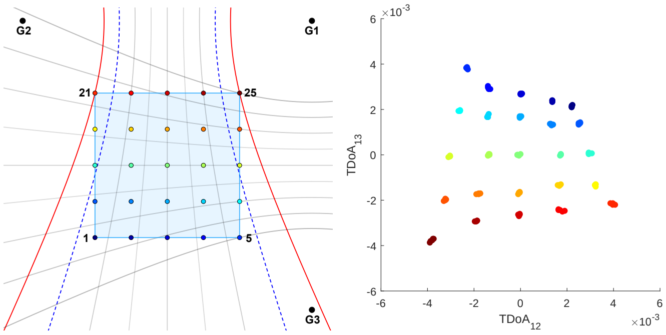


Fig. 3: The measured TDoA values of 25 locations on a glass broad show a linear pattern.



(a) Hyperbola layout with three sensors

(b) TDoA pattern in a irregular quadrilateral shape

Fig. 4: The spatio-temporal relationship between tap locations and their corresponding TDoA values.

Fine-grained: The motivation of MM-Tap is to provide fine-grained tap localization at any physical surface. The localization accuracy of current vibration-based approaches does not fulfill the needs of practical use. We aim to push the limits of tap sensing accuracy from unstable cm-level to stable mm-level. However, as shown in our preliminary study, the traditional TDoA-based localization scheme is not applicable. We need to propose a new localization scheme.

Adaptive and scalable: Classification-based tap localization systems rely on collecting samples at pre-defined points. The trained model will fail to work when moving the system to a new environment. In addition, a user-friendly interaction system should not put too much burden on users to re-train the system. It is non-trivial to find an appropriate adaptation scheme when the system is deployed on a new surface or scaled to an arbitrary size of sensing area.

Easy to deploy: Infrastructure-based tap localization systems require a delicate setting of ambient sensors. To guarantee localization accuracy, users have to measure the distance between sensors. We want to release this restriction and support quick and easy deployment even if users do not know the exact coordinate of ambient sensors.

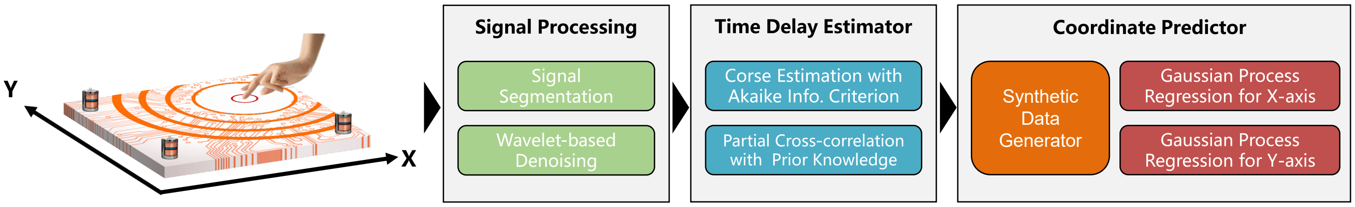

Y
X
Signal Processing
Signal Segmentation
Wavelet-based Denoising
Time Delay Estimator
Corse Estimation with Akaike Info. Criterion
Partial Cross-correlation with Prior Knowledge
Coordinate Predictor
Synthetic Data Generator
Gaussian Process Regression for X-axis
Gaussian Process Regression for Y-axis

Fig. 5: System overview of MM-Tap.

Low-latency: Last but not least, we should give tap feedback to the user without noticeable latency (¡100 ms [50]). This indicates that we need to achieve the above-mentioned design goals while not adopting complex algorithms that may increase the interaction latency.

B. *System Workflow*

Figure 5 shows the system architecture of MM-Tap, which comprises three main modules to enable an adaptive, scalable, and fine-grained tap sensing system using passive vibration signals.

Signal Processing: The signal processing module takes charge of detecting tap-induced vibration signals across sensor channels. A wavelet-based filter is designed to combat the dispersive and reflective nature of solid mediums, extracting the direct path waveform with L-infinity norm.

Two-stage Time Delay Estimator: Measuring the accurate time delay of directly arriving signals is critical for tap sensing with mm-level accuracy in this paper. We propose a two-stage time delay estimation scheme that can extract more precise TDoA values than previous work. The first stage is a coarse picking of tap-induced vibration signals with the Akaike Information Criterion (AIC). AIC can also provide prior knowledge, that is, the signal arriving order at sensors. In the second stage, we adopt an effective trick that selects partial signals for calculating cross-correlation coefficients. With the help of prior knowledge in the first stage, we can further exclude the false choice of time delay and get a more accurate estimation of TDoA values in the end.

Coordinate Predictor: Since the model-based localization algorithm (i.e., TDoA triangulation) is not applicable. We treat the problem as coordinate regression using the obtained TDoA features. It is impossible to ask users to collect training samples by tapping all the target locations before using them. In addition, the regression model should also be able to adapt to different surface materials and sensing scales. We address these challenges by proposing a model-based synthetic data generator, which can characterize the TDoA pattern of the whole sensing area with extremely low human effort. Then, the synthetic TDoA data is fed into two Gaussian process regression models for training. Finally, given the TDoA measurement from any tap location, the system can predict its 2D coordinate with mm-level accuracy.

IV. MM-Tap System

A. *Signal Processing*

1) Tap Detection: The vibration wave caused by finger tap is dominated by Rayleigh waves (decaying $\propto distance^{1/2}$) which have a slower velocity, lower frequency but stronger power compared to other mechanical waves like P- and S-waves [27]. Three geophone sensors are deployed to capture tap-induced vibration since they are more sensitive to the wave we collect. The tap-induced vibration has a pulse-like waveform, and we can apply an energy-based sliding window algorithm [15] to get the segment efficiently. The segment length is set to be 2000 sample points, which is long enough to cover the useful vibration signals. We denote the segmented vibration signals of the i-th channel as $x_i(t)$.

2) Wavelet-based Denoising: The tap-induced vibration signals are nonlinear and non-stationary time-series. Wavelet-based decomposition is well suited for denoising such signals [29]–[31]. In order to obtain high resolution in both time and frequency domains, we utilize continuous wavelet transform (CWT) [28] to decompose the tap-induced vibration signals. Specifically, CWT can be expressed as:

$$CWT_{x_i}(\alpha, \tau; \Psi) = \frac{1}{\sqrt{\alpha}} \int_{-\infty}^{+\infty} x_i(t) \Psi(\frac{t-\tau}{\alpha}) dt \qquad (2)$$

where $\Psi(t)$ is the wavelet base function with scaling factor α that controls the width of the wavelet and translation parameter τ that controls its time location.

We select the Ricker wavelet as the wavelet base function because it is frequently used to model seismic data (i.e., vibration signals) [33], [34], [36]. Figure 6 shows an example of tap-induced vibration signals and their CWT spectrum. For each channel, we will select one scale with the highest energy out of 50 scales. The index of the selected scale may be different for each channel. Then we will share the indexes across all channels for signal reconstruction. Instead of using L2 norm to calculate the energy as usual, we use L-infinity norm in this paper. The reason is that L2 norm sums up across the whole time-series and will incorporate the energy of undesired reflected paths that appear in the latter part of the signals. Therefore, we can increase the TDoA estimation accuracy of the direct path by using L-infinity norm instead. The third column in Figure 6 shows the vibration signals after CWT filtering. We can see that the clutter level of reconstructed signals is lower.

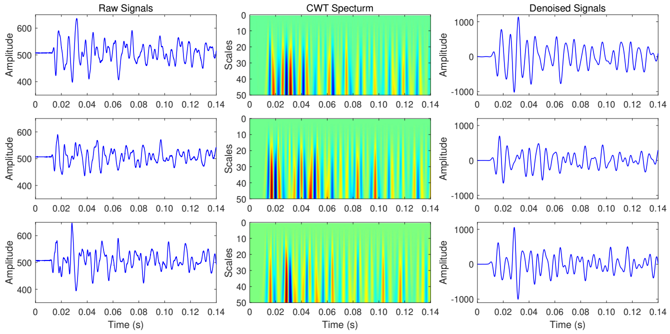


Fig. 6: The raw signals of the three channels are reconstructed using CWT.

B. Two-stage Time Delay Estimator

We design a two-stage time delay estimator that can more precisely estimate the TDoA values between sensors. First, we estimate the coarse-grained arrival time of the vibration signal using AIC. This step denotes a onset point of the signals and determines the signal arrival order at each sensor. The arriving orders at each sensor are essential prior knowledge that can improve localization accuracy. Second, we extract the partial signals (including the first two peaks) for cross-correlation to mitigate the multi-path effect. With the prior knowledge obtained in the first step, we can select the optimal correlation coefficient corresponding to the correct time delay.

1) Coarse Estimation with AR-AIC: The segmented vibration signals comprise two intervals. One is the ambient noise, and the other is the tap-induced vibration. We can express the segmented signals of the i-th channel as an autoregressive (AR) model.

$$x_i(t) = \sum_{m=1}^{M} a_j(m) x_i(t-m) + \varepsilon_j(t) \tag{3}$$

where M is the order of an AR process fitting the data, $a_j(m)$ are constant coefficients of the j-th interval, and $\varepsilon_j(t)$ is stationary white noise with zero mean and variance σ_j^2. Given the segment length N, we have $t \in [1, M]$ for interval 1 and $t \in [N-M+1, N]$ for interval 2.

Our target is to detect the exact onset time of interval 2 so that we can get a more fine-grained time delay estimation. To this end, we can calculate the Akaike information criterion (AIC) [37], which is represented as:

$$AIC(t) = (t-M)\log(\sigma^2_{1,\max}) + (N-M-t)\log(\sigma^2_{2,\max}) \tag{4}$$

The global minimum of the AIC curve indicates the onset point of interval 2. Figure 7 shows an example AIC curve of a segment and the onset point we find. In order to calculate the AIC efficiently, the Maeda method [38] is applied. However, the onset point is not evident, and the detection is inaccurate when the signal-to-noise ratio (SNR) is low (e.g., the tap position is far away from one of the sensors). Therefore, AIC can only provide a coarse estimation of the onset point, but this prior knowledge is helpful for the final estimation in the next stage.

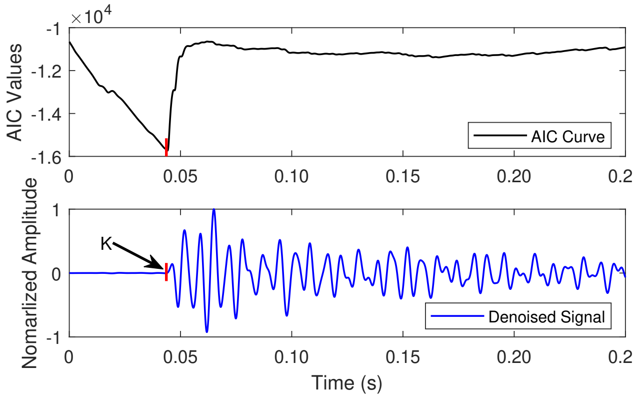


Fig. 7: A sample AIC curve. The minimum of AIC indicates the onset of tap-induced vibration.

2) PCC with Prior Knowledge: It is typical to calculate the cross-correlation between signals of two channels for time delay estimation [39], [40]. However, tap-induced vibration signals suffer reflection and refraction inside the surface. As shown in Figure 8, the latter part of the waveform compromises the influence of multipath and has a high clutter level, which makes the time delay estimation drift a lot. To avoid such influence, we only consider the initial periods (i.e., sample index from around 500 to 700 in Figure 8) for cross-correlation calculation. Specifically, the onset point is detected by AIC, and we set the endpoint at the second peak of the signals. Figure 8 shows the cross-correlation coefficients calculated between two extracted partial signals. Generally, the time delay is determined as the offsets corresponding to the maximum coefficients. However, this is sometimes not the case in practice. The maximum coefficient N_{max} in Figure 8 indicates a negative offset of -91. In fact, the positive offset of 23 (i.e., shift signals of geophone 2 toward the positive direction for 23 sample points) gives the correct time delay estimation when calculating the cross-correlation $xcorr(Geo\ 1, Geo\ 2)$ [41], because we know that the ground truth location of the touchpoint is closer to geophone 2. Actually, we can leverage the prior knowledge (i.e., the arriving order of tap signals at each sensor) obtained by AIC to eliminate this estimation error. The prior knowledge can help decide which peak in the cross-correlation coefficient curve is the correct one and avoid the localization with large errors due to inaccurate time delay estimation. The effectiveness of partial cross-correlation (PCC) with prior knowledge will be evaluated in Section V.

C. Synthetic Data Generation

Let us assume that there is a mechanical wave propagating in a straight line, and it propagates through two geophone sensors G_1 and G_2 one after the other at time T_1 and T_2, respectively. If we ignore the velocity parameter by setting it as 1, then we can define a new distance metric called time-difference distance (TDD) to characterize the distance between sensors. For example, the distance between G_1 and G_2 will be $T_1 - T_2$ TDD. TDD is measured by the time difference, and its physical meaning is the time it takes for a mechanical

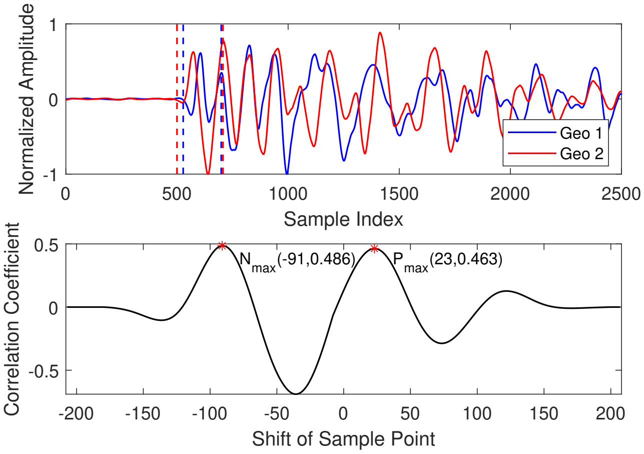


Fig. 8: Illustration of partial cross-correlation with prior knowledge. The dotted lines denote the partial signals used for cross-correlation.

wave to pass through two points on a straight line at a unit velocity. In this section, we set TDD as the default distance metric.

We now assume that there are three geophones G_1, G_2, and G_3. The distances between every two geophones are a and b, as shown in Figure 9(a). We now find any point M on the line G_1G_2, and tap at that point. The three geophones will receive the vibration signal and can calculate the time difference of arrival. We can set the distance between M and G_1 as x_1. The TDoA value measured at G_1 and G_2 can be expressed as:

$$TDoA_{M_{12}} = T_{M_1} - T_{M_2} = x_1 - (a - x_1) = 2x_1 - a \quad (5)$$

The TDoA value measured at G_1 and G_3 is $TDoA_{M_{13}} = x_1 - |MG_3|$. Therefore, we have $|MG_3| = |x_1 - TDoA_{M_{13}}|$. According to the Pythagorean theorem, we have:

$$x_1^2 + b^2 = (x_1 - TDoA_{M_{13}})^2 \quad (6)$$

We can find another point N on the line G_1G_3 to tap and set the distance between N and G_1 as x_2. Similarly, we can get another two equations:

$$TDoA_{N_{13}} = T_{N_1} - T_{N_3} = x_2 - (b - x_2) = 2x_2 - b \quad (7)$$

$$x_2^2 + a^2 = (x_2 - TDoA_{N_{12}})^2 \quad (8)$$

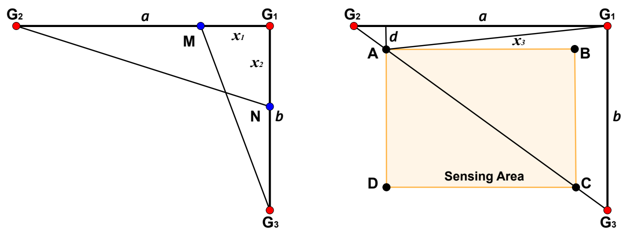
Solving Eq (5)-(8), we can know the length of a and b in TDD.

The next step is to determine the sensing area size, as shown in Figure 9(b). The solution is to find a third point A on the line G_2G_3, and we denote the distance between A and G_1 as x_3 TDD. Then we have $|AG_2| = |x_3 - TDoA_{A_{12}}|$ and $|AG_3| = |x_3 - TDoA_{A_{13}}|$. Based on Pythagorean theorem, we can get:

$$a^2 + b^2 = [(x_3 - TDoA_{A_{12}}) + (x_3 - TDoA_{A_{13}})]^2 \quad (9)$$

Leveraging the similar triangle theorem, we can get the distance d from the point A to the line G_1G_2 by:

$$\frac{d}{b} = \frac{x_3 - TDoA_{A_{12}}}{(x_3 - TDoA_{A_{12}}) + (x_3 - TDoA_{A_{13}})} \quad (10)$$



(a) Sensor Intervals (b) Sensing Area

Fig. 9: MM-Tap exploits a geometrical model to synthesize data with a few calibration points.

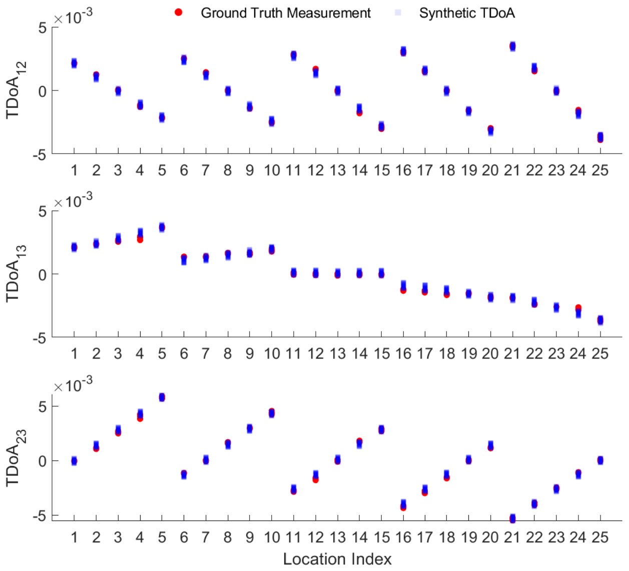
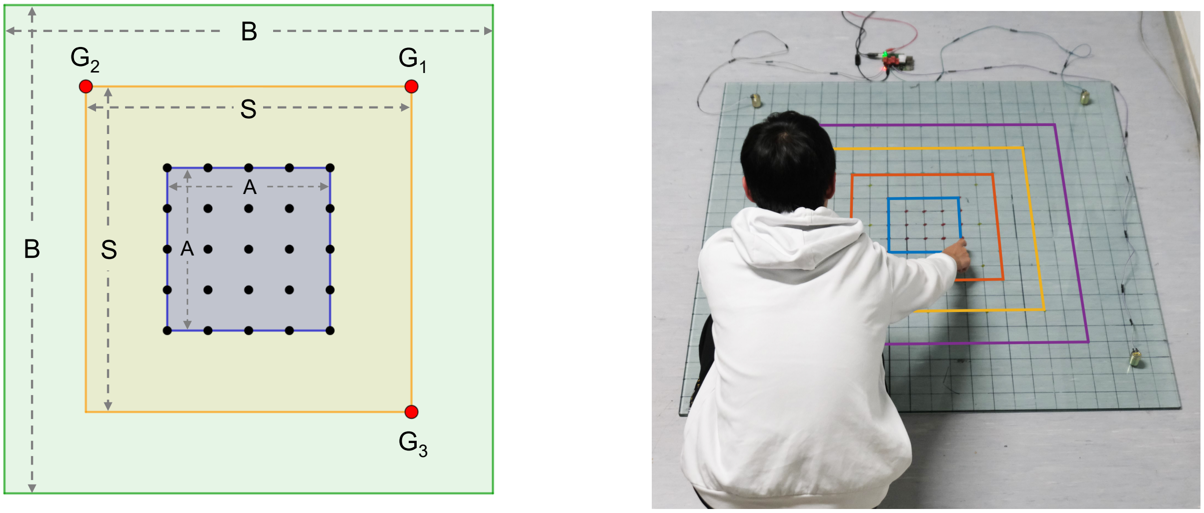


Fig. 10: A comparison between synthetic TDoA values and ground truth measurements from 25 locations on an acrylic board.

Ultimately, we can infer the width and length of sensing area as $|AD| = b - 2d$ and $|AB| = \frac{a}{b}(b - 2d)$, respectively. Taking G_1 as the origin, the coordinate of point A is $(-(x_3^2 - d^2)^{1/2}, -d)$, which can be estimated using the actual measurement of TDoA values. Then, we can synthesize the TDoA values of any location in the sensing area and learn a regression model to predict the tap location. However, synthesizing data in this way can not provide a satisfactory localization performance. This is because the pattern of the synthetic data with TDoA measurement from only one point A is a regular quadrangle, which is contradicted to the irregular quadrilateral pattern of ground truth measurement using three geophone sensors (see Figure 4(b)). We need to refine our method further.

Actually, we need at least four vertices to pinpoint the irregular quadrilateral. Since the sensor intervals a and b is known, let us assume the coordinates of the vertices of sensing area are $A(x_a, y_a)$, $B(x_b, y_b)$, $C(x_c, y_c)$, $D(x_d, y_d)$, which can be obtained by solving Chan equations [23] without considering velocity. The coordinates of the remaining locations in the sensing area can be inferred based on these four vertices. Assuming that the $\vec{AD}$ is divided into $n - 1$ segments and



(a) Denotation of surface layout.

(b) Real world setup.

Fig. 11: (a) The setting of surface layout, where B=120 cm is the board size, S=100 cm is the sensor intervals, and A is the sensing area size. The red and black dots denote geophone sensor locations and 25 tap locations, respectively. (b) The experiment setup of MM-Tap prototype under different A in the real world.

a total of n points, then for the i-th equal division point K^{i}_{AD}, we have $\vec{AK^{i}_{AD}} = \frac{i}{n-1}\vec{AD}$. Therefore, the coordinate of the i-th point on $\vec{AD}$ is $K^{i}_{AD} = \frac{i}{n-1}(D-A)+A, i \in [0, n-1]$. More generally, the coordinate of the point in the i-th row and the j-th column is:

$$X_{ij} = \frac{j}{m-1}(K^{i}_{BC} - K^{i}_{AD}) + K^{i}_{AD}, \quad j \in [0, m-1] \quad (11)$$

Figure 10 shows the synthetic TDoA values of 5×5 points using the samples from 6 calibration points, which matches the ground truth measurement well. In practical use, vertices of the sensing area can be easily indicated for users on a projected interface. The mathematical theory builds upon the Pythagorean theorem, but our system does not require a perfectly orthogonal layout in practice (see validation in Section V). Note that the proposed synthetic data generation model is applicable when the number of the sensor is 3. Without introducing extra human effort, a fourth sensor does not provide additional information and will not improve localization accuracy. Therefore, we only investigate the case using three sensors in compliance with our design goals. In summary, with our method, users no longer need to measure the sensor intervals to get the exact coordinate of ambient sensors when deploying the tap sensing system. A large amount of synthetic data can be generated within a few seconds, supporting rapid adaptation to new environments and sensing scales. Actually, we generate 15 synthetic samples for each location. Gaussian noise with a mean value of 0 and standard deviation of 10^{-5} is added to the synthetic data for better generalization.

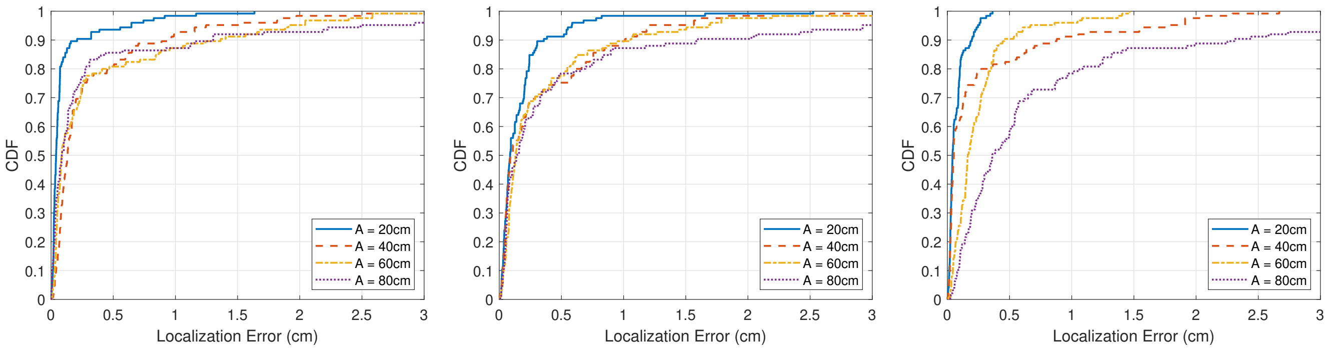
D. Tap Localization

With sufficient synthetic TDoA data, we can learn the spatio-temporal mapping relationship of the whole surface sensing area using the regression model. In this paper, we adopt Gaussian process regression (GPR) [43] as our learning model. GPR has been successfully applied in fingerprint-based indoor localization systems using received signal strength indicator (RSSI) [?], [44], [45]. It is a non-parametric model that does not build upon the discrete representation of space and is able to represent arbitrary probabilistic models. The three-channel synthetic TDoA data is used to train the GPR model with the linear basic function and the rational quadratic kernel. The model is optimized with the Bayesian optimization scheme and iterated for 30 epochs. Two GPR models are trained to predict the tap location's x-axis and y-axis coordinates, respectively.

V. Evaluation

A. Implementation

Three geophone sensors (LCT-20D100) are used to detect tap-induced vibration on the surface. The three-channel analog signals are then amplified by three amplifiers (BOB-09816), respectively. A 10-bit 8-channel analog-to-digital converter (MCP3008) is used for digitizing the analog signals. The above-mentioned components are connected to a daughter-board, which can directly plug into a Raspberry Pi 3B. We configure the sampling rate as 30 kHz using the BCM2832 library in C. The real-time vibration signals are transmitted to a laptop (Dell G7 7588-R1745) for further processing in the MATLAB platform. An interactive projector is implemented by connecting a COTS projector (XGIMI XK03E) to the laptop. The projected surface is completely overlapped with the sensing area supported by MM-Tap. We conduct a user study and ask volunteers to play video games using this interactive projector.



(a) Glass

(b) Acrylic

(c) Wooden

Fig. 12: Localization errors across different sensing scales on three different types of surfaces.

B. Experimental Setup

Without loss of generality, we prepare three different surface materials for evaluation, including toughened glass, acrylic, and wood[1]. These surfaces have the same dimension of 120 cm × 120 cm × 1.5 cm. The layout setting is shown in Figure 11. Three geophone sensors are fixed on the corner of the surface to cover an area of 100 cm × 100 cm. For an effective sensing area, we arrange the sensing area size A from 20 cm to 80 cm with a step size of 20 cm. We uniformly distribute 25 touchpoints on the effective sensing area where each touchpoint is separated by $D = A/4$ cm on both the x-axis and y-axis. For a specific sensing area on a surface, we use the fingertip to tap on each touchpoint 10 times. Then, we form a main dataset with $3 \times 4 \times 25 \times 10 = 3000$ tap samples in total.

To study the impact of other factors (e.g., tap tool material, etc.), we collect other sub-datasets with a default setting where only acrylic surface is considered and A = 60 cm. We will specify more about the setup details in the corresponding experiments.

C. Accuracy

1) Over performance: We first demonstrate the overall localization errors of MM-Tap across three types of surfaces and four sensing area sizes. Figure 13 shows the CDF of localization errors, which reveals following findings:

(1) Overall, we can see that the 80th percentile localization errors fall within 1 cm, which indicates that MM-Tap can facilitate mm-level tap sensing on ubiquitous surfaces.

(2) For different sizes of sensing area, we can observe that the tap in the center (e.g., A = 20 cm) has a better localization performance. This is because the distance between the touchpoints and sensors is relatively similar, and the signal quality is high (i.e., much less attenuation and dispersion) for all channels. With the increase of the sensing area, the localization errors slightly increase.

(3) For different surface materials, the wooden surface shows more fluctuating localization errors under different sizes of sensing area. The primary reason is that the vibration wave in the wooden board has a higher velocity and velocity fluctuation range.

(4) For the best case (A = 20 cm), the 90th percentile localization errors are 2.09 mm, 3.59 mm, amd 1.96 mm for glass, acrylic, and wooden surface, respectively. For the worst case (A = 80 cm), the 90th percentile localization error is 1.30 cm, 1.60 cm, and 2.3 cm for glass, acrylic, and wooden surfaces, respectively.

2) Comparison: We also compare MM-Tap with two baselines. The Fingerprints baseline is evaluated using leave-six-out cross-validation. The SurfaceVibe baseline is the SOTA vibration-based tap sensing system. Figure 13 plots the localization errors by taking all the test samples in different sensing area sizes into consideration. We can see that the Fingerprints baseline provides the best performance, but the tedious data collection process is infeasible in the actual application. Overall, MM-Tap outperforms the SOTA vibration-based tap sensing system and provides much more accurate and stable localization results.

3) Effectiveness of Time Delay Estimator: In this experiment, we evaluate the proposed modification for estimating time delay discussed in Section IV-B. Figure 14 compares the median localization errors using three different methods for time delay estimation. We can observe that calculating cross-correlation across the whole time-series (i.e., Corr) yields the worst performance. The errors further increase with the increase of sensing area size, where the multipath effect has more impact on the signal waveform. This effect is mitigated by only considering the front part of the segment (i.e., PCC). We can see a further improvement when leveraging the prior knowledge obtained by AIC (i.e., PCC w/ PK), which validates our method's effectiveness for fine-grained time delay estimation.

D. Adaptability and Scalability

1) Localization across Different Surfaces: One of the design goals of MM-Tap is to guarantee the adaptation on ubiquitous surfaces. In this section, we first investigate how the environment changes (i.e., different surface materials) impact the system performance. For example, we use the samples

[1] We measure the wave velocity (m/s) of different points on our boards, where $VFR_{glass} \in [108, 171]$, $VFR_{acrylic} \in [57, 137]$, $VFR_{wood} \in [141, 269]$.

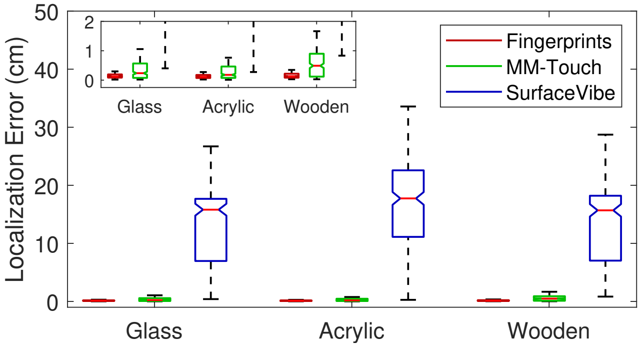


Fig. 13: Localization errors comparison against different baselines.

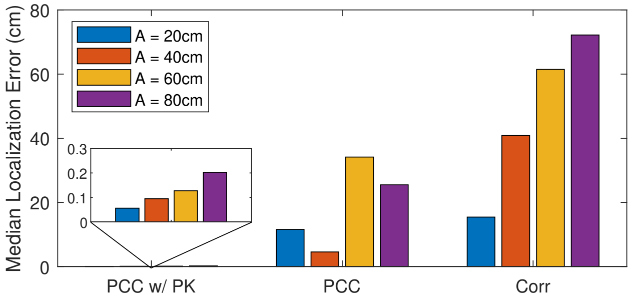


Fig. 14: Comparison between different time delay estimation methods.

collected from one surface material to test the regression models trained on another surface material. As shown in Figure 15(a), the trained regression models yield unacceptable median localization errors over 9 cm when testing with samples from unseen environments. This indicates that the synthetic data can not generalize across all the surface materials, and we need to calibrate the model. Figure 15(b) shows how the system performance is recovered with a different number of samples from each calibration point. With one tap at each calibration point, the median localization errors can recover to below 6 mm across all surfaces.

2) Localization across Different Scales: We also want MM-Tap to be responsive for different sizes of sensing area. Similar to the previous experiment, Figure 16 shows how the scale changes impact the system performance. Again, MM-Tap fails to work across different scales, but the system can quickly adapt and recover to a standard performance after the calibration. Interestingly, using more samples for calibration is not necessary for higher localization accuracy. In practical use, users can scale the interactive interface to an arbitrary size by tapping each calibration point one time only.

E. Generalizability

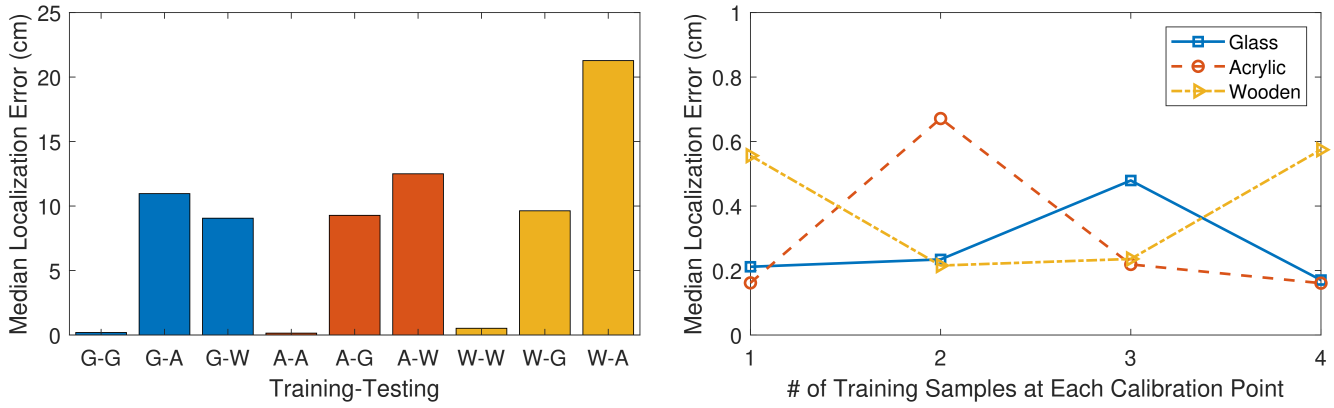
1) Impact of sensor displacement: One of the key features of MM-Tap is that users do not need to measure the exact coordinate of ambient sensors. Without the restriction of sensor deployment, users may introduce errors in calculating synthetic data. To simulate the human error, we keep sensor G_1 still and shift sensor G_2 and G_3 along the line G_2G_3 simultaneously. Both sensors are shifted inwards to the board center and outwards to the board edge for 8 cm with a step size of 2 cm. Figure 17 shows the localization performance when shifting sensors to different levels. The results validate that MM-Tap is resilient to around 8 cm sensor displacement, having a high fault tolerance to sensor deployment.

2) Impact of Tap Strength: In this experiment, we investigate how tap strength will affect the system performance. Specifically, we consider two different strength levels–”heavy” and ”gentle.” The average SNR of vibration signals collected with heavy tap strength is considerably higher than that of gentle ones. Figure 18 shows the localization error when applying different tap strengths. It is obvious that the system performance suffers no degradation. The reason is that the variation of signal pattern due to different tap strength levels is not an influence factor for extracting exact TDoA values.

3) Impact of Tap Materials: In practice, users may utilize different tap tools when interacting with the touchscreen. In this experiment, we iterate six kinds of tap tools for interaction, including the fingertip, the fingernail, an eraser, an iron pen, a plastic marker pen, and a wooden pencil. As shown in Figure 19, MM-Tap maintains stable and high localization accuracy for most of the tap materials compared to the baseline (fingertip). However, we can see that MM-Tap has high errors sometimes when using an iron pen as the tap tool. We analyze the iron pen-induced vibration signals and find that the frequency band is much broader than others. The dispersion effect degrades the estimation accuracy of TDoA values.

4) Impact of Surrounding Objects: In this experiment, we simulate the practical scenario where the objects exist nearby the surface. Specifically, we consider three kinds of objects with different sizes and weights (e.g., A 1.3kg 13-inch laptop, a 194g 6.81-inch smartphone, and a 2kg A4 book). We divide the surface into three areas (i.e., blue, orange, and green areas in Figure 11(a)) to indicate different influence levels. Three objects are randomly placed in different locations of the corresponding area. Figure 20 compares the localization errors with or without objects placed on the surface. MM-Tap shows high robustness against the environmental changes above surfaces. Unlike classification-based localization system [14], [15] that relies on a stable signal pattern of each location, MM-Tap can still extract fine-grained TDoA values of directly arriving path even if the pattern is changed.

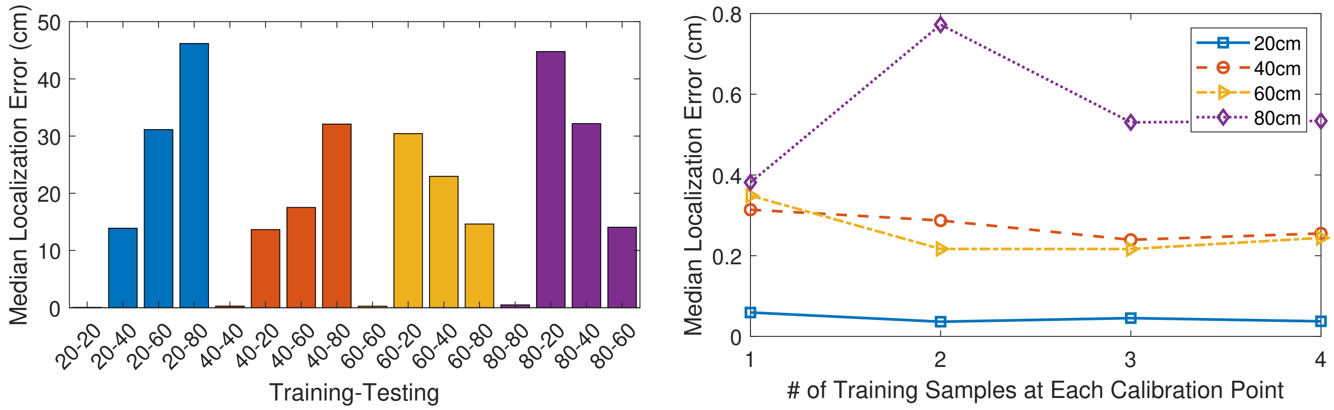
5) Impact of Ambient Noise: We consider the impact of two types of ambient noise, namely, air-borne and solid-borne noise. A Xiaomi smart speaker Pro is placed one meter away on the other table to generate air-borne noise. The solid-borne noise is generated by placing the speaker at the corner that has no sensor. A Happy Birthday song is played when collecting data. We measure the sound pressure level at the speaker using a sound meter (AR844). The sound pressure level ranges from 60 dB to 100 dB with a step size of 10 dB. Figure 21 shows that MM-Tap is resilient to both air-borne and solid-borne noise.



(a) Without adaptation

(b) With adaptation

Fig. 15: Localization errors across different surfaces, where ”G” for glass, ”A” for acrylic, and ”W” for wood.



(a) Without adaptation

(b) With adaptation

Fig. 16: Localization errors across different scales, where sensing area size A = {20, 40, 60, 80} (unit=cm).

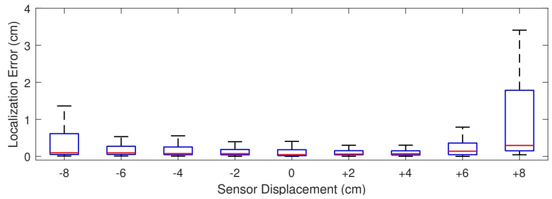


Fig. 17: Impact of sensor displacement (”+” for inwards, while ”-” for outwards).

F. User Study

An anonymous demo showing our user study’s game applications is available at: https://youtu.be/nQBnXOpntsc.

1) User Study Setup: We recruit ten volunteers (2 of them are female) from our university for the user study. The user study contains four sessions, including two sessions of normative tests by tapping randomly generated points and two sessions of playing video games. A projector and three geophone sensors are used to create a 20-inch digitally augmented interface on a wooden table. A 10-inch tablet (Surface Go) is used as a baseline for comparing MM-Tap with a capacitive touchscreen. Volunteers are asked to alternately use MM-Tap and touchscreen to play 10 rounds of each game. During the user study, a camera is on to record every interactive operation.

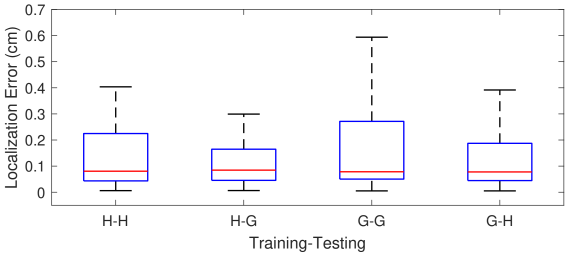


Fig. 18: Impact of tap strength (”H” indicates ”heavy”, while ”G” indicates ”gentle”).

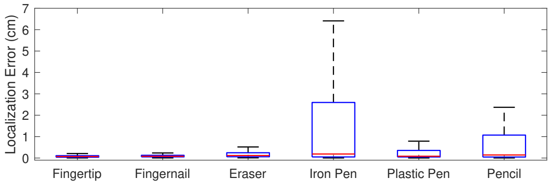


Fig. 19: Impact of different tap materials.

2) Normative Test: Tapping Random Points: We conduct this experiment under two calibration settings: self-calibration

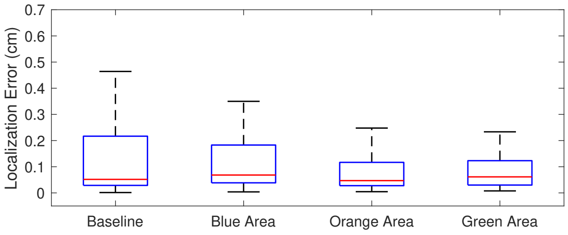


Fig. 20: Impact of surrounding objects.

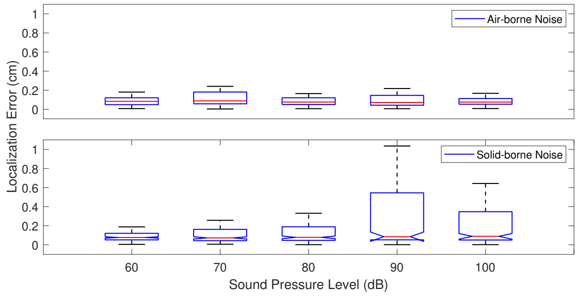


Fig. 21: Impact of ambient noise.

and calibration by others. For self-calibration, the volunteers are taught to set up the MM-Tap system by themselves. For the other session, the instructor will calibrate the system before volunteers use it. For each session, a circle with a diameter of 1 cm will randomly pop up on the interface 300 times. The volunteers are asked to tap this projected circle. Overall, the average time used for self-calibration is 10.5 s. Figure 22 shows the localization errors of each volunteer. All the volunteers can achieve mm-level tap localization accuracy using MM-Tap when calibrating the system by themselves. However, some of the volunteers (e.g., V2, V3, V7, and V10) yield unstable localization results when using the system calibrated by others. But still, the median localization errors fall within 1 cm for these four volunteers. We also counted the miss detection rate and false alarm rate during this normative test in Table I, which shows a high detection accuracy of MM-Tap.

3) Game 1: Whack a Mole: Whack a Mole [49] is a well-known tap-based game, which requires players to tap the randomly pop-up moles within a certain time for scoring. In this study, we ask volunteers to play Whack a Mole for one minute per round, and each effective hit accounts for one score. Figure 23(a) shows the average scores for each volunteer. Most of the volunteers can get comparable scores when using both interaction methods. Due to the larger interface, they feel it will be easier to score a hit on MM-Tap. However, we notice that some of the volunteers (e.g., V6, V7, and V8) have much lower scores when using MM-Tap. They report that moles have a short existence time, and the screen is too large to respond sometimes, and many hits were missed.

TABLE I: The detection accuracy of MM-Tap.

Item	Self-Cali.	Cali. by others
Miss Detection Rate(%)	3.06%	2.60%
False Alarm Rate(%)	0%	0%

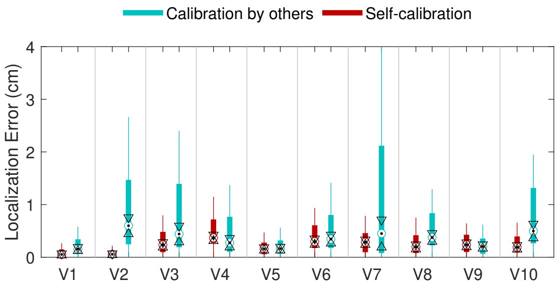
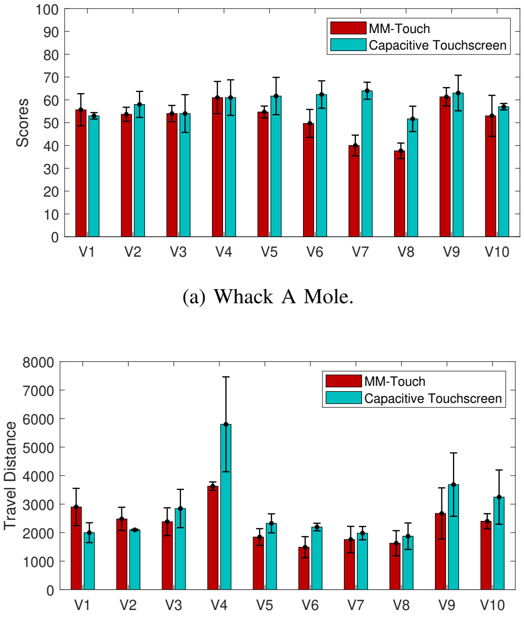


Fig. 22: The localization errors with respect to two calibration settings.



(a) Whack A Mole.

(b) Smash Hit.

Fig. 23: Comparison of user scores between MM-Tap and the capacitive touchscreen.

4) Game 2: Smash Hit: Smash Hit [42] is a 3D shooting game involving an increasingly-moving view through a passageway. The players need to tap the screen to aim and shoot limited metal balls to smash the crystals or obstacles to get bonus balls for surviving a longer time. If a player does not smash the obstacles encountered in time, he will lose some balls and die when the ball pool is empty. In this study, we ask volunteers to play as long as they can. Figure 23(b) presents the travel distance given by the game when each player dies. Overall, the score of playing Smash Hit on MM-Tap will be slightly lower than that of using the touchscreen. Volunteers report that the latency of MM-Tap is a little bit higher and more unstable than that of capacitive touchscreen, and the predictive aiming for shooting is more challenging when the moving speed of view is faster in the later period of the game. For each tap, MM-Tap shows an average calculation latency of 82.9 ms with a standard deviation of 28.7 ms inside the MATLAB. We notice a large transmission latency from the laptop to the projector in our run-time demo, and we believe the latency can be much lower with a better design of the prototype in the future.

VI. Discussion

A. Finger Swipe Tracking

Previous work [16] tracks the swipe trajectory by continuously localizing finger location using multiple signal segments within a swipe process. The swipe-induced vibration wave is dominated by body waves whose attenuation is proportional to the propagation distance. Therefore, it has a lower signal-noise ratio when reaching the sensor compared to tap-induced vibration waves. The swipe-induced is hard to detect for a large-scale tap sensing system like MM-Tap. In addition, the latency will be unacceptably high due to the computation of continuous localization. A new tracking method needs to be proposed in the future.

B. Irregular Surface

For irregular surfaces, we can discuss three situations. The first one is the surface is flat, but the shape is irregular (e.g., a star shape or a cross shape). MM-Tap can still work when sensors cover a regular quadrangle area because of having a learnable TDoA pattern. The second one is rough and uneven surfaces (e.g., hills and valleys), which can be considered as a 3D surface. However, MM-Tap can only predict the 2D coordinate on the flat surface. There is also a surface that is formed by multiple splicing plates. TDoA values are inaccurate in this case and can not provide a good tap sensing performance. The errors will be higher when the splicing plates are of different materials.

C. Multiple Points Interaction

Recognizing multiple vibration signals is a typical cocktail-party problem [46]. We have to separate multiple vibration signals using the recorded temporal signals of geophone sensors. [48] localizes up to 3 people's footstep vibration signals in an indoor environment by assuming that the moments when different people's feet hit on the ground are not exactly the same. Therefore, signals are separable in the time domain with high sampling frequency. However, vibration signals of the multi-touch interaction are typically overlapped and hard to be separated directly. We plan to leverage blind source separation techniques [46], [47] to recover the original vibration signals of different sources and support multiple points interaction in the future.

VII. Related Work

We focus on the related work that senses the direct contact between fingers and the physical surface. The tap interaction system can be broadly classified into two classes: instrument-based and instrument-free.

A. Instrument-based

Instrument-based tap sensing systems require users to wear special equipment, which is typically a signal source generator. VersaTouch [1] deploys a vibration microphone on the user's fingernail to transmit active signals to nearby piezoelectric receivers and realize mm-level finger tracking within a circular area of 40cm diameter. Acustico [3] specializes a wrist-worn device with four acoustic sensors to detect and localize finger taps across different surfaces. However, it requires a 1 MHz sampling rate for accurate TDoA estimation, and its interaction is limited since the user's hand cannot move around. ElectroRing [4] presents a wearable ring with electrodes and an IMU sensor. It can detect subtle finger pinch and track the finger touch on conductive surfaces. ItrackU [5] designs a surface tracking system with the fusion of ultra-wide band (UWB), inertial measurement unit (IMU), and pressure sensor on a pen, achieving a 90th percentile error of 7 mm in an area of 2.5 m × 2m. Instrument-based systems do provide a better interaction experience in terms of detection and localization accuracy, but the extra device may sometimes be cumbersome to the users, and the power supply is another critical issue for such systems. In contrast, MM-Tap supports instant interaction without wearing any extra devices.

B. Instrument-free

Instrument-free systems attempt to enable sensing by modifying the surface itself [7]–[9], crafting an interface similar to traditional capacitive touchscreens. In addition, vision-based techniques [10]–[13] are also widely adopted to detect finger touch on surfaces. Another trend is to deploy ambient sensors around the sensing area and analyze the readings when users are interacting on the surface. UbiK [14] utilizes dual microphones on a mobile device to capture the acoustic signal of finger taps. It further extracts location-dependent features from the signals to train a classification model and realize a virtual keyboard on ubiquitous surfaces. VibSense [15] deploys a single piezoelectric sensor to receive the surface vibration of a finger tap and applies SVM to determine the keystroke location. VSkin [18] supports fine-grained 1D finger tracking and tapping recognition on the back surface of a mobile phone by analyzing the amplitude and phase of sound signals. PACE [53] deploys a 6-mic array to collect structure-borne and air-borne footstep impact sounds (FIS) for small-scale indoor scenarios, demonstrating a sub-meter localization accuracy with a median error of 30 cm. Ubitap [17] exploits the dispersion phenomenon and collects both surface-borne and air-borne acoustic signals of finger taps with accelerometer and microphone sensors, respectively. The system uses three standalone smartphones to cover an area of 24cm× 36cm and can perform accurate TDoA triangulation of finger taps on different surfaces. Acoustic-based systems have high accuracy, but it is hard to scale up the sensing range. In contrast, MM-Tap supports a much larger sensing range up to 80 cm × 80 cm with similar performance. The most relevant work to us is SurfaceVibe [16], which deploys four geophone sensors to conduct TDoA triangulation and estimate the finger tap location and swipe trajectory with cm-level accuracy. On the other hand, MM-Tap adopts a new localization scheme that can realize mm-level tap sensing with fewer geophone sensors. In addition, users no longer need to measure the exact coordinate of ambient sensors when deploying MM-Tap.

VIII. Conclusion

In this paper, we propose MM-Tap push the limits of vibration-based tap sensing on ubiquitous surfaces with mm-level accuracy. MM-Tap can transform a normal flat surface into a scalable virtual interface in a low-cost manner. MM-Tap builds upon a novel localization scheme that constructs a mapping relationship between TDoA values and tap locations. We exploit the geometry of the sensor layout and propose a novel synthetic data generator and an effortless calibration scheme. Our comprehensive experiments validate that MM-Tap can adapt to varying surface material and scale to arbitrary sensing area size after calibration within a few seconds.

References

[1] Y. Shi and H. Zhang, et al.,”VersaTouch: A versatile plug-and-play system that enables touch interactions on everyday passive surfaces,” in Proceedings of the Augmented Humans International Conference, 2020, pp. 1-12.

[2] Y. Gu and C. Yu, et al.,”Accurate and low-latency sensing of touch contact on any surface with finger-worn imu sensor,” in Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, 2019, pp. 1059-1070.

[3] J. Gong and A. Gupta, et al.,”Acustico: Surface tap detection and localization using wrist-based acoustic tdoa sensing,” in Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology, 2020, pp. 406-419.

[4] W. Kienzle and E. Whitmire, et al.,”ElectroRing: Subtle pinch and touch detection with a ring,” in Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 2021, pp. 1-12.

[5] Y. Cao and A. Dhekne, et al.,”ITrackU: tracking a pen-like instrument via UWB-IMU fusion,” in Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services, 2021, pp. 453-466.

[6] M. Meier and P. Streli, et al.,”TaplD: Rapid touch interaction in virtual reality using wearable sensing,” in 2021 IEEE Virtual Reality and 3D User Interfaces (VR), 2021, pp. 519-528.

[7] Y. Zhang and G. Laput, et al.,”Electrick: Low-cost touch sensing using electric field tomography,” in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017, pp. 1-14.

[8] S. Swaminathan and J. Fagert, et al., ”OptiStructures: Fabrication of room-Scale interactive structures with embedded fiber bragg Grating Optical Sensors and Displays,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 4, no. 2, pp. 1-21, 2020.

[9] Y. Zhang and C. Yang, et al.,”Wall++ room-scale interactive and context-aware sensing,” in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1-15.

[10] Y. Zhang and W. Kienzle, et al.,”ActiTouch: Robust touch detection for on-skin AR/VR interfaces,” in Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology, 2019, pp. 1151-1159.

[11] R. Xiao and S. Hudson, et al., ”Supporting responsive cohabitation between virtual interfaces and physical objects on everyday surfaces,” Proceedings of the ACM on Human-Computer Interaction, vol. 1, no. EICS, pp. 1-17, 2017.

[12] R. Xiao and J. Schwarz, et al., ”MRTouch: Adding touch input to head-mounted mixed reality,” IEEE transactions on visualization and computer graphics, vol. 24, no. 4, pp. 1653-1660, 2018.

[13] C. Harrison and H. Benko, et al.,”OmniTouch: wearable multitouch interaction everywhere,” in Proceedings of the 24th annual ACM symposium on User interface software and technology, 2011, pp. 441-450.

[14] J. Wang and K. Zhao, et al.,”Ubiquitous keyboard for small mobile devices: harnessing multipath fading for fine-grained keystroke localization,” in Proceedings of the 12th annual international conference on Mobile systems, applications, and services, 2014, pp. 14-27.

[15] J. Liu and Y. Chen, et al.,”Vibsense: Sensing touches on ubiquitous surfaces through vibration,” in 2017 14th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), 2017, pp. 1-9.

[16] S. Pan and C. G. Ramirez, et al.,”Surfacevibe: vibration-based tap & swipe tracking on ubiquitous surfaces,” in 2017 16th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), 2017, pp. 197-208.

[17] H. Kim and A. Byanjankar, et al.,”UbiTap: Leveraging acoustic dispersion for ubiquitous touch interface on solid surfaces,” in Proceedings of the 16th ACM Conference on Embedded Networked Sensor Systems, 2018, pp. 211-223.

[18] K. Sun and T. Zhao, et al.,”Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals,” in Proceedings of the 24th Annual International Conference on Mobile Computing and Networking, 2018, pp. 591-605.

[19] J. Xu and M. Ma, et al.,”Position estimation using UWB TDOA measurements,” in 2006 IEEE International Conference on Ultra-Wideband, 2006, pp. 605-610.

[20] S. Y. Jung and S. Hann, et al., ”TDOA-based optical wireless indoor localization using LED ceiling lamps,” IEEE Transactions on Consumer Electronics, vol. 57, no. 4, pp. 1592-1597, 2011.

[21] J. Smith and J. Abel, ”Closed-form least-squares source location estimation from range-difference measurements,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 35, no. 12, pp. 1661-1669, 1987.

[22] H. C. Schau and A. Z. Robinson, ”Passive source localization employing intersecting spherical surfaces from time-of-arrival differences,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 35, no. 8, pp. 1223-1225, 1987

[23] Y. T. Chan and K. C. Ho, ”A simple and efficient estimator for hyperbolic location,” IEEE Transactions on signal processing, vol. 42, no. 8, pp. 1905-1915, 1994.

[24] L. Yang and K. C. Ho, ”An approximately efficient TDOA localization algorithm in closed-form for locating multiple disjoint sources with erroneous sensor positions,” IEEE Transactions on Signal Processing, vol. 57, no. 12, pp. 4598-4615, 2009.

[25] I. A. Viktrov, ”Rayleigh and Lamb waves: physical theory and applications,” Chapter II, 1967.

[26] R. Jota and A. Ng, et al.,”How fast is fast enough? a study of the effects of latency in direct-touch pointing tasks,” in Proceedings of the sigchi conference on human factors in computing systems, 2013, pp. 2291-2300.

[27] Y. Huang and K. Wu, ”Vibration-based pervasive computing and intelligent sensing,” CCF Transactions on Pervasive Computing and Interaction, pp. 1-21, 2020.

[28] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, Princeton University Press, 2009.

[29] P. S. Addison, The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance, CRC press, 2017.

[30] H. Jeong and Y. S. Jang, ”Wavelet analysis of plate wave propagation in composite laminates,” Composite Structures, vol. 49, no. 4, pp. 443-450, 2000.

[31] H. Jeong and Y. S. Jang, ”Fracture source location in thin plates using the wavelet transform of dispersive waves,” IEEE transactions on ultrasonics, ferroelectrics, and frequency control, vol. 47, no. 3, pp. 612-619, 2000.

[32] M. Mirshekari and S. Pan, et al.,”Characterizing wave propagation to improve indoor step-level person localization using floor vibration,” in Sensors and smart structures technologies for civil, mechanical, and aerospace systems 2016, 2016, pp. 980305.

[33] A. Chakraborty and D. Okaya, ”Frequency-time decomposition of seismic data using wavelet-based methods,” Geophysics, vol. 60, no. 6, pp. 1906-1916, 1995.

[34] Ricker wavelet. [Online]. Available: https://en.wikipedia.org/wiki/Ricker_wavelet

[35] Number of IoT connected devices worldwide 2019-2030. [Online]. Available: https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/

[36] S. Sinha and P. S. Routh, et al., ”Spectral decomposition of seismic data with continuous-wavelet transform,” Geophysics, vol. 70, no. 6, pp. P19-P25, 2005.

[37] R. Sleeman and T. Van Eck, ”Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings,” Physics of the earth and planetary interiors, vol. 113, no. 1-4, pp. 265-275, 1999.

[38] N. Maeda, ”A method for reading and checking phase times in auto-processing system of seismic wave data,” Zisin, vol. 38, pp. 365-379, 1985.

[39] J. Benesty, et al., ”Time-delay estimation via linear interpolation and cross correlation,” IEEE Transactions on Speech and Audio Processing, vol. 12, no. 5, pp. 509-519, 2004.

[40] J. Ianniello, ”Time delay estimation via cross-correlation in the presence of large estimation errors,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 30, no. 6, pp. 998-1003, 1982.

[41] xcorr:Cross-correlation. [Online]. Available: https://ww2.mathworks.cn/help/matlab/ref/xcorr.html?lang=en

[42] Smash Hit. [Online]. Available: http://www.smashhitgame.com/

[43] C. E. Rasmussen,”Gaussian processes in machine learning,” in Summer school on machine learning, 2003, pp. 63-71.

[44] S. Yiu and M. Dashti, et al., ”Wireless RSSI fingerprinting localization,” Signal Processing, vol. 131, pp. 235-244, 2017.

[45] S. He and S. H. G. Chan, ”Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons,” IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 466-490, 2015. hahnel2006gaussianGRP3 B. F. D. Hahnel and D. Fox,”Gaussian processes for signal strength-based location estimation,” in Proceeding of robotics: science and systems, 2006.

[46] A. Hyvarinen and E. Oja, ”Independent component analysis: algorithms and applications,” Neural networks, vol. 13, pp.411-430, 2000.

[47] P. Comon and C. Jutten, Handbook of blind source separation: Independent component analysis and applications, Academic press, 2010.

[48] L. Shi and M. Mirshekari, et al.,”Device-free multiple people localization through floor vibration,” in Proceedings of the 1st ACM International Workshop on Device-Free Human Sensing, 2019, pp. 57-61.

[49] Whac-A-Mole. [Online]. Available: https://en.wikipedia.org/wiki/Whac-A-Mole

[50] K. Lee and D. Chu, et al.,”Outatime: Using speculation to enable low-latency continuous interaction for mobile cloud gaming,” in Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, 2015, pp. 151-165.

[51] O. Nishizawa and G. Kitagawa, ”An experimental study of phase angle fluctuation in seismic waves in random heterogeneous media: time-series analysis based on multivariate AR model,” Geophysical Journal International, vol. 169, no. 1, pp. 149-160, 2007.

[52] H. Sato, M. Fehler, and T. Maeda, ”Seismic wave propagation and scattering in the heterogeneous earth,” Springer Science & Business Media, 2012.

[53] C. Cai, H. Pu, P. Wang, Z. Chen and J. Luo, ”We Hear Your PACE: Passive acoustic localization of multiple walking persons ,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 5, no. 2, pp. 1-24, 2021.

Yandao Huang is currently working toward the Ph.D. degree at the Hong Kong University of Science and Technology. Before that, he received the bachelor’s degree in computer science and technology from Shenzhen University, China, in 2020. His research interests include Mobile and Ubiquitous Computing, Human Activities Recognition (HAR), Human-computer Interaction (HCI), and Smart Healthcare.

Cong Li received the BEng degree and BS degree from Shenzhen University, Shenzhen, China, in 2021. He is working toward the postgraduate degree in the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. His research interests include mobile computing and human-computer interaction.

Fuwen Chen is working toward the master degree in the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China.

Kaishun Wu received his PhD degree in computer science and engineering at the Hong Kong University of Science and Technology. Before joining HKUST(GZ) as a full professor at DSA Thrust and IoT Thrust in 2022, he was a distinguished Professor and Director of Guangdong Provincial Wireless Big Data and Future Network Engineering Center at Shenzhen University. Prof. Wu is an active researcher with more than 200 papers published on major international academic journals and conferences, as well as more than 100 invention patents, including 12 from the USA. He received the 2012 Hong Kong Young Scientist Award, the 2014 Hong Kong ICT Awards: Best Innovation, and 2014 IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award. He is an IEEE, IET, and AAIA Fellow.

Qian Zhang (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in computer science from Wuhan University, Wuhan, China, in 1994, 1996, and 1999, respectively. She is currently a Tencent Professor of Engineering and the Chair Professor with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong, where she is also serving as the Co-Director of Huawei-HKUST innovation lab and the Director with the Digital Life Research Center. Before that, she was with Microsoft Research Asia, Beijing, China, where she was the Research Manager with the Wireless and Networking Group. She has published more than 400 refereed papers in international leading journals and key conferences in the areas of wireless/Internet multimedia networking, wireless communications and networking, wireless sensor networks, and overlay networking. She is the inventor of more than 50 granted international patents. Her current research interests include Internet of Things, smart health, mobile computing and sensing, wireless networking, as well as cybersecurity.